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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 February 2020 (27.02.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/041751 A1

(51) International Patent Classification:
C12N 9/00 (2006.01) *C12N 9/24* (2006.01)
C12N 9/22 (2006.01) *C12N 15/62* (2006.01)

(21) International Application Number:
PCT/US2019/047996

(22) International Filing Date:
23 August 2019 (23.08.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/722,057 23 August 2018 (23.08.2018) US
62/886,937 14 August 2019 (14.08.2019) US

(71) Applicants: **THE BROAD INSTITUTE, INC.** [US/US]; 415 Main Street, Cambridge, MA 02142 (US). **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US).

(72) Inventors: **LIU, David, R.**; 3 Whitman Circle, Lexington, MA 02420 (US). **WANG, Tina**; 730 Williamson Street, Apt. 420, Madison, WI 53703 (US). **MILLER, Shannon**; 22 Concord Avenue, Somerville, MA 02143 (US).

(74) Agent: **MCCOOL, Gabriel, J.** et al.; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

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

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

(54) Title: CAS9 VARIANTS HAVING NON-CANONICAL PAM SPECIFICITIES AND USES THEREOF

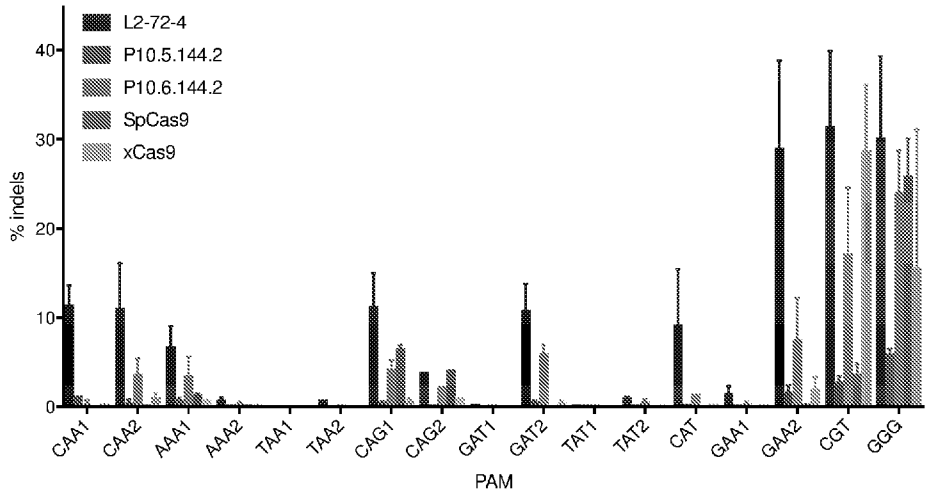


FIG. 18

(57) Abstract: Some aspects of this disclosure provide strategies, systems, reagents, methods, and kits that are useful for engineering Cas9 and Cas9 variants that have increased activity on target sequences that do not contain the canonical PAM sequence. In some embodiments, fusion proteins comprising such Cas9 variants and nucleic acid editing domains, e.g., deaminase domains, are also provided.

WO 2020/041751 A1

[Continued on next page]

Published:

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

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

CAS9 VARIANTS HAVING NON-CANONICAL PAM SPECIFICITIES AND USES THEREOF

RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application No. 62/722,057 filed August 23, 2018, and to U.S. Provisional Patent Application No. 62/886,937, filed August 14, 2019, each of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] CRISPR-Cas systems, and especially systems based on the Cas9 enzyme from *Streptococcus pyogenes* (SpCas9) have successfully been engineered for genome editing and base editing in a wide range of organisms. As one example, base editors have been developed that convert Cas endonucleases into programmable nucleotide deaminases [1, 2, 3], thus facilitating the introduction of C-to-T mutations (by C-to-U deamination) or A-to-G mutations (by A-to-I deamination) without induction of a double-strand break [4, 5].

[0003] One drawback of current genome and base engineering tools (e.g., ZNFs, TALENS, and CRISPR/Cas9) is that they are limited with respect to the DNA sequences that can be targeted. For example, ZNF and TALENS are limited because each system requires the design of a specific DNA-binding portion, the amino acid sequence of which being a function of each individual target nucleotide sequence. CRISPR/Cas9 technologies are also limited. While Cas9 can be programmably targeted to virtually any target sequence by providing a suitable guide RNA, Cas9 strictly requires the presence of a protospacer-adjacent motif (PAM)-- which is typically the canonical nucleotide sequence 5′-NGG-3′ (*e.g.*, for SpCas9)--immediately adjacent to the 3′-end of the targeted nucleic acid sequence in order for the Cas9 to bind and act upon the target sequence. This requirement for a PAM sequence effectively limits the nucleotide sequences which can be efficiently targeted by Cas9.

[0004] Accordingly, there is a need for nucleic acid programmable DNA binding proteins, such as Cas9, that are capable of binding target nucleotide sequences that lack canonical PAMs(*e.g.*, 5′-NGG-3′ for SpCas9) in order to expand the scope and flexibility of genome and base editing.

SUMMARY OF THE INVENTION

[0005] The clustered regularly interspaced short palindromic repeat (CRISPR) system is a prokaryotic adaptive immune system that has been modified to enable robust genome and nucleobase engineering in a variety of organisms and cell lines. CRISPR-Cas (CRISPR-associated) systems are protein-RNA complexes that use an RNA molecule (sgRNA) as a guide to localize the complex to a target nucleic acid sequence via base-pairing. In the natural systems, a Cas protein then acts as an endonuclease to cleave the targeted DNA sequence. The target nucleic acid sequence must be both complementary to the sgRNA and also contain a "protospacer-adjacent motif" (PAM) at the 3′-end of the complementary region in order for the system to function. The requirement for a PAM sequence limits the use of Cas9 technology, especially for applications that require precise Cas9 positioning, such as base editing, which requires a PAM approximately 13-17 nucleotides from the target base and some forms of homology-directed repair, which are most efficient when DNA cleavage occurs ~ 10-20 base pairs away from a desired alteration. To address this limitation, researchers have harnessed natural CRISPR nucleases with different PAM requirements and engineered existing systems to accept variants of naturally recognized PAMs. Other natural CRISPR nucleases shown to function efficiently in mammalian cells include *Staphylococcus aureus* Cas9 (SaCas9), *Acidaminococcus sp.* Cpf1 (AsCpf1), *Lachnospiraceae* bacterium Cpf1, *Campylobacter jejuni* Cas9, *Streptococcus thermophilus* Cas9, and Neisseria meningitides Cas9. None of these mammalian cell-compatible CRISPR nucleases, however, offers a PAM that occurs as frequently as that of SpCas9.

[0006] Some aspects of the disclosure relate to novel Cas9 mutants that are capable of binding to target sequences that do not include a canonical PAM sequence (5′-NGG-3′, where N is any nucleotide) at the 3′-end. The disclosure also provides methods of generating and identifying novel

Cas9 variants, *e.g.,* using Phage Assisted Continuous Evolution (PACE) and/or Phage Assisted Non-Continuous Evolution (PANCE), that are capable of recognizing (*e.g.,* binding to) target sequences encompassing the a variety of PAM sequences . In particular, methods and compositions have been developed for targeting sequences that have an adenine (A) at the second nucleic acid position of the PAM (*e.g.,* 5′-NAN-3′). It should be appreciated that target sequences having PAMs that lack one or more guanines (Gs) are particularly difficult to target given the paucity of SpCas9 activity (*e.g.,* binding activity) on such sequences. One goal of the disclosure is to provide a repertoire of SpCas9 variants that could be selected from for use in genome and/or base editing applications that are specific for a target nucleic acid sequence (*e.g.*, DNA sequence) based on a particular PAM sequence. Such a catalogue/library of SpCas9 variants would be useful for expanding the scope of genome and base editing, so as not to be restricted by any particular PAM requirement.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] **Figures 1A-1C** show schematic representations of Phage Assisted Continuous Evolution (PACE) of Cas9 and results of SpCas9 vs xCas9 evolution. **Figure 1A**, PACE takes place in a fixed-volume "lagoon" that is continuously diluted with fresh host *E. coli* cells. Upon infection, each selection phage (SP) that encodes a Cas9 variant capable of binding the target PAM and protospacer on the accessory plasmid (AP) induces expression of gene III, resulting in infectious progeny phage that propagate the active Cas9 variant in subsequent host cells. **Figure 1B**, accessory plasmids representing each of 64 PAM sequences are used to select for Cas9 variants capable of binding to the PAM/protospacer sequences, where RNAP fused to the Cas9 variant induces express ion of gene III upon binding to the sequence having the specific PAM. **Figure 1C,** data (luciferase assay) for overnight phage propagation reveals on which PAMs SpCas9 and xCas9 have binding activity. xCas9 has a less strict PAM requirement as compared to SpCas9.

[0008] **Figures 2A-B** show a schematic representation of a Cas9 64 PAM Phage Assisted Non-Continuous Evolution (PANCE) and results of SpCas9 vs xCas9 PANCE evolution. **Figure 2A**, 96

well PANCE format allowed for simultaneous evolution of all 64 PAM sequences. PANCE is lower stringency than PACE as it is not continuous flow, thereby allowing for evolution from low activity. **Figure 2B**, data (luciferase assay) for PANCE evolution at passage 2 (P2), passage 12 (P12), and passage 16 (P16) for SpCas9 (wt) or xCas9 show an increase in the ability to bind additional PAM sequences.

[0009] **Figures 3A-B** show clones resulting from PANCE evolution experiments using SpCas9 (N3) after passage 12, including the activity for selected clones. **Figure 3A**, is a table listing individual clones and their mutations as compared to nuclease inactive SpCas9. The nomenclature of each clone indicates the PAM on which the clone was evolved. For example, clones CAA-2, CAA-3, and CAA-4 were evolved using a 5′-CAA-3′-PAM sequence. **Figure 3B**, shows activity for clones SpCas9, CAA-3, GAT-2, ATG-2, ATG-3, and AGC-3, using a luciferase assay. Clones were obtained from PANCE evolution experiments using SpCas9 (N3) after passage 12.

[0010] **Figures 4A-B** show clones resulting from PANCE evolution experiments using SpCas9 (N3) after passage 19, including the activity for selected clones. **Figure 4A**, is a table listing individual clones and their mutations as compared to nuclease inactive SpCas9. The nomenclature of each clone indicates the PAM on which the clone was evolved. For example, clones ACG-1, ACG-2, ACG-3, and ACG-4 were evolved using a 5′-ACG-3′-PAM sequence. **Figure 4B**, shows activity for clones SpCas9, N3.19.CAA1, N3.19.CAA2, N3.19.GAA1, N3.19.GAA2, N3.19.GAC5, N3.19.GAT1, N3.19.GAT3, N3.19.ACG1, N3.19.ACG3, N3.19.ACG6, N3.19.ATG3, and N3.19.ATG6 using a luciferase assay. Clones were obtained from PANCE evolution experiments using SpCas9 (N3) after passage 19.

[0011] **Figures 5A-B** show clones resulting from PANCE evolution experiments using xCas9 3.7 (N4) after passage 12, including the activity for selected clones. **Figure 5A**, is a table listing individual clones and their mutations as compared to xCas9 3.7. The table indicates the PAM on which each of the clones was evolved. For example, clones N4.12.10 TAT1, N4.12.10 TAT2, and

N4.12.10 TAT3 were evolved using a 5′-TAT-3′-PAM sequence. **Figure 5B**, shows activity for clones xCas9 (xCas9 3.7), TAT-1, TAT-3, GTA-1, GTA-3, and CAC-2 using a luciferase assay. Clones were obtained from PANCE evolution experiments using xCas9 3.9 (N4) after passage 12.

[0012] **Figures 6A-B** show clones resulting from PANCE evolution experiments using xCas9 3.7 (N4) after passage 19, including the activity for selected clones. **Figure 6A**, is a table listing individual clones and their mutations as compared to xCas9 3.7. The table indicates the PAM on which each of the clones was evolved. For example, clones N4.19.AAA1, N4.19.AAA2, N4.19.AAA4, and N4.19.AAA7 were evolved using a 5′-AAA-3′-PAM sequence. **Figure 6B**, shows activity for N4.19.AAA1, N4.19.TAA2, N4.19.TAA5, N4.19.TAT5, N4.19.CAC5, N4.19.CAC6, N4.19.GTA2, N4.19.GTA7, N4.19.GCC2, N4.19.GCC5, and N4.19.GCC8 using a luciferase assay. Clones were obtained from PANCE evolution experiments using xCas9 3.9 (N4) after passage 19.

[0013] **Figure 7** shows the results of mammalian cell editing using cytidine base editor BE3 having various evolved Cas9 clones (top). Indel formation for each of the clones as nuclease active Cas9s is also provided (bottom).

[0014] **Figure 8** shows activity data (luciferase assay) for PANCE evolution experiments after passage 2 (N6.2), passage 12 (N6.12) and passage 16 (N6.16) using N4.12.TAT1 as the starting clone (N6). Increased shading indicates increased activity as described in Figure 1C.

[0015] **Figures 9A-B** show the mutations of TAT1 well as activity data (luciferase assay) on all 64 possible PAM sequences. **Figure 9A** provides the individual mutations of N4.12.TAT1 (TAT1) as compared to SpCas9. **Figure 9B** shows activity of TAT1 on all 64 possible PAM sequences. Increased shading indicates increased activity as described in Figure 1C.

[0016] **Figure 10** shows clones of resulting from PANCE evolution experiments using N4.12.TAT1 (N6) after passage 12. The individual mutations in clones N6.12.6, N6.12.7, N6.12.25, and N6.12.28, are shown as compared to TAT1.

[0017] **Figure 11** shows clones of resulting from PANCE evolution experiments using N4.12.TAT1 (N6) after passage 18. The individual mutations for each of the listed clones (*e.g.*, N6.18.1-1, N6, 18.1-2, etc.), are shown as compared to TAT1.

[0018] **Figure 12** shows activity for N6.18.17-2, N6.18.18-2, N6.18.18-3, N6.18.28-2, N6.18.33-3, N6.18.39-1, N6.18.39-3, N6.18.39-4, N6.18.40-2, N6.18.40-3, N6.18.44-1, SP047a, and SpCas9. using a luciferase assay. Clones were obtained from PANCE evolution experiments using N4.12.TAT1 (N6) after passage 18 (See Figure 11).

[0019] **Figures 13A-B** show a split-intein PACE configuration to allow evolution of two separate activities of interest. **Figure 13A** shows that the bacteriophage gIII gene that produces the pIII protein is split into N-terminal (g3N) and C-terminal (g3C) fragments in two separate accessory plasmids (AP1 and AP2). AP1 and AP2 have the same PAM, but a different protospacer (it is not required that they have the same PAM, i.e., both the PAM and protospacer could be changed). **Figure 13B** shows the workflow for using a split-intein PACE configuration of the gIII gene.

[0020] **Figures 14A-C** show the evolution and activity of SpCas9 resulting from PACE experiments using two separate protospacers and split-intein fusion (two allow evolution on two protospacers) as in Figures 13A-B. **Figure 14A** shows clones resulting from PACE evolution experiments using two protospacers with SpCas9 after passage 4 (P4). **Figure 14B** shows the ability of the P4 SpCas9 variants incorporated into a BE4max base-editor to support conversion of C to T in CAG, CAT, GAT, CAA, GAA, CGT, or GGG PAMs. **Figure 14C** shows the ability of the L2-72-4 SpCas9 P4 clone to form insertions or deletions in CAA1, CAA2, AAA1, AAA2, TAA1, TAA2, CAG1, CAG2, GAT1, GAT2, TAT1, TAT2, CAT, GAA1, GAA2, CGT, and GGG PAMs.

[0021] **Figures 15A-B** show a split-intein PACE configuration (whereby Cas9 is divided into two parts to limit Cas9 concentration) to allow evolution of Cas9 proteins of interest. **Figure 15A** shows that increasing the SpCas9 concentration increases cleavage of alternative (NAG) PAMs (as reported in Karvelis, T., Gasiunas, G., Young, J., Bigelyte, G., Silanskas, A., Cigan, M., and Siksnys, V. (2015).

Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. Genome Biol. 16, 253). **Figure 15B** shows that the amount of Cas9 protein may be limited in PACE by splitting the inactive Cas9 protein (dCas9) into an N-terminal fragment (dCas9 (1-573)) and a C-terminal fragment (dCas9 (573-end)) and producing the N-terminal fragment from a low-copy number plasmid with a weak promoter (rpoZ).

[0022] **Figure 16** shows clones resulting from PACE evolution when a split-intein Cas9 protein with the P4.2.72.4. mutations Experiment P10). The individual mutations for each of the listed clones (*e.g.*, L5.144.2, L5.144.6, etc.), are shown as compared to spCas9 and spCas9 with the P4.2.72.4. mutations.

[0023] **Figure 17** shows the ability of the P10 SpCas9 variants from Figure 16 incorporated into a BE4max base-editor to support conversion of C to T in CAG, GAT, TAT, CAT, GAA, CAA-1, or CAA-2 PAMs.

[0024] **Figure 18** shows the ability of two P10 SpCas9 variants (P10.5.144.2 and P10.6.144.2) to form insertions or deletions in CAA1, CAA2, AAA1, AAA2, TAA1, TAA2, CAG1, CAG2, GAT1, GAT2, TAT1, TAT2, CAT, GAA1, GAA2, CGT, and GGG PAMs compared to a P4 variant (L2-72-4), SpCas9, and xCas9.

[0025] **Figures 19A-C** show characterization of a P10 SpCas9 variant with PAM depletion in *E. coli*. **Figure 19A** shows a workflow for PAM depletion in *E. coli*, wherein *E. coli* containing a Cas9 variant (e.g., P10) are transformed with a library of negative selection plasmids (e.g., pUC ampR with HEK3 protospacer followed by NNNN). See Kleinstiver et al., Engineered CRISPR-Cas9 nucleases with altered PAM specificities, *Natur,* 523: 481-485. The transformed cells are recovered and Cas9 expression is induced for 1-4 hours. The cells are then plated on carbenicillin media. The plates are then scraped and surviving colonies are sequenced for mutations. Colonies that survive and are sequenced contain PAMs that the P10 Cas9 variant protein could not cut. **Figure 19B** shows the frequency of PAM sequences present in surviving colonies, wherein more shaded PAM sequences

occur more frequently (left), and the activity of P10 Cas9 variant protein on the PAM sequences in a luciferase assay (right). **Figure 19C** the activity of the P10 SpCas9 variants were characterized by PAM depletion incorporated into a BE4max base-editor to support conversion of C to T in CAG, CAT, GAT, CAA, GAA, CGT, or GGG PAMs

[0026] **Figure 20** shows a characterization of the P10 SpCas9 variant protein following PAM depletion as in Figures 19A-19C. The P10 SpCas9 variant protein (left) and xCas9 variant proteins (middle) show preference for the fourth nucleotide in the PAM, wherein C is the most preferred and G is the least preferred. The spCas9 protein (right) does not show this preference. Higher Cas9 protein activity is denoted by darker shading.

[0027] **Figure 21** shows clones resulting from split-intein PACE evolution of Cas9 with the P4.2.72.4 mutations Experiment P11) with a AAA PAM. The individual mutations for each of the listed clones (*e.g.*, P11.1.139-2, P11.1.139-4, etc.), are shown as compared to spCas9 with the P4.2.72.4. mutations.

[0028] **Figure 22** shows the ability of the P11 SpCas9 variants from Figure 16 incorporated into a BE3 base-editor to support conversion of C to T in CAG, GAT, CAT, GAA, AAA-1, AA1-2, CAA-1, CAA-2, or GGG PAMs.

[0029] **Figure 23** shows the ability of two P11 SpCas9 variants (P11-SacB-1 and P11-SacB-2) to form insertions or deletions in CAA1, CAA2, AAA1, AAA2, TAA1, TAA2, CAG1, CAG2, GAT1, GAT2, TAT1, TAT2, CAT, GAA1, GAA2, CGT, and GGG PAMs compared to a P4 variant (L2-72-4), SpCas9, and xCas9.

[0030] **Figures 24A-B** show clones resulting from split-intein PACE evolution of Cas9 with P12 mutations on AAT (**FIG. 24A**) or TAT (**FIG. 24B**) PAMs. The individual mutations for each of the listed clones (*e.g.*, P12.3.b9-2, P12.3.b10-2 etc.), are shown as compared to spCas9 protein.

[0031] **Figures 25A-B** show the ability of the P12 SpCas9 variants from Figures 24A-B incorporated into a BE3 base-editor to support conversion of C to T at sites s893, s1073, s1081,

s1140, b3, e1, e2, f1, f2, s33, s34, s35, s36, s37, s38, s39, s40, s41, s43, s44, s45, or s46. Darker shading indicates a higher % of C to T editing (**FIG. 25A**). **Figure 25B** shows the average C to T editing on NATA, NATT, NATC, or NATG PAMs. pSM060ax is clone P12.3.b9-8 and pSM060ay is clone P12.3.b10-6.

[0032] **Figures 26A-B** show the ability of two P12 SpCas9 variants (P12.3.b9-8 and P12.3.b10-6) to cleave DNA in bacterial PAM depletion in AAA, AAC, AAT, AAG, CAA, CAC, CAT, CAG, TAA, TAC, TAT, TAG, GAA, GAT, GAG, AGA, AGC, AGT, AGG, CGA, CGC, CGT, CGG, TGA, TGC, TGT, TGG, GGA, GGC, GGT, or GGG PAMs. PPDV is the PAM frequency after Cas9 cutting/frequency of input library, wherein lower numbers signify more active Cas9 proteins.

[0033] **Figures 27A-B** show a split-intein PACE configuration to allow evolution of Cas9 proteins of interest with 2 protospacers. **Figure 27A** shows evolution of a split-intein Cas9 using selection on 2 protospacers. A second gene (gVI) is removed from the phage and is used as a selection marker on AP2. AP1 and AP2 have the same PAM, but different protospacers and a different nucleotide immediately 3' of the PAM. **Figure 27B** shows clones resulting from split-intein PACE evolution of Cas9 as in **Figure 27A**. The individual mutations for each of the listed clones (*e.g.*, L2-120-1, L2-120-2, etc.), are shown as compared to spCas9 protein.

[0034] **Figure 28** shows survival-based selection for isolating nuclease-active Cas9 variant proteins. In this selection, cutting identifies nuclease-active PACE variants. SacB is lethal in the presence of sucrose unless it is cut by Cas9, sfGFP loses fluorescence if Cas9 cutting occurs, and kanR confers survival on kanamycin medium if no cutting occurs.

[0035] **Figures 29A-B** show nuclease-active TAT variants that were identified by SacB selection as in **Figure 28**. The original spCas9 TAT variant was isolated from PANCE evolution on a TAT PAM (N4.TAT.1), but had no nuclease activity. This N4.TAT.1 (TAT1) Cas9 variant was subcloned from the pool of N4.TAT SP (H840-onward) into a Cas9 plasmid and selected for variants that could cut a SacB selection plasmid with a TAT PAM after a 4 hour induction. **Figure 29A** shows clones resulting from

SacB selection of nuclease-inactive TAT. The individual mutations for each of the listed clones (*e.g.*, SacB-TAT-1, SacB-TAT-2), are shown as compared to SpCas9 and TAT SpCas9 variant proteins. **Figure 29B** shows the location of mutations in the TAT SpCas9 variant proteins.

[0036] **Figures 30A-B** show the activity of the TAT SpCas9 variant proteins identified in **Figure 29A**. **Figure 30A** shows the ability of the nuclease-active TAT SpCas9 variants (SacB-TAT1 and SacB-TAT2) incorporated into a BE4max base-editor to support conversion of C to T in CAG, GAT, TAT, CAT, GAA-1, GAA-2, CAA-1, CAA-2, or GGG PAMs. **Figure 30B** shows ability of the SacB-TAT1 and SacB-TAT2 variants to form PAM depletion in CAA1, CAA2, AAA1, AAA2, TAA1, TAA2, CAG1, CAG2, GAT1, GAT2, TAT1, TAT2, CAT, GAA1, GAA2, CGT, or GGG PAMs.

[0037] **Figure 31** shows the ability of the SacB-TAT-1 SpCas9 protein variant to form insertions or deletions in AAA, AAC, AAT, AAG, CAA, CAC, CAT, CAG, TAA, TAC, TAT, TAG, GAA, GAT, GAG, AGA, AGC, AGT, AGG, CGA, CGC, CGT, CGG, TGA, TGC, TGT, TGG, GGA, GGC, GGT, or GGG PAMs. PPDV is the PAM frequency after Cas9 cutting/frequency of input library, wherein lower numbers signify more active Cas9 proteins.

[0038] **Figure 32** shows the location of frequently mutagenized residues by PAM selection. Positions commonly mutated in SpCas9 variants obtained when evolving on NAN PAMs include: D1135, E1219, D1332.

[0039] **Figures 33A-33D** show C to T base editing with evolved variants on PAMs. C to T base editing with SpCas9 variants were incorporated into Be4MAX architecture in HEK293T cells. **Figure 33A** shows C to T base editing with NAA PAMs. **Figure 33B** shows C to T base editing with NAC PAMs. **Figure 33C** shows C to T base editing with NAT PAMs. **Figure 33D** shows C to T base editing with NAG PAMs. Each bar represents the average of 3 independent experiments, and the error bars represents the standard deviation. The "es" SpCas9 variant protein works best on NARH PAMs, with some activity on NARG and NGN PAMS, the "fn" SpCas9 variant protein works best on NRCH

PAMs, with some activity on NRCG and NGN PAMs, and the "ax" SpCas9 variant protein works best on NRTH PAMs, with some activity on NRTG and NGN PAMs.

[0040] **Figures 34A-34B** show C to T base editing with evolved SpCas9 variants on PAMs. C to T base editing with SpCas9 variants were incorporated into BE4MAX architecture in HEK293T cells. **Figure 34A** shows C to T base editing on NAA, NAC, and NAT PAMs. **Figures 34B** shows C to T base editing on NAAH, NACH, and NATH PAMs, where H is any base except for G. Each bar represents the average of 3 independent experiments, and the error bars represents the standard deviation.

[0041] **Figures 35A-35C** show A to G base editing with evolved SpCas9 variants on PAMs. A to G base editing with SpCas9 variants incorporated into ABEMAX architecture in HEK293T cells. **Figure 35A** shows A to G base editing on NAA/NGA PAMs with es variant SpCas9. **Figure 35B** shows A to G base editing on NAC/NGC PAMs with fn variant SpCas9. **Figure 35C** shows A to G base editing on NAG/NGG PAMs with es and fn variant SpCas9 proteins. Each bar represents the average of 2 independent experiments, and the error bars represent the standard deviation.

[0042] **Figure 36** show phage-assisted non-continuous evolution (PANCE) of SpCas9 binding activity on non-G PAMs. (A) Original selection scheme for Cas9 DNA binding. w-dSpCas9 expressed by DgIII selection phage (SP) binds to a designated protospacer/PAM sequence upstream of gIII on an accessory plasmid (AP) in host *E. coli* cells. Host cells and infecting SP are continuously mutagenized by a mutagenesis plasmid (MP). (B) Fold propagation of SP expressing w-dSpCas9 or w-dxCas9 on APs encoding each of all 64 NNN PAM sequences upstream of gIII. (C) Schematic overview of PANCE workflow. Host cells containing an AP and MP are grown to log phase in a deep well plate or tube before being infected with SP. Mutagenesis is induced and SP are allowed to propagate for 6-18 hours before cells are pelleted and the SP-containing supernatant is collected. The SP pool is then used to infect host cells in the next iteration of PANCE. (D) Consensus mutations

arising from evolution of w-dSpCas9 (N1) or w-dxCas9 (N2) on NAA (red), NAT (blue), or NAC (green) PAM sequences.

[0043] **Figures 37A-37E** shows multiple new PACE schemes utilizing a split-intein Cas9 and/or two protospacers. **Figure 37A** shows new PACE schemes to limit the concentration of spCas9 protein and/or increase the number of Cas9 binding sites. **Figure 37B** shows SpCas9 individual NAA mutations for each of the listed clones (*e.g.,* N3.GAA-3, N3.GAA-4, etc.), are shown as compared to SpCas9 protein. **Figure 37C** shows a timecourse of the NAA variants from **Figure 37B** through evolution. **Figure 37D** shows SpCas9 individual NAC mutations for each of the listed clones (*e.g.,* N4.CAC-1, N4.CAC-5, etc.), are shown as compared to SpCas9 protein. Also shown is D1135N, R1114G, V1139A, E1219V, Q1221H, R1320V, and R1333K mapped to the SpCas9 crystal structure 4un3. **Figure 37E** shows SpCas9 individual NAT mutations for each of the listed clones (*e.g.,* SacB.N4.TAT-1, SacB.N4-TAT-3, etc.), are shown as compared to SpCas9 protein. Also shown is D1135N, R1114G, E1219V, H1349R, S1338T, R1335Q, and D1332N mapped to the SpCas9 crystal structure 4un3 (left, lower structure). The lower right structure also shows D1135N, R1114G, E1219V, G1218S, Q1221H, P1321S, R1335, and D1332G mapped to the SpCas9 crystal structure 4un3.

[0044] **Figures 38A-38D** show characterization of evolved variants and SpCas9-NG through bacterial PAM depletion and mammalian cell indel formation. **Figure 38A** shows bacterial PAM depletion of SpCas9-NRRH, -NRTH, -NRCH, and SpCas9-NG using a bacterial NNNN PAM library. The inverse of the depletion score was used to generate enrichment scores of activity on each NNNN PAM, which were then used to create sequence logos (WebLogo3.0). **Figure 38B** shows indel formation in HEK293T cells across 64 endogenous mammalian sites containing NANN PAMs for SpCas9-NRRH, -NRTH, -NRCH, and SpCas9-NG. Mean and SE of three independent biological replicates are shown. **Figure 38C** provides a summary of indel formation efficiencies in HEK293T cells across 48 endogenous mammalian sites containing NANH (H=non-G) PAMs for SpCas9-NRRH,

-NRTH, -NRCH, and SpCas9-NG. Mean and standard deviation (SD) of all individual values of three independent biological replicates are plotted. **Figure 38D** shows DNA targeting specificity of SpCas9, xCas9, and evolved variants SpCas9-NRRH, -NRTH-, and NRCH as determined by % on-target reads resulting from GUIDE-seq analysis using HEK target site 4 in U2OS cells.

[0045] **Figure 39A-39E** show mammalian C to T and A to G base editing activity of evolved variants and SpCas9-NG. **Figure 39A** shows cytosine base editing in HEK293T cells across 64 endogenous mammalian sites containing NANN PAMs for BE4-NRRH, BE4-NRTH, BE4-NRCH, and BE4-NG. Mean and SE of three independent biological replicates are shown. **Figure 39B** shows a summary of cytosine base editing in HEK293T cells across 48 endogenous mammalian sites containing NANH (H=non-G) PAMs for BE4-NRRH, BE4-NRTH, BE4-NRCH, and BE4-NG. Mean and SE of three independent biological replicates are shown. **Figure 39C** shows adenine base editing in HEK293T cells across 27 endogenous mammalian sites containing NANN PAMs for ABE-NRRH, ABE-NRTH, ABE-NRCH, and ABE-NG. Mean and SE of three independent biological replicates are shown. **Figure 39D** shows the fraction of pathogenic SNPs in the ClinVar Database that could in principle be corrected by a C•G to T•A (left) or A•T to G•C (right) base conversion using NR PAMs. **Figure 39E** shows the number of possible sgRNAs capable of targeting pathogenic SNPs in the ClinVar Database using NR, NG, or NGG PAMs.

[0046] **Figures 40A-40G** shows a characterization of PAM preferences using a genomically integrated human cell base editing target sequence library. **Figure 40A** is a schematic overview of a mammalian cell base editing library experiment. A library of matched sgRNA/protospacer target sites spanning all NNNN PAMs is stably genomically integrated in HEK293T cells. Library cells are then transfected with and selected for genomic integration of plasmids encoding BE4 variants. After antibiotic selection, cells are lysed and the integration of plasmids encoding BE4 variants. After antibiotic selection, cells are lysed and the integrated sgRNA/protospacer site is PCR amplified for HTS analysis. **Figure 40B** provides a heat map of base editing activity on the NNNN PAM

library in HEK293T cells, with positions 2, 3, and 4 of the PAM defined. For each construct, the mean editing across all sites containing the designated PAM over two independent biological replicates, internally normalized against the highest editing value for each construct, is shown. **Figure 40C-E** shows the average base editing activity on the NNNN PAM library in HEK293T cells by BE4, BE4-NRRH, BE4-NRTH, and BE4-NRCH, with PAM positions 2 (C), position 3 (D), or position 4 (E) fixed. Mean and SE for individual editing values (averaged across two independent biological replicates) at all relevant library sequences are shown. **Figure 40F-40G** show the effect of sgRNA length and 5'G mismatches on the base editing efficiency of profiled SpCas9 variants. The percentage decrease of editing efficiency from using a 21 nt sgRNA with either a mached (F) or mismatched (G) 5'G compared to using a matched 20 nt sgRNA is shown for BE4, BE4-NRRH, BE4-NRCH, BE4-NRTH, and BE4-NG on all library sequences containing NAN, NRN, NGN, or NGG PAMs. The mean and SE are plotted.

[0047] **Figure 41A-41C** shows evolved SpCas9 variants allow correction of pathogenic SNPs using non-G PAMs. **Figure 41A** provides an overview of adenine base editing strategy for correcting the sickle hemoglobin (*HbS*) SNP. In HbS, the Glu (GAG codon) at position 6 of normal b-globin (*HBB*) is mutated to a Val (GTG codon). Targeting this SNP with A·T to G·C base editing on the reverse strand enables a Val to Ala (GTG to GCG) base conversion, leading to the Makassar b-globin variant (HbG) which produces phenotypically normal b-globin. **Figure 41B** shows A·T to G·C base editing in HEK293T cells engineered with the *HbS* mutation using a CACC PAM by ABE-NRRH, ABE-NRTH, ABE-NRCH, and ABE-NG. This PAM places the target A at position 7, and an off-target A, which leads to a silent pro (CCT) to pro (CCC) mutation, at position 9. Mean and SE of three independent biological replicates are shown. **Figure 41C** shows A·T to G·C base editing in HEK293T cells engineered with the *HbS* mutation using a CATG PAM by ABE-NRRH, ABE-NRTH, ABE-NRCH, and ABE-NG. This PAM places the target A at position 4, and an off-target A, which

leads to a silent pro (CCT) to pro (CCC) mutation, at position 6. Mean and SE of three independent biological replicates are shown.

[0048] **Figure 42** provides a table of NRNN PAM targeting potential by SpCas9 and SaCs9 variants described herein. The variants SpCas9-NRRH, SpCas9-NRTH, and SpCas9-NRCH are disclosed and discussed herein.

[0049] **Figure 43A-43F** depicts additional details of Cas9:DNA binding PACE and Cas9 nuclease selections. **Figure 43A** shows dual AP selection where ω-dSpCas9 binds two distinct protospacer/PAM sequences to drive either half of split-intein pIII. **Figure 43B** shows split-intein Cas9 limits total Cas9 concentration in host cells, thus avoiding saturation of protospacer/PAM binding sites. Residues 574-1368 of Cas9 fused to NpuC is expressed by ΔgIII SP and ω-dSpCas9(1-573) fused to NpuN is encoded on a low copy complimentary plasmid (CP) in host cells. **Figure 43C** shows a combination of the selection principles from (A) and (B) through use of gVI as an additional PACE-compatible selection marker for phage propagation and ΔgIIIΔgVI SP. **Figure 43D** shows overnight propagation assay of selection phage (SP) encoding dSpCas9C on host cells containing a complimentary plasmid (CP) providing either ω-dSpCas9$_N$ or ω-dSpCas9$_{N\text{-}mut}$ and an AP encoding either a AAA or CAA PAM. **Figure 43E and 43F** show a scheme of survival based selection for Cas9 nuclease activity. Cells containing a high-copy selection plasmid encoding a protospacer/ PAM sequence, sfGFP, and the conditionally lethal protein SacB are transformed with a library of nuclease-active Cas9s encoded on a low-copy plasmid that also includes the matching sgRNA. Binding and cleavage of the designated PAM/protospacer by Cas9 leads to destruction of the selection plasmid, resulting in loss of both sfGFP and SacB expression, allowing cells to survive on sucrose-containing media.

[0050] **Figure 44A-44C** show the effects of mutations on PAM recognition by SpCas9 variants. **Figure 44A** shows the addition of the Y1131C mutation, which was enriched in the later phases of the NAT evolution trajectory, inactivates BE3-NRTH in HEK293T cells. Mean and SE of three

independent biological replicates are shown. **Figure 44B** shows the N-terminal mutations of SpCas9-NRRH, -NRCH, and -NRTH mapped to the SpCas9 crystal structure (4UN3). **Figure 44C** shows CBE activity of BE3-NRRH, BE3-NRTH, and BE3-NRCH with and without the N-terminal mutations shown in (B) in HEK293T cells. Mean and SE of three independent biological replicates are shown.

[0051] **Figure 45A-45D** is a characterization of SpCas9, xCas9, and evolved variants (SpCa9-NRTH, SpCas9-NRCH, and SpCas9-NRRH) in bacterial PAM depletion and mammalian indel formation experiments. **Figure 45A** shows bacterial PAM depletion of SpCas9-NRRH, -NRCH, -NRTH, and SpCas9-NG on a bacterial NNNN PAM library with 1 h, 3 h, and overnight Cas9 induction. The inverse of the depletion score was used to generate enrichment scores of activity on each NNNN PAM, which were then used to create sequence logos (WebLogo3.0). **Figure 45B** shows indel formation in HEK293T cells across endogenous mammalian sites containing NANN PAMs for xCas9, SpCas9-NRRH, -NRTH, -NRCH, and SpCas9-NG. Mean and SE of three independent biological replicates are shown. **Figure 45C** shows indel formation in HEK293T cells across endogenous mammalian sites containing NGNN PAMs for SpCas9-NRRH, -NRTH, -NRCH, SpCas9-NG, and SpCas9. Mean and SE of three independent biological replicates are shown. **Figure 45D** shows GUIDE-seq analysis of SpCas9, xCas9, and evolved variants SpCas9-NRRH, -NRTH, and -NRCH targeting HEK site 4 in U2OS cells. GUIDE-seq on-target (indicated by the asterisk) and off-target reads that are greater than or equal to 1% total reads are shown.

[0052] **Figure 46A-46C** shows the characterization of SpCas9 (BE4), SpCas9-NG (BE4-NG), and evolved CBE and ABE variants in mammalian base editing experiments. **Figure 46A** shows CBE in HEK293T cells across endogenous mammalian sites containing NGNN PAMs for BE4-NRRH, BE4-NRTH, BE4-NRCH, BE4-NG, and BE4. Mean and SE of three independent biological replicates are shown. **Figure 46B** shows ABE in HEK293T cells across endogenous mammalian sites containing NGNN PAMs for ABE-NRRH, ABE-NRTH, ABE-NRCH, and ABE-NG. Mean and SE of three

independent biological replicates are shown. For target sites with NGA, NGC, and NGT PAMs, only ABE-NRRH, ABE-NRTH, and ABE-NRCH are shown, respectively, in addition to SpCas9-NG. **Figure 46C** shows the fraction of pathogenic SNPs in the ClinVar Database with either a single targetable base within the window or multiple targetable bases that could in principle be corrected by a C•G to T•A (top left) or A•T to G•C (top right) base conversion using NR PAMs or C•G to T•A (bottom left) or A•T to G•C (bottom right) base conversion using NG PAMs.

[0053] **Figure 47A-47D** shows the characterization of PAM preferences of BE4, BE4-NRRH, BE4-NRCH, and BE4-NG using a genomically integrated human cell base editing target sequence library **Figure 47A** shows the distribution of the number of target sites per PAM within the integrated sgRNA library. **Figure 47B** shows the PAM preferences for BE4, BE4-NRRH, BE4-NRTH, and BE4-NRCH as determined by base editing on the target sequence library integrated in HEK293T cells. Sequence logos for each construct were created from the CBE activity on each NNNN PAM contained in the library (WebLogo3.0). **Figure 47C** Average base editing activity on the NNNN PAM library in HEK293T cells by BE4, BE4-NRRH, BE4-NRTH, and BE4-NRCH, with PAM position 1 fixed. Mean and SE for individual editing values (averaged across two independent biological replicates) at all relevant library sequences are shown. **Figure 47C-47D** shows effect of sgRNA length and 5'G mismatch on base editing efficiency of profiled SpCas9 variants. Average base editing on the NNNN PAM library in HEK293T cells by BE4, BE4-NRRH, BE4-NRTH, and BE4-NRCH is grouped by sites containing a 20-nt sgRNA with a 5'G matched to the target sequence, a 21-nt sgRNA with a 5'G matched to the target sequence, or a 21-nt sgRNA with a mismatched 5' nucleotide. Average editing activity of constructs on NGN (**Figure 47D**), NAN (**Figure 47E**), and NGG (**Figure 47F**) PAMs are shown. Mean and SE for individual editing values (averaged across two independent biological replicates) at all relevant library sequences are shown. ns, not significant; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$ (Student's t test).

[0054] **Figure 48A-48C** shows high-throughput sequencing analysis of sickle cell locus editing by SpCas9 variant-derived ABEs. **Figure 48A** shows Crispresso2 output showing the *HbS* mutation in a engineered HEK293T cell line. HEK293T cells were treated with nickase-SpCas9, sgRNA (binding shown in grey), and ssODN containing the point mutation. After two rounds of transfection, sorting, and growth, the cell line sequenced above was isolated and identified to have 100% conversion to the sickle cell anemia allele. **Figure 48B** shows Crispresso2 output showing ABE activity of ABE-NRRH, ABE-NRTH, ABE-NRCH, and ABE-NG in *HbS* engineered HEK293T cells using a sgRNA (gray bar) targeting a CATG PAM. **Figure 48C** shows Crispresso2 output showing ABE activity of ABE-NRRH, ABE-NRTH, ABE-NRCH, and ABE-NG in *HbS* engineered HEK293T cells using a sgRNA (gray bar) targeting a CACC PAM.

DEFINITIONS

[0055] As used herein and in the claims, the singular forms "a," "an," and "the" include the singular and the plural reference unless the context clearly indicates otherwise. Thus, for example, a reference to "an agent" includes a single agent and a plurality of such agents.

[0056] The term "base editor (BE)," or "nucleobase editor (NBE)," as used herein, refers to an agent comprising a polypeptide that is capable of making a modification to a base (*e.g.,* A, T, C, G, or U) within a nucleic acid sequence (*e.g.,* DNA or RNA). In some embodiments, the base editor is capable of deaminating a base within a nucleic acid. In some embodiments, the base editor is capable of deaminating a base within a DNA molecule. In some embodiments, the base editor is capable of deaminating a cytosine (C) in DNA. In some embodiments, the base editor is a fusion protein comprising a nucleic acid programmable DNA binding protein fused to a nucleic acid editing domain. In some embodiments, the base editor is a fusion protein comprising a nucleic acid programmable DNA binding protein (napDNAbp) fused to a cytidine deaminase domain. In some embodiments, the base editor comprises a Cas9 domain (*e.g.,* an evolved Cas9 domain), or an evolved version of a CasX, CasY, Cpf1, C2c1, C2c2, C2c3, or Argonaute protein that comprises one or more mutations

homologous to the mutations provided herein fused to a cytidine deaminase. In some embodiments, the base editor comprises a Cas9 nickase (Cas9n) fused to an cytidine deaminase domain. In some embodiments, the base editor comprises a nuclease-inactive Cas9 (dCas9) fused to a cytidine deaminase domain. In some embodiments, the base editor includes an inhibitor of base excision repair, for example, a UGI domain or a dISN domain.

[0057] In some embodiments, the base editor is capable of deaminating an adenosine (A) in DNA. In some embodiments, the base editor is a fusion protein comprising a nucleic acid programmable DNA binding protein fused to a nucleic acid editing domain. In some embodiments, the base editor is a fusion protein comprising a nucleic acid programmable DNA binding protein (napDNAbp) fused to an adenosine deaminase domain. In some embodiments, the base editor is a fusion protein comprising a nucleic acid programmable DNA binding protein (napDNAbp) fused to one or more adenosine deaminase domains. In some embodiments, the base editor is a fusion protein comprising a nucleic acid programmable DNA binding protein (napDNAbp) fused to two adenosine deaminase domains. In some embodiments, the base editor comprises a Cas9 (*e.g.,* an evolvedCas9 domain), or an evolved version of a CasX, CasY, Cpf1, C2c1, C2c2, C2c3, or Argonaute protein that comprises one or more mutations homologous to the mutations provided herein fused to an adenosine deaminase domain. In some embodiments, the base editor comprises a Cas9 nickase (Cas9n) fused to an adenosine deaminase domain. In some embodiments, the base editor comprises a Cas9 nickase (Cas9n) fused to two adenosine deaminase domains. In some embodiments, the base editor comprises a nuclease-inactive Cas9 (dCas9) fused to an adenosine deaminase domain. In some embodiments, the base editor comprises a nuclease-inactive Cas9 (dCas9) fused to two adenosine deaminase domains. In some embodiments, the base editor is fused to an inhibitor of base excision repair, for example, a UGI domain, or a dISN domain.

[0058] The term "nucleic acid programmable DNA binding protein" or "napDNAbp" refers to a protein that associates with a nucleic acid (*e.g.,* DNA or RNA), such as a guide nucleic acid (*e.g.,*

gRNA), that guides the napDNAbp to a specific nucleic acid sequence, for example, by hybridizing to the target nucleic acid sequence. For example, a Cas9 domain can associate with a guide RNA that guides the Cas9 domain to a specific DNA sequence that has complementary to the guide RNA. In some embodiments, the napDNAbp is a class 2 microbial CRISPR-Cas effector. In some embodiments, the napDNAbp is a Cas9 domain, for example, a nuclease active Cas9, a Cas9 nickase (Cas9n), or a nuclease inactive Cas9 (dCas9). Examples of nucleic acid programmable DNA binding proteins include, without limitation, Cas9 (*e.g.,* an evolved Cas9 domain), or an evolved version of a CasX, CasY, Cpf1, C2c1, C2c2, C2c3, or Argonaute protein that comprises one or more mutations homologous to the mutations provided herein. It should be appreciated, however, that nucleic acid programmable DNA binding proteins also include nucleic acid programmable proteins that bind RNA. For example, the napDNAbp may be associated with a nucleic acid that guides the napDNAbp to an RNA. Other nucleic acid programmable DNA binding proteins are also within the scope of this disclosure, though they may not be specifically described in this Application.

[0059] As used herein, the term "circular permutant" refers to a protein or polypeptide (e.g., a Cas9) comprising a circular permutation, which is change in the protein's structural configuration involving a change in order of amino acids appearing in the protein's amino acid sequence. In other words, circular permutants are proteins that have altered N- and C-termini as compared to a wild-type counterpart, e.g., the wild-type C-terminal half of a protein becomes the new N-terminal half. Circular permutation (or CP) is essentially the topological rearrangement of a protein's primary sequence, connecting its N- and C-terminus, often with a peptide linker, while concurrently splitting its sequence at a different position to create new, adjacent N- and C-termini. The result is a protein structure with different connectivity, but which oftern can have the same overall similar three-dimensional (3D) shape, and possibly include improved or altered characteristics, including, reduced proteolytic susceptibility, improved catalytic activity, altered substrate or ligand binding, and/or improved thermostability. Circular permutant proteins can occur in nature (e.g., concanavalin A and

lectin). In addition, circular permutation can occur as a result of posttranslational modifications or may be engineered using recombinant techniques.

[0060] The term "circularly permuted Cas9" refers to any Cas9 protein, or variant thereof, that has been occurs as a circular permutant, whereby its N- and C-termini have been topically rearranged. Such circularly permuted Cas9 proteins ("CP-Cas9"), or variants thereof, retain the ability to bind DNA when complexed with a guide RNA (gRNA). See, Oakes et al., "Protein Engineering of Cas9 for enhanced function," *Methods Enzymol*, 2014, 546: 491–511 and Oakes et al., "CRISPR-Cas9 Circular Permutants as Programmable Scaffolds for Genome Modification," *Cell*, January 10, 2019, 176: 254-267, each of are incorporated herein by reference. The instant disclosure contemplates any previously known CP-Cas9 or use a new CP-Cas9 so long as the resulting circularly permuted protein retains the ability to bind DNA when complexed with a guide RNA (gRNA).

[0061] In some embodiments, the napDNAbp is an "RNA-programmable nuclease" or "RNA-guided nuclease." The terms are used interchangeably herein and refer to a nuclease that forms a complex with (*e.g.*, binds or associates with) one or more RNA(s) that is not a target for cleavage. In some embodiments, an RNA-programmable nuclease, when in a complex with an RNA, may be referred to as a nuclease:RNA complex. Typically, the bound RNA(s) is referred to as a guide RNA (gRNA). Guide RNAs can exist as a complex of two or more RNAs, or as a single RNA molecule. Guide RNAs that exist as a single RNA molecule may be referred to as single-guide RNAs (sgRNAs), though "gRNA" is also used to refer to guide RNAs that exist as either single molecules or as a complex of two or more molecules. Typically, gRNAs that exist as a single RNA species comprise two domains: (1) a domain that shares homology to a target nucleic acid (*i.e.*, directs binding of a Cas9 complex to the target); and (2) a domain that binds a Cas9 domain. In some embodiments, domain (2) corresponds to a sequence known as a tracrRNA and comprises a stem-loop structure. In some embodiments, domain (2) is identical or homologous to a tracrRNA as provided in Jinek *et al.*, *Science* 337:816-821 (2012), the entire contents of which is incorporated herein by reference. Other

examples of gRNAs (*e.g.*, those including domain 2) can be found in International Patent Application PCT/US2014/054252, filed September 5, 2014, entitled "Switchable Cas9 Nucleases And Uses Thereof," and International Patent Application PCT/US2014/054247, filed September 5, 2014, entitled "Delivery System For Functional Nucleases," the entire contents of each are hereby incorporated by reference in their entirety. In some embodiments, a gRNA comprises two or more of domains (1) and (2), and may be referred to as an "extended gRNA." For example, an extended gRNA will bind two or more Cas9 domains and bind a target nucleic acid at two or more distinct regions, as described herein. The gRNA comprises a nucleotide sequence that complements a target site, which mediates binding of the nuclease/RNA complex to said target site, providing the sequence specificity of the nuclease:RNA complex. In some embodiments, the RNA-programmable nuclease is the (CRISPR-associated system) Cas9 endonuclease, for example, Cas9 (also known as Csn1) from *Streptococcus pyogenes* (see, *e.g.*, "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607 (2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. Science 337:816-821 (2012), the entire contents of each of which are incorporated herein by reference).

[0062] Because RNA-programmable nucleases (*e.g.*, Cas9) use RNA:DNA hybridization to target DNA cleavage sites, these proteins are able to target, in principle, any sequence specified by the guide RNA. Methods of using RNA-programmable nucleases, such as Cas9, for site-specific cleavage (*e.g.*, to modify a genome) are known in the art (see *e.g.*, Cong, L. *et al.*, Multiplex genome engineering

using CRISPR/Cas systems. *Science* **339**, 819-823 (2013); Mali, P. *et al.*, RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013); Hwang, W.Y. *et al.*, Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* **31**, 227-229 (2013); Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013); Dicarlo, J.E. *et al.*, Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Research* (2013); Jiang, W. *et al.*, RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* **31**, 233-239 (2013); the entire contents of each of which are incorporated herein by reference).

[0063] In general, a "CRISPR system" refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or other sequences and transcripts from a CRISPR locus. The tracrRNA of the system is complementary (fully or partially) to the tracr mate sequence present on the guide RNA.

[0064] The term "Cas9" or "Cas9 nuclease" refers to an RNA-guided nuclease comprising a Cas9 domain, or a fragment thereof (*e.g.*, a protein comprising an active or inactive DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A "Cas9 domain" as used herein, is a protein fragment comprising an active or inactive cleavage domain of Cas9 and/or the gRNA binding domain of Cas9. A "Cas9 protein" is a full length Cas9 protein. A Cas9 nuclease is also referred to sometimes as a casn1 nuclease or a CRISPR (Clustered Regularly Interspaced Short Palindromic Repeat)-associated nuclease. CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements, and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids.

CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc) and a Cas9 domain. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, then trimmed 3′-5′ exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA", or simply "gNRA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. See, *e.g.*, Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of which are hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art (see, *e.g.*, "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti *et al.*, J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P., Qian Y., Jia H.G., Najar F.Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E., Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C.M., Gonzales K., Chao Y., Pirzada Z.A., Eckert M.R., Vogel J., Charpentier E., Nature 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J.A., Charpentier E. *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Additional suitable Cas9 nucleases and sequences will be apparent to those of skill in the art based on this disclosure, and such Cas9 nucleases and sequences include Cas9 sequences from the organisms and loci disclosed in Chylinski, Rhun, and

Charpentier, "The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems" (2013) *RNA Biology* 10:5, 726-737; the entire contents of which are incorporated herein by reference. In some embodiments, a Cas9 nuclease comprises one or more mutations that partially impair or inactivate the DNA cleavage domain.

[0065] A nuclease-inactivated Cas9 domain may interchangeably be referred to as a "dCas9" protein (for nuclease-"dead" Cas9). Methods for generating a Cas9 domain (or a fragment thereof) having an inactive DNA cleavage domain are known (see, *e.g.*, Jinek *et al.*, *Science*. 337:816-821(2012); Qi *et al.*, "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) *Cell*. 28;152(5):1173-83, the entire contents of each of which are incorporated herein by reference). For example, the DNA cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvC1 subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvC1 subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations D10A and H840A completely inactivate the nuclease activity of *S. pyogenes* Cas9 (Jinek *et al.*, *Science*. 337:816-821(2012); Qi *et al.*, *Cell*. 28;152(5):1173-83 (2013)). In some embodiments, proteins comprising fragments of Cas9 are provided. For example, in some embodiments, a protein comprises one of two Cas9 domains: (1) the gRNA binding domain of Cas9; or (2) the DNA cleavage domain of Cas9. In some embodiments, proteins comprising Cas9 or fragments thereof are referred to as "Cas9 variants." A Cas9 variant shares homology to Cas9, or a fragment thereof. For example, a Cas9 variant is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, at least about 99.8% identical, or at least about 99.9% identical to wild type Cas9 (e.g., SpCas9 of SEQ ID NO: 2). In some embodiments, the Cas9 variant may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35,

36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more amino acid changes compared to wild type Cas9 (e.g., SpCas9 of SEQ ID NO: 2). In some embodiments, the Cas9 variant comprises a fragment of Cas9 (*e.g.*, a gRNA binding domain or a DNA-cleavage domain), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to the corresponding fragment of wild type Cas9 (e.g., SpCas9 of SEQ ID NO: 2). In some embodiments, the fragment is at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identical, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid length of a corresponding wild type Cas9 (e.g., SpCas9 of SEQ ID NO: 2).

[0066] In some embodiments, proteins comprising fragments of Cas9 are provided. In some embodiments, the fragment is at least 100 amino acids in length. In some embodiments, the fragment is at least 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, or at least 1300 amino acids in length. For example, in some embodiments, a protein comprises one of two Cas9 domains: (1) the gRNA binding domain of Cas9; or (2) the DNA cleavage domain of Cas9. In some embodiments, proteins comprising Cas9 or fragments thereof are referred to as "Cas9 variants." A Cas9 variant shares homology to Cas9. For example a Cas9 variant is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to wild type Cas9 (e.g., SpCas9 of SEQ ID NO: 2). In some embodiments, the Cas9 variant comprises a fragment of Cas9 (*e.g.*, a gRNA binding domain or a DNA-cleavage domain), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about

98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to the corresponding fragment of wild type Cas9 (e.g., SpCas9 of SEQ ID NO: 2). In some embodiments, wild type Cas9 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC_017053.1, SEQ ID NO: 1 (nucleotide); SEQ ID NO: 2 (amino acid)).

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGA
TGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAA
ATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCT
CGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGC
GAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATG
AACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTAT
CATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGC
GCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGG
ACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGT
AGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGC
TCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCC
CTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGAT
GATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAA
TTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTAT
CAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGA
CAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAGGTTATAT
TGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTA
CTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGC
TCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCC
ATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTC
CATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGG
AATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGA
TAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATA
ACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAG
AAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGA
TTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTT
CATTAGGCGCCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAAT
GAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGGGATGATTGAGGAAAG
ACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTG
GTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTA
GATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGAC
ATTTAAAGAAGATATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCTA

ACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGTC
AAAGTAATGGGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAA
GGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGA
TTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAA
AATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGATCA
CATTGTTCCACAAAGTTTCATTAAAGACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAA
ATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAA
CTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTT
GAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATG
TGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTT
AAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACG
TGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTA
AGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATG
ATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAA
CTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATG
GGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATG
CCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACC
AAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTG
ATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTA
AAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGA
CTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTC
TTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAG
CTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAG
TCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTG
AGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCA
TATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGAC
GAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTA
CAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGAT
TTGAGTCAGCTAGGAGGTGACTGA (SEQ ID NO: 1)

MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPINAS
RVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ

KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELV
KVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV
KVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKM
IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSM
PQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKL
KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSA
YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID
LSQLGGD (SEQ ID NO: 2) (encoded by SEQ ID NO: 1)
(single underline: HNH domain; double underline: RuvC domain)

[0067] In some embodiments, wild type Cas9 corresponds to, or comprises SEQ ID NO: 3 (nucleotide) and/or SEQ ID NO: 4 (amino acid):

ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGA
TGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGA
ATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCT
CGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGC
CAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATG
AACGGCACCCCATCTTTGGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTAT
CACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGC
CCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAACTCGGATGTCG
ACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATAAATGCAAGT
GGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGC
ACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACAC
CAAATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGAT
GACGATCTCGACAATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAA
CCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTAT
CCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGT
CAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTATAT
TGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGA
CGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAACGGT
AGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCC
GTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGAC

CCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGG
AATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGA
CAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCACAGTGTACA
ATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAG
AAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGA
CTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGT
CACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAAT
GAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAG
ACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGG
GCTGGGGACGATTGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTC
GATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAAC
CTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGA
ATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTT
AAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCA
GAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCC
AGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTA
CAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGACGTCGA
TCACATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATA
AGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGG
CAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGG
CTTGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGC
ATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAA
GTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGT
TAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCA
TTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAG
ATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTAT
GAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCA
ATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCC
ATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCT
TCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCT
TCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAA
CTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCAT
CGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACCAAAGTATA
GTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAAC
GAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGG
TTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCA
TAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGC
GCATACAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCT

TACCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTT
CTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATA
GATTTGTCACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGA
CCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA
(SEQ ID NO: 3) (wild type Cas9).

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGDGSPKKKRKVSSDYKDHDGDYKDHDIDYKDDDDKAAG (SEQ ID NO: 4) (wild type Cas9 encoded by SEQ ID NO: 3).

[0068] In some embodiments, wild type Cas9 corresponds to Cas9 from *Streptococcus pyogenes* (NCBI Reference Sequence: NC_002737.2, SEQ ID NO: 5 (nucleotide); and Uniport Reference Sequence: Q99ZW2, SEQ ID NO: 6 (amino acid).

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGA
TGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAA
ATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCT
CGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGC
GAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATG
AACGTCATCCTATTTTTGGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTAT

CATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGC
GCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGG
ACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGT
GGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGC
TCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGTTTGACCC
CTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGAT
GATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAA
TTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAGGCTCCCCTAT
CAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGA
CAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAACGGATATGCAGGTTATAT
TGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTA
CTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGC
TCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCC
ATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTC
CATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGG
AATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGA
TAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATA
ACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAG
AAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGA
TTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTT
CATTAGGTACCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAAT
GAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAG
ACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTG
GTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTA
GATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGAC
ATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAACATATTGCAA
ATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTTGATGAATTGGTC
AAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGACAACTCA
AAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTC
AGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTC
CAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTGATTATGATGTCGA
TCACATTGTTCCACAAAGTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATA
AAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGA
CAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGG
TTTGAGTGAACTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGC
ATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAG
GTTAAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGT
ACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGTTGGAACTGCTTTGA

TTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAA
ATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCAT
GAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTA
ATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCC
ATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTT
ACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTT
TTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAG
TTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGAT
TGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTTAATCATTAAACTACCTAAATATA
GTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAAT
GAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGG
TAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTA
TTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGT
GCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTT
GACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGT
CTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATT
GATTTGAGTCAGCTAGGAGGTGACTGA (SEQ ID NO: 5)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS

AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD (SEQ ID NO: 6) (single underline: HNH domain; double underline: RuvC domain)

[0069] In some embodiments, Cas9 refers to a Cas9 nickase having a D10A substitution (e.g., *S. pyogenes* Cas9 Q99ZW2 (D10A)) (SEQ ID NO: 7):

MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD (SEQ ID NO: 7) (comprises D10A substitution).

(single underline: HNH domain; double underline: RuvC domain)

[0070] In other embodiments, Cas9 refers to a Cas9 nickase having a H840A substitution (e.g., *S. pyogenes* Cas9 Q99ZW2 (H840A)) (SEQ ID NO: 8):

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ

KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVD**A**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD (SEQ ID NO: 8) (single underline: HNH domain; double underline: RuvC domain; H840A mutation shown in **bold**)

[0071] In still other embodiments, Cas9 refers to a dead Cas9 having D10A and H840A substitutions (e.g., *S. pyogenes* Cas9 Q99ZW2 (D10A) (H840A)) (SEQ ID NO: 9):

MDKKYSIGL**A**IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVD**A**IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS

AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD (SEQ ID NO: 9)

(D10A and H840A mutations shown in **bold**; see, *e.g.*, Qi *et al.*, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*. 2013; 152(5):1173-83, the entire contents of which are incorporated herein by reference).

[0072] In some embodiments, Cas9 refers to Cas9 protein derived from: *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquisI* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1), *Listeria innocua* (NCBI Ref: NP_472073.1), *Campylobacter jejuni* (NCBI Ref: YP_002344900.1); *Geobacillus stearothermophilus* (NCBI Ref: NZ_CP008934.1); or *Neisseria meningitidis* (NCBI Ref: YP_002342100.1) or to a Cas9 from any other organism.

[0073] In some embodiments, a Cas9 domain comprising one or more mutations provided herein is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 92%, at least about 95% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to SEQ ID NO: 2. In some embodiments, variants of a Cas9 domain comprising one or more mutations provided herein are provided having amino acid sequences which are shorter, or longer than SEQ ID NO: 2, by about 5 amino acids, by about 10 amino acids, by about 15 amino acids, by about 20 amino acids, by about 25 amino acids, by about 30 amino acids, by about 40 amino acids, by about 50 amino acids, by about 75 amino acids, by about 100 amino acids, or more.

[0074] In some embodiments, the Cas9 domain comprises a D10A mutation, while the residue at position 840 remains a histidine relative to the amino acid sequence as provided in SEQ ID NO: 2, or at corresponding positions in any of the amino acid sequences provided in SEQ ID NO: 2. Without

wishing to be bound by any particular theory, the presence of the catalytic residue H840 restores the activity of the Cas9 to cleave the non-edited (*e.g.,* non-deaminated) strand containing a G opposite the targeted C. Restoration of H840 (*e.g.,* from A840) does not result in the cleavage of the target strand containing the C. Such Cas9 variants are able to generate a single-strand DNA break (nick) at a specific location based on the gRNA-defined target sequence, leading to repair of the non-edited strand, ultimately resulting in a base change (*e.g.*, a G to A change) on the non-edited strand. Briefly, the C of a C-G base pair can be deaminated to a U by a deaminase, *e.g.,* an APOBEC deaminase. Nicking the non-edited strand, the strand having the G, facilitates removal of the G via mismatch repair mechanisms. Uracil-DNA glycosylase inhibitor protein (UGI) inhibits Uracil-DNA glycosylase (UDG), which prevents removal of the U.

[0075] In other embodiments, dCas9 variants having mutations other than D10A and H840A are provided, which, *e.g.*, result in nuclease inactivated Cas9 (dCas9). Such mutations, by way of example, include other amino acid substitutions at D10 and H820, or other substitutions within the nuclease domains of Cas9 (*e.g.*, substitutions in the HNH nuclease subdomain and/or the RuvC1 subdomain).

[0076] The term "Cas9 nickase" or "Cas9n" or "nCas9" as used herein, refers to a Cas9 domain that is capable of cleaving one strand of the duplexed nucleic acid molecule (*e.g.*, a duplexed DNA molecule). In some embodiments, a Cas9 nickase comprises a D10A mutation and has a histidine at position H840 of SEQ ID NO: 2, or a corresponding mutation in any of SEQ ID NOs: 2. For example, in some embodiments, a Cas9 nickase comprises the amino acid sequence as set forth in SEQ ID NO: 8 comprising the H840A substitution. Such a Cas9 nickase (Cas9n) has an active HNH nuclease domain and is able to cleave the non-targeted strand of DNA, *i.e.*, the strand bound by the gRNA. Further, such a Cas9 nickase has an inactive RuvC nuclease domain and is not able to cleave the targeted strand of the DNA, *i.e.*, the strand where base editing is desired. In some embodiments, any of the Cas9 domains provided herein comprises a D10A mutation (e.g., SEQ ID NO: 7). In

some embodiments, any of the Cas9 domains provided herein comprises a H840A mutation (SEQ ID NO: 8). Exemplary Cas9 nickases are shown below. However, it should be appreciated that additional Cas9 nickases that generate a single-stranded DNA break of a DNA duplex would be apparent to the skilled artisan and are within the scope of this disclosure.

[0077] In some embodiments, Cas9 fusion proteins as provided herein comprise the full-length amino acid sequence of a Cas9 domain, *e.g.*, one of the sequences provided above. In other embodiments, however, fusion proteins as provided herein do not comprise a full-length Cas9 sequence, but only a fragment thereof. For example, in some embodiments, a Cas9 fusion protein provided herein comprises a Cas9 fragment, wherein the fragment binds crRNA and tracrRNA or a sgRNA, but does not comprise a functional nuclease domain, *e.g.*, it comprises only a truncated version of a nuclease domain or no nuclease domain at all. Exemplary amino acid sequences of suitable Cas9 domains and Cas9 fragments are provided herein, and additional suitable sequences of Cas9 domains and Cas9 fragments will be apparent to those of skill in the art. In some embodiments, a Cas9 fragment is at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identical, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid length of a corresponding wild type Cas9 domain. In some embodiments, a Cas9 fragment comprises at least at least 100 amino acids in length. In some embodiments, the Cas9 fragment is at least 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, or at least 1600 amino acids of a corresponding wild type Cas9 domain. In some embodiments, the Cas9 fragment comprises an amino acid sequence that has at least 10, at least 15, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 150, at least 200, at least 250, at least 300, at least 350, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1100, or at least 1200 identical contiguous amino acid residues of a corresponding wild type Cas9 domain. In

some embodiments, the wild-type protein is *S. pyogenes* Cas9 (SpCas9) of SEQ ID NO: 2.

[0078] In some embodiments, Cas9 fusion proteins as provided herein comprise the full-length amino acid sequence of a Cas9 domain, *e.g.*, one of the Cas9 sequences provided herein. In other embodiments, however, fusion proteins as provided herein do not comprise a full-length Cas9 sequence, but only a fragment thereof. For example, in some embodiments, a Cas9 fusion protein provided herein comprises a Cas9 fragment, wherein the fragment binds crRNA and tracrRNA or sgRNA, but does not comprise a functional nuclease domain, *e.g.*, in that it comprises only a truncated version of a nuclease domain or no nuclease domain at all. Exemplary amino acid sequences of suitable Cas9 domains and Cas9 fragments are provided herein, and additional suitable sequences of Cas9 domains and fragments will be apparent to those of ordinary skill in the art. In some embodiments, Cas9 refers to Cas9 from: *Corynebacterium ulcerans* (NCBI Refs: NC_015683.1, NC_017317.1); *Corynebacterium diphtheria* (NCBI Refs: NC_016782.1, NC_016786.1); *Spiroplasma syrphidicola* (NCBI Ref: NC_021284.1); *Prevotella intermedia* (NCBI Ref: NC_017861.1); *Spiroplasma taiwanense* (NCBI Ref: NC_021846.1); *Streptococcus iniae* (NCBI Ref: NC_021314.1); *Belliella baltica* (NCBI Ref: NC_018010.1); *Psychroflexus torquis I* (NCBI Ref: NC_018721.1); *Streptococcus thermophilus* (NCBI Ref: YP_820832.1); *Geobacillus stearothermophilus* (NCBI Ref: NZ_CP008934.1); *Listeria innocua* (NCBI Ref: NP_472073.1); *Campylobacter jejuni* (NCBI Ref: YP_002344900.1); or *Neisseria. meningitidis* (NCBI Ref: YP_002342100.1).

[0079] The term "deaminase" or "deaminase domain," as used herein, refers to a protein or enzyme that catalyzes a deamination reaction. In some embodiments, the deaminase or deaminase domain is a naturally-occurring deaminase from an organism, such as a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse. In some embodiments, the deaminase or deaminase domain is a variant of a naturally-occurring deaminase from an organism, that does not occur in nature. For example, in some embodiments, the deaminase or deaminase domain is at least 50%, at least 55%, at least 60%, at least

65%, at least 70%, at least 75% at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring deaminase from an organism.

[0080] In some embodiments, the deaminase or deaminase domain is a cytidine deaminase, catalyzing the hydrolytic deamination of cytidine or deoxycytidine to uridine or deoxyuridine, respectively. In some embodiments, the deaminase or deaminase domain is a cytidine deaminase domain, catalyzing the hydrolytic deamination of cytosine to uracil. In some embodiments, the cytidine deaminase catalyzes the hydrolytic deamination of cytidine or cytosine in deoxyribonucleic acid (DNA). In some embodiments, the cytidine deaminase domain comprises the amino acid sequence of any one disclosed herein. In some embodiments, the cytidine deaminase or cytidine deaminase domain is a naturally-occurring cytidine deaminase from an organism, such as a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse. In some embodiments, the cytidine deaminase or cytidine deaminase domain is a variant of a naturally-occurring cytidine deaminase from an organism that does not occur in nature. For example, in some embodiments, the cytidine deaminase or cytidine deaminase domain is at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75% at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring cytidine deaminase from an organism, such as a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse.

[0081] In some embodiments, the deaminase or deaminase domain is an adenosine deaminase, which catalyzes the hydrolytic deamination of adenine or adenosine. In some embodiments, the deaminase or deaminase domain is an adenosine deaminase, catalyzing the hydrolytic deamination of adenosine or deoxyadenosine to inosine or deoxyinosine, respectively. In some embodiments, the adenosine deaminase catalyzes the hydrolytic deamination of adenine or adenosine in deoxyribonucleic acid (DNA). The adenosine deaminases (*e.g.*, engineered adenosine deaminases, evolved adenosine deaminases) provided herein may be from any organism, such as a bacterium. In

some embodiments, the deaminase or deaminase domain is a variant of a naturally-occurring deaminase from an organism. In some embodiments, the deaminase or deaminase domain does not occur in nature. For example, in some embodiments, the deaminase or deaminase domain is at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75% at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring deaminase. In some embodiments, the adenosine deaminase is from a bacterium, such as *E.coli*, *S. aureus*, *S. typhi*, *S. putrefaciens*, *H. influenzae*, or *C. crescentus*. In some embodiments, the adenosine deaminase is a TadA deaminase. In some embodiments, the TadA deaminase is an *E. coli* TadA deaminase (ecTadA). In some embodiments, the TadA deaminase is a truncated *E. coli* TadA deaminase. For example, the truncated ecTadA may be missing one or more N-terminal amino acids relative to a full-length ecTadA. In some embodiments, the truncated ecTadA may be missing 1, 2, 3, 4, 5 ,6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 N-terminal amino acid residues relative to the full length ecTadA. In some embodiments, the truncated ecTadA may be missing 1, 2, 3, 4, 5 ,6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 6, 17, 18, 19, or 20 C-terminal amino acid residues relative to the full length ecTadA. In some embodiments, the ecTadA deaminase does not comprise an N-terminal methionine.

[0082] In some embodiments, the TadA deaminase is an N-terminal truncated TadA. In certain embodiments, the adenosine deaminase comprises the amino acid sequence:

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGI
LADECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 77).

[0083] In some embodiments the TadA deaminase is a full-length *E. coli* TadA deaminase. For example, in certain embodiments, the adenosine deaminase comprises the amino acid sequence:

MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPG
MNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 78)

[0084] It should be appreciated, however, that additional adenosine deaminases useful in the present application would be apparent to the skilled artisan and are within the scope of this disclosure. For example, the adenosine deaminase may be a homolog of an ADAT. Exemplary ADAT homologs include, without limitation:

Staphylococcus aureus TadA:

MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAK VLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQSNFNHRAIVDKGV LKEACSTLLTTFFKNLRANKKSTN (SEQ ID NO: 79)

Bacillus subtilis TadA:

MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACKALGT WRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNHQAEVVSGVLEEE CGGMLSAFFRELRKKKKAARKNLSE (SEQ ID NO: 63)

Salmonella typhimurium TadA:

MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDPTAH AEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPG MNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV (SEQ ID NO: 80)

Shewanella putrefaciens TadA:

MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGKKLENYR LLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQVEVTSGVLAEACS AQLSRFFKRRRDEKKALKLAQRAQQGIE (SEQ ID NO: 81)

Haemophilus influenzae F3031 (*H. influenzae*) TadA:

MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHAEIIALR NGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHFFDDYKMNHTLEI TSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK (SEQ ID NO: 82)

Caulobacter crescentus TadA:

MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIAAMR AAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKFFAQPTCHWRPEV TGGVLADESADLLRGFFRARRKAKI (SEQ ID NO: 83)

Geobacter sulfurreducens TadA:

MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHAEMIAIR QAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDLSADPRLNHQVRL SPGVCQEECGTMLSDFFRDLRRRKKAKATPALFIDERKVPPEP (SEQ ID NO: 84)

[0085] The term "effective amount," as used herein, refers to an amount of a biologically active agent

that is sufficient to elicit a desired biological response. For example, in some embodiments, an effective amount of a nuclease may refer to the amount of the nuclease that is sufficient to induce cleavage of a target site specifically bound and cleaved by the nuclease. In some embodiments, an effective amount of a fusion protein provided herein, *e.g.*, of a fusion protein comprising a Cas9 domain and a nucleic acid editing domain (*e.g.*, a deaminase domain) may refer to the amount of the fusion protein that is sufficient to induce editing of a target site specifically bound and edited by the fusion protein. As will be appreciated by the skilled artisan, the effective amount of an agent, *e.g.*, a fusion protein, a nuclease, a deaminase, a recombinase, a hybrid protein, a protein dimer, a complex of a protein (or protein dimer) and a polynucleotide, or a polynucleotide, may vary depending on various factors such as, for example, on the desired biological response, *e.g.*, on the specific allele, genome, or target site to be edited; on the cell or tissue being targeted; and on the agent (*e.g.*, Cas9 domain, fusion protein, vector, cell, *etc.*) being used.

[0086] The term "immediately adjacent" as used in the context of two nucleic acid sequences refers to two sequences that directly abut each other as part of the same nucleic acid molecule and are not separated by one or more nucleotides. Accordingly, sequences are immediately adjacent, when the nucleotide at the 3′-end of one of the sequences is directly connected to nucleotide at the 5′-end of the other sequence via a phosphodiester bond.

[0087] The term "linker," as used herein, refers to a chemical group or a molecule linking two molecules or moieties, *e.g.*, two domains of a fusion protein, such as, for example, a nuclease-inactive Cas9 domain and a nucleic acid editing domain (*e.g.*, a deaminase domain). A linker may be, for example, an amino acid sequence, a peptide, or a polymer of any length and composition. In some embodiments, a linker joins a gRNA binding domain of an RNA-programmable nuclease, including a Cas9 nuclease domain, and the catalytic domain of a nucleic-acid editing protein. In some embodiments, a linker joins a dCas9 and a nucleic-acid editing protein. In some embodiments, a linker joins a Cas9n and a nucleic-acid editing protein. In some embodiments, a linker joins an RNA-

programmable nuclease domain and a UGI domain. In some embodiments, a linker joins a dCas9 and a UGI domain. In some embodiments, a linker joins a Cas9n and a UGI domain. Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker is an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 1-100 amino acids in length, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, or 150-200 amino acids in length. Longer or shorter linkers are also contemplated. In some embodiments, a linker comprises the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 89), which may also be referred to as the XTEN linker. In some embodiments, a linker comprises the amino acid sequence SGGS (SEQ ID NO: 90). In some embodiments, a linker comprises the amino acid sequence (SGGS)2-SGSETPGTSESATPES-(SGGS)2 (SEQ ID NO: 96), which may also be referred to as (SGGS)2-XTEN-(SGGS)2. In some embodiments, a linker comprises (SGGS)n (SEQ ID NO: 92), (GGGS)n (SEQ ID NO: 94), (GGGGS)n (SEQ ID NO: 96), (G)n (SEQ ID NO: 97), (EAAAK)n (SEQ ID NO: 99), (GGS)n (SEQ ID NO: 101), SGGS(GGS)n (SEQ ID NO: 103), (SGGS)n-SGSETPGTSESATPES-(SGGS)n (SEQ ID NO: 98), or $(XP)_n$ motif, or a combination of any of these, wherein n is independently an integer between 1 and 30, and wherein X is any amino acid. In some embodiments, n is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15. In some embodiments, n is 1, 3, or 7. In some embodiments, the linker comprises the amino acid sequence:

GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG

TSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS (SEQ ID NO: 110);

SGGSSGGSSGSETPGTSESATPES (SEQ ID NO: 111);

SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS (SEQ ID NO: 109);

SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ

ID NO: 107); or

PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGSEPATS (SEQ ID NO: 108).

[0088] The term "mutation," as used herein, refers to a substitution of a residue within a sequence, *e.g.*, a nucleic acid or amino acid sequence, with another residue, or a deletion or insertion of one or more residues within a sequence. Mutations are typically described herein by identifying the original residue followed by the position of the residue within the sequence and by the identity of the newly substituted residue. Various methods for making the amino acid substitutions (mutations) provided herein are well known in the art, and are provided by, for example, Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)).

[0089] The terms "nucleic acid" and "nucleic acid molecule," as used herein, refer to a compound comprising a nucleobase and an acidic moiety, *e.g.*, a nucleoside, a nucleotide, or a polymer of nucleotides. Typically, polymeric nucleic acids, *e.g.*, nucleic acid molecules comprising three or more nucleotides are linear molecules, in which adjacent nucleotides are linked to each other via a phosphodiester linkage. In some embodiments, "nucleic acid" refers to individual nucleic acid residues (*e.g.* nucleotides and/or nucleosides). In some embodiments, "nucleic acid" refers to an oligonucleotide chain comprising three or more individual nucleotide residues. As used herein, the terms "oligonucleotide" and "polynucleotide" can be used interchangeably to refer to a polymer of nucleotides (*e.g.*, a string of at least three nucleotides). In some embodiments, "nucleic acid" encompasses RNA as well as single and/or double-stranded DNA. Nucleic acids may be naturally occurring, for example, in the context of a genome, a transcript, mRNA, tRNA, rRNA, siRNA, snRNA, a plasmid, cosmid, chromosome, chromatid, or other naturally occurring nucleic acid molecule. On the other hand, a nucleic acid molecule may be a non-naturally occurring molecule, *e.g.*, a DNA or RNA, an artificial chromosome, an engineered genome, or fragment thereof, or a

synthetic DNA, RNA, DNA/RNA hybrid, or including non-naturally occurring nucleotides or nucleosides. Furthermore, the terms "nucleic acid," "DNA," "RNA," and/or similar terms include nucleic acid analogs, *e.g.,* analogs having other than a phosphodiester backbone. Nucleic acids can be purified from natural sources, produced using expression systems and optionally purified, chemically synthesized, *etc.* Where appropriate, *e.g.*, in the case of chemically synthesized molecules, nucleic acids can comprise nucleoside analogs such as analogs having chemically modified bases or sugars, and backbone modifications. A nucleic acid sequence is presented in the 5′ to 3′ direction unless otherwise indicated. In some embodiments, a nucleic acid is or comprises natural nucleosides (*e.g.* adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (*e.g.*, 2-aminoadenosine, 2-thiothymidine, inosine, pyrrolo-pyrimidine, 3-methyl adenosine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-aminoadenosine, 7-deazaadenosine, 7-deazaguanosine, 8-oxoadenosine, 8-oxoguanosine, O(6)-methylguanine, and 2-thiocytidine); chemically modified bases; biologically modified bases (*e.g.*, methylated bases); intercalated bases; modified sugars (*e.g.*, 2′-fluororibose, ribose, 2′-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (*e.g.*, phosphorothioates and 5′-*N*-phosphoramidite linkages). In some embodiments, an RNA is an RNA associated with the Cas9 system. For example, the RNA may be a CRISPR RNA (crRNA), a trans-encoded small RNA (tracrRNA), a single guide RNA (sgRNA), or a guide RNA (gRNA).

[0090] The term "nucleic acid editing domain," as used herein refers to a protein or enzyme capable of making one or more modifications (*e.g.*, deamination of a cytidine residue) to a nucleic acid (*e.g.*, DNA or RNA). Exemplary nucleic acid editing domains include, but are not limited to a deaminase, a nuclease, a nickase, a recombinase, a methyltransferase, a methylase, an acetylase, an acetyltransferase, a transcriptional activator, or a transcriptional repressor domain. In some embodiments, the nucleic acid editing domain is a deaminase, a nuclease, a nickase, a recombinase, a

methyltransferase, a methylase, an acetylase, an acetyltransferase, a transcriptional activator, or a transcriptional repressor domain. In some embodiments, the nucleic acid editing domain is a deaminase domain (*e.g.*, a cytidine deaminase, such as an APOBEC or an AID deaminase, or an adenosine deaminase, such as ecTadA). In some embodiments, the nucleic acid editing domain is a cytidine deaminase domain (*e.g.*, an APOBEC or an AID deaminase). In some embodiments, the nucleic acid editing domain is an adenosine deaminase domain (*e.g.*, an ecTadA).

[0091] The term "nuclear localization sequence" or "NLS" refers to an amino acid sequence that promotes import of a protein into the cell nucleus, for example, by nuclear transport. Nuclear localization sequences are known in the art and would be apparent to the skilled artisan. For example, NLS sequences are described in Plank *et al.*, international PCT application, PCT/EP2000/011690, filed November 23, 2000, published as WO/2001/038547 on May 31, 2001, the contents of which are incorporated herein by reference for its disclosure of exemplary nuclear localization sequences. In some embodiments, a NLS comprises the amino acid sequence PKKKRKV (SEQ ID NO: 113) or MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 114).

[0092] The term "proliferative disease," as used herein, refers to any disease in which cell or tissue homeostasis is disturbed in that a cell or cell population exhibits an abnormally elevated proliferation rate. Proliferative diseases include hyperproliferative diseases, such as pre-neoplastic hyperplastic conditions and neoplastic diseases. Neoplastic diseases are characterized by an abnormal proliferation of cells and include both benign and malignant neoplasias. Malignant neoplasia is also referred to as cancer.

[0093] The terms "protein," "peptide," and "polypeptide" are used interchangeably herein, and refer to a polymer of amino acid residues linked together by peptide (amide) bonds. The terms refer to a protein, peptide, or polypeptide of any size, structure, or function. Typically, a protein, peptide, or polypeptide will be at least three amino acids long. A protein, peptide, or polypeptide may refer to an individual protein or a collection of proteins. One or more of the amino acids in a protein, peptide, or

polypeptide may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, *etc.* A protein, peptide, or polypeptide may also be a single molecule or may be a multi-molecular complex. A protein, peptide, or polypeptide may be just a fragment of a naturally occurring protein or peptide. A protein, peptide, or polypeptide may be naturally occurring, or synthetic, or any combination thereof. The term "fusion protein" as used herein refers to a hybrid polypeptide which comprises protein domains from at least two different proteins, or at least two identical protein domains (*i.e.*, a homodimer). One protein may be located at the amino-terminal (N-terminal) portion of the fusion protein or at the carboxy-terminal (C-terminal) protein thus forming an "amino-terminal fusion protein" or a "carboxy-terminal fusion protein," respectively. A protein may comprise different domains, for example, a nucleic acid binding domain (*e.g.*, the gRNA binding domain of Cas9 that directs the binding of the protein to a target site) and a nucleic acid cleavage domain or a catalytic domain of a nucleic acid editing protein. In some embodiments, a protein comprises a proteinaceous part, *e.g.*, an amino acid sequence constituting a nucleic acid binding domain, and an organic compound, *e.g.*, a compound that can act as a nucleic acid cleavage agent. In some embodiments, a protein is in a complex with, or is in association with, a nucleic acid, *e.g.*, RNA. Any of the proteins provided herein may be produced by any method known in the art. For example, the proteins provided herein may be produced via protein expression and purification, which is especially suited for fusion proteins comprising a peptide linker. Methods for protein expression and purification are well known, and include those described by Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), the entire contents of which are incorporated herein by reference.

[0094] The term "subject," as used herein, refers to an individual organism, for example, an individual mammal. In some embodiments, the subject is a human. In some embodiments, the subject

is a non-human mammal. In some embodiments, the subject is a non-human primate. In some embodiments, the subject is a rodent. In some embodiments, the subject is a sheep, a goat, a cattle, a cat, or a dog. In some embodiments, the subject is a vertebrate, an amphibian, a reptile, a fish, an insect, a fly, or a nematode. In some embodiments, the subject is a plant or a fungus. In some embodiments, the subject is a research animal (*e.g.*, a rat, a mouse, or a non-human primate). In some embodiments, the subject is genetically engineered, *e.g.*, a genetically engineered non-human subject. The subject may be of either sex, of any age, and at any stage of development.

[0095] The term "target site" refers to a nucleic acid sequence or a nucleotide within a nucleic acid that is targeted or modified by an effector domain that is fused to a napDNAbp. In some embodiments, a "target site" is a sequence within a nucleic acid molecule that is deaminated by a deaminase or a fusion protein comprising a deaminase, (*e.g.,* a dCas9-deaminase fusion protein or a Cas9n-deaminase fusion protein provided herein). In some embodiments, the target site refers to a sequence within a nucleic acid molecule that is cleaved by a napDNAbp (*e.g.,* a nuclease active Cas9 domain) provided herein. The target site is contained within a target sequence (*e.g.,* a target sequence comprising a reporter gene, or a target sequence comprising a gene located in a safe harbor locus).

[0096] The terms "treatment," "treat," and "treating," refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. As used herein, the terms "treatment," "treat," and "treating" refer to a clinical intervention aimed to reverse, alleviate, delay the onset of, or inhibit the progress of a disease or disorder, or one or more symptoms thereof, as described herein. In some embodiments, treatment may be administered after one or more symptoms have developed and/or after a disease has been diagnosed. In other embodiments, treatment may be administered in the absence of symptoms, *e.g.*, to prevent or delay onset of a symptom or inhibit onset or progression of a disease. For example, treatment may be administered to a susceptible individual prior to the onset of symptoms (*e.g.*, in light of a history of symptoms and/or in light of genetic or other susceptibility factors). Treatment may also

be continued after symptoms have resolved, for example, to prevent or delay their recurrence.

[0097] The term "pharmaceutical composition," as used herein, refers to a composition that can be administrated to a subject in the context of treatment of a disease or disorder. In some embodiments, a pharmaceutical composition comprises an active ingredient, e.g., a nuclease or a nucleic acid encoding a nuclease, and a pharmaceutically acceptable excipient.

[0098] The term "uracil glycosylase inhibitor" or "UGI," as used herein, refers to a protein that is capable of inhibiting a uracil-DNA glycosylase base-excision repair enzyme. In some embodiments, a UGI domain comprises a wild-type UGI or a UGI as set forth in SEQ ID NO: 115-120. In some embodiments, the UGI proteins provided herein include fragments of UGI and proteins homologous to a UGI or a UGI fragment. For example, in some embodiments, a UGI domain comprises a fragment of the amino acid sequence set forth in SEQ ID NO: 115-120. In some embodiments, a UGI fragment comprises an amino acid sequence that comprises at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid sequence as set forth in SEQ ID NO: 115-120. In some embodiments, a UGI comprises an amino acid sequence homologous to the amino acid sequence set forth in SEQ ID NO: 115-120, or an amino acid sequence homologous to a fragment of the amino acid sequence set forth in SEQ ID NO: 115-120. In some embodiments, proteins comprising UGI or fragments of UGI or homologs of UGI or UGI fragments are referred to as "UGI variants." A UGI variant shares homology to UGI, or a fragment thereof. For example a UGI variant is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, at least 99% identical, at least 99.5% identical, or at least 99.9% identical to a wild type UGI or a UGI as set forth in SEQ ID NO: 115-120. In some embodiments, the UGI variant comprises a fragment of UGI, such that the fragment is at least 70% identical, at least 80% identical, at least 90% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, at

least 99% identical, at least 99.5% identical, or at least 99.9% to the corresponding fragment of wild-type UGI or a UGI as set forth in SEQ ID NO: 115-120. In some embodiments, the UGI comprises the amino acid sequence of SEQ ID NO: 115, as set forth below. Exemplary Uracil-DNA glycosylase inhibitor (UGI; >sp|P14739|UNGI_BPPB2)

MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPW ALVIQDSNGENKIKML (SEQ ID NO: 115).

[0099] The term "catalytically inactive inosine-specific nuclease," or "dead inosine-specific nuclease (dISN)," as used herein, refers to a protein that is capable of inhibiting an inosine-specific nuclease. Without wishing to be bound by any particular theory, catalytically inactive inosine glycosylases (*e.g.,* alkyl adenine glycosylase [AAG]) will bind inosine, but will not create an abasic site or remove the inosine, thereby sterically blocking the newly-formed inosine moiety from DNA damage/repair mechanisms. In some embodiments, the catalytically inactive inosine-specific nuclease may be capable of binding an inosine in a nucleic acid but does not cleave the nucleic acid. Exemplary catalytically inactive inosine-specific nucleases include, without limitation, catalytically inactive alkyl adenosine glycosylase (AAG nuclease), for example, from a human, and catalytically inactive endonuclease V (EndoV nuclease), for example, from *E. coli*. In some embodiments, the catalytically inactive AAG nuclease comprises an E125Q mutation as shown in SEQ ID NO: 40, or a corresponding mutation in another AAG nuclease. In some embodiments, the catalytically inactive AAG nuclease comprises the amino acid sequence set forth in SEQ ID NO: 40. In some embodiments, the catalytically inactive EndoV nuclease comprises an D35A mutation as shown in SEQ ID NO: 41, or a corresponding mutation in another EndoV nuclease. In some embodiments, the catalytically inactive EndoV nuclease comprises the amino acid sequence set forth in SEQ ID NO: 41. It should be appreciated that other catalytically inactive inosine-specific nucleases (dISNs) would be apparent to the skilled artisan and are within the scope of this disclosure. Various examples include:

Truncated AAG (*H. sapiens*) nuclease (E125Q); mutated residue shown in **bold**.

KGHLTRLGLEFFDQPAVPLARAFLGQVLVRRLPNGTELRGRIVET**Q**AYLGPEDEAAHSRGGRQTPRNR
GMFMKPGTLYVYIIYGMYFCMNISSQGDGACVLLRALEPLEGLETMRQLRSTLRKGTASRVLKDRELC
SGPSKLCQALAINKSFDQRDLAQDEAVWLERGPLEPSEPAVVAAARVGVGHAGEWARKPLRFYVRGSP
WVSVVDRVAEQDTQA (SEQ ID NO: 116); and

EndoV nuclease (D35A); mutated residue shown in bold.

DLASLRAQQIELASSVIREDRLDKDPPDLIAGA**A**VGFEQGGEVTRAAMVLLKYPSLELVEYKVARIAT
TMPYIPGFLSFREYPALLAAWEMLSQKPDLVFVDGHGISHPRRLGVASHFGLLVDVPTIGVAKKRLCG
KFEPLSSEPGALAPLMDKGEQLAWVWRSKARCNPLFIATGHRVSVDSALAWVQRCMKGYRLPEPTRWA
DAVASERPAFVRYTANQP (SEQ ID NO: 117).

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS

[00100] *Streptococcus pyogenes* Cas9 (SpCas9) is a widely-utilized genome-editing tool, but is restricted in genome targeting by the requirement for an NGG PAM sequence, which can be limiting for precision genome editing applications such as base editing, homology-directed repair, and predictable template-free genome editing. While SpCas9 variants with alternative PAM requirements have been previously reported, their targeting scope remains restricted primarily to G-containing PAMs.

[00101] The present application provides three SpCas9 variants capable of recognizing NRTH, NRRH, and NRCH PAMs, respectively, using an improved phage-assisted continuous evolution (PACE) Cas9 binding selection. These PAM sequence preferences are provided for these SpCas9 variants, along with the previously reported SpCas9-NG variant, by cytosine base editing, indel formation, and adenine base editing in a panel of 64 mammalian potential cell target sites. In further aspects, the present application provides the editing efficiencies of the SpCas9 variants on a mammalian cell library of ~12,000 genomically integrated sgRNA/protospacer targets.

[00102] Some aspects of this disclosure provide Cas9 proteins (e.g., SgCas9) that efficiently target nucleic acid sequences that do not include the canonical PAM sequence (5′-NGG-3′, where N is any nucleotide, for example A, T, G, or C) at their 3'-ends. It should be appreciated that the phrase "Cas9 proteins" can refer to isolated Cas9 proteins or Cas9 domains as part of fusion proteins. In some

embodiments, the Cas9 domains provided herein comprise one or more mutations identified in directed evolution experiments using a target sequence library comprising randomized PAM sequences. The non-PAM restricted Cas9 domains provided herein are useful for targeting DNA sequences that do not comprise the canonical PAM sequence at their 3'-end and thus greatly extend the applicability and usefulness of Cas9 technology for gene editing. The evolution of Cas9 domains that are not restricted to the canonical 5′-NGG-3′ PAM sequence has been previously described, for example, in International Patent Application No., PCT/US2016/058345, filed October 22, 2016, and published as Patent Publication No. WO 2017/070633, published April 27, 2017, entitled "Evolved Cas9 Proteins for Gene Editing" which is herein incorporated by reference in its entirety. In addition to the Cas9 mutations identified and proteins listed in Publication No. WO 2017/070633, provided herein are novel additional mutations and Cas9 domains that have activity on target sequences comprising non-canonical PAM sequences. It should be understood that any of the mutations listed in Patent Publication No. WO 2017/070633 may be combined with or used in lieu of any of the mutations or Cas9 domains disclosed herein, unless explicity stated otherwise.

[00103] Some aspects of this disclosure provide fusion proteins that comprise a Cas9 domain and an effector domain, for example, a nucleic acid editing domain, such as a deaminase domain, a nuclease domain, a nickase domain, a recombinase domain, a methyltransferase domain, a methylase domain, an acetylase domain, an acetyltransferase domain, a transcriptional activator domain, or a transcriptional repressor domain.

[00104] The deamination of a nucleobase by a deaminase can lead to a point mutation at the specific residue, which is referred to herein as nucleic acid editing. Fusion proteins comprising a Cas9 domain or variant thereof and a nucleic acid editing domain can thus be used for the targeted editing of nucleic acid sequences. Such fusion proteins are useful for targeted editing of DNA in vitro, e.g., for the generation of mutant cells or animals; for the introduction of targeted mutations, e.g., for the correction of genetic defects in cells ex vivo, e.g., in cells obtained from a subject that are

subsequently re-introduced into the same or another subject; and for the introduction of targeted mutations, e.g., the correction of genetic defects or the introduction of deactivating mutations in disease-associated genes in a subject in vivo. Typically, the Cas9 domain of the fusion proteins described herein is a Cas9 domain comprising one or more mutations provided herein (e.g., an "xCas9" domain) that has impaired nuclease activity (e.g., a nuclease-inactive xCas9 domain). For example, in some embodiments, the Cas9 domain comprises a D10A and/or a H840A mutation in the amino acid sequence provided in SEQ ID NO: 2. Methods for the use of fusion proteins comprising Cas9 as described herein are also provided.

[00105] Additional suitable nuclease-inactive Cas9 domains will be apparent to those of skill in the art based on this disclosure. Such additional exemplary suitable nuclease-inactive Cas9 domains include, but are not limited to, D10A, D839A, H840A, N863A, D10A/D839A, D10A/H840A, D10A/N863A, D839A/H840A, D839A/N863A, D10A/D839A/H840A, and D10A/D839A/H840A/N863A mutant proteins (See, e.g., Prashant et al., "Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering," *Nature Biotechnology*, 2013; 31(9): 833-838, the entire contents of which are incorporated herein by reference). In some embodiments, the Cas9 domain comprises a D10A mutation of the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in any of the amino acid sequences provided in SEQ ID NOs: 2.

[00106] The base editors disclosed herein may also comprise a circular permutant Cas9 variant. The term "circularly permuted Cas9" refers to any Cas9 protein, or variant thereof, that occurs or has been modify to occur as a circular permutant, whereby its N- and C-termini have been topically rearranged. Such circularly permuted Cas9 proteins ("CP-Cas9"), or variants thereof, retain the ability to bind DNA when complexed with a guide RNA (gRNA). See, Oakes et al., "Protein Engineering of Cas9 for enhanced function," *Methods Enzymol*, 2014, 546: 491–511 and Oakes et al., "CRISPR-Cas9 Circular Permutants as Programmable Scaffolds for Genome Modification," *Cell*,

January 10, 2019, 176: 254-267, each of are incorporated herein by reference. The instant disclosure contemplates any previously known CP-Cas9 or use a new CP-Cas9 so long as the resulting circularly permuted protein retains the ability to bind DNA when complexed with a guide RNA (gRNA).

[00107] Any of the Cas9 proteins described herein, including any variant, ortholog, or naturally occurring Cas9 or equivalent thereof, may be reconfigured as a circular permutant variant.

[00108] In various embodiments, the circular permutants of Cas9 may have the following structure: N-terminus-[original C-terminus] – [optional linker] – [original N-terminus]-C-terminus.

[00109] As an example, the present disclosure contemplates the following circular permutants of *S. pyogenes* Cas9 (based on 1368 amino acids of UniProtKB - Q99ZW2 (CAS9_STRP1) of SEQ ID NO: 6:

[00110] N-terminus-[1268-1368]-[optional linker]-[1-1267]-C-terminus;

[00111] N-terminus-[1168-1368]-[optional linker]-[1-1167]-C-terminus;

[00112] N-terminus-[1068-1368]-[optional linker]-[1-1067]-C-terminus;

[00113] N-terminus-[968-1368]-[optional linker]-[1-967]-C-terminus;

[00114] N-terminus-[868-1368]-[optional linker]-[1-867]-C-terminus;

[00115] N-terminus-[768-1368]-[optional linker]-[1-767]-C-terminus;

[00116] N-terminus-[668-1368]-[optional linker]-[1-667]-C-terminus;

[00117] N-terminus-[568-1368]-[optional linker]-[1-567]-C-terminus;

[00118] N-terminus-[468-1368]-[optional linker]-[1-467]-C-terminus;

[00119] N-terminus-[368-1368]-[optional linker]-[1-367]-C-terminus;

[00120] N-terminus-[268-1368]-[optional linker]-[1-267]-C-terminus;

[00121] N-terminus-[168-1368]-[optional linker]-[1-167]-C-terminus;

[00122] N-terminus-[68-1368]-[optional linker]-[1-67]-C-terminus; or

[00123] N-terminus-[10-1368]-[optional linker]-[1-9]-C-terminus, or the corresponding circular permutants of other Cas9 proteins (including other Cas9 orthologs, variants, etc).

[00124] In particular embodiments, the circular permuant Cas9 has the following structure (based on *S. pyogenes* Cas9 (SEQ ID NO: 6):

[00125] N-terminus-[102-1368]-[optional linker]-[1-101]-C-terminus;

[00126] N-terminus-[1028-1368]-[optional linker]-[1-1027]-C-terminus;

[00127] N-terminus-[1041-1368]-[optional linker]-[1-1043]-C-terminus;

[00128] N-terminus-[1249-1368]-[optional linker]-[1-1248]-C-terminus; or

[00129] N-terminus-[1300-1368]-[optional linker]-[1-1299]-C-terminus, or the corresponding circular permutants of other Cas9 proteins (including other Cas9 orthologs, variants, etc).

[00130] In still other embodiments, the circular permutant Cas9 has the following structure (based on *S. pyogenes* Cas9 (SEQ ID NO: 6):

[00131] N-terminus-[103-1368]-[optional linker]-[1-102]-C-terminus;

[00132] N-terminus-[1029-1368]-[optional linker]-[1-1028]-C-terminus;

[00133] N-terminus-[1042-1368]-[optional linker]-[1-1041]-C-terminus;

[00134] N-terminus-[1250-1368]-[optional linker]-[1-1249]-C-terminus; or

[00135] N-terminus-[1301-1368]-[optional linker]-[1-1300]-C-terminus, or the corresponding circular permutants of other Cas9 proteins (including other Cas9 orthologs, variants, etc).

[00136] In some embodiments, the Cas9 variant comprises a fragment of Cas9 (e.g., a gRNA binding domain or a DNA-cleavage domain), such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to the corresponding fragment of wild type Cas9. In some embodiments, the fragment is is at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% identical, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid length of a corresponding wild type Cas9. In some embodiments,

the Cas9 fragment is at least 100 amino acids in length. In some embodiments, the fragment is at least 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, or at least 1300 amino acids in length. In some embodiments, the fragment binds crRNA and tracrRNA or sgRNA, but does not comprise a functional nuclease domain, e.g., in that it comprises only a truncated version of a nuclease domain or no nuclease domain at all.

[00137] In some embodiments, the circular permutant can be formed by linking a C-terminal fragment of a Cas9 to an N-terminal fragment of a Cas9, either directly or by using a linker, such as an amino acid linker. In some embodiments, The C-terminal fragment may correspond to the C-terminal 95% or more of the amino acids of a Cas9 (e.g., amino acids about 1300-1368), or the C-terminal 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, or 5% or more of a Cas9 (e.g., SEQ ID NO: 6). The N-terminal portion may correspond to the N-terminal 95% or more of the amino acids of a Cas9 (e.g., amino acids about 1-1300), or the N-terminal 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10%, or 5% or more of a Cas9 (e.g., of SEQ ID NO: 6).

[00138] In some embodiments, the circular permutant can be formed by linking a C-terminal fragment of a Cas9 to an N-terminal fragment of a Cas9, either directly or by using a linker, such as an amino acid linker. In some embodiments, the C-terminal fragment that is rearranged to the N-terminus, includes or corresponds to the C-terminal 30% or less of the amino acids of a Cas9 (*e.g.,* amino acids 1012-1368 of SEQ ID NO: 6). In some embodiments, the C-terminal fragment that is rearranged to the N-terminus, includes or corresponds to the C-terminal 30%, 29%, 28%, 27%, 26%, 25%, 24%, 23%, 22%, 21%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% of the amino acids of a Cas9 (*e.g.,* the Cas9 of SEQ ID NO: 6). In some embodiments, the C-terminal fragment that is rearranged to the N-terminus, includes or corresponds to the C-terminal 410 residues or less of a Cas9 (*e.g.,* the Cas9 of SEQ ID NO: 6). In some embodiments, the C-terminal portion that is rearranged to the N-terminus, includes or

corresponds to the C-terminal 410, 400, 390, 380, 370, 360, 350, 340, 330, 320, 310, 300, 290, 280, 270, 260, 250, 240, 230, 220, 210, 200, 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 20, or 10 residues of a Cas9 (*e.g.,* the Cas9 of SEQ ID NO: 6). In some embodiments, the C-terminal portion that is rearranged to the N-terminus, includes or corresponds to the C-terminal 357, 341, 328, 120, or 69 residues of a Cas9 (e.g., the Cas9 of SEQ ID NO: 6).

[00139] In other embodiments, circular permutant Cas9 variants may be defined as a topological rearrangement of a Cas9 primary structure based on the following method, which is based on *S. pyogenes* Cas9 of SEQ ID NO: 6: (a) selecting a circular permutant (CP) site corresponding to an internal amino acid residue of the Cas9 primary structure, which dissects the original protein into two halves: an N-terminal region and a C-terminal region; (b) modifying the Cas9 protein sequence (e.g., by genetic engineering techniques) by moving the original C-terminal region (comprising the CP site amino acid) to preceed the original N-terminal region, thereby forming a new N-terminus of the Cas9 protein that now begins with the CP site amino acid residue. The CP site can be located in any domain of the Cas9 protein, including, for example, the helical-II domain, the RuvCIII domain, or the CTD domain. For example, the CP site may be located (relative to the *S. pyogenes* Cas9 of SEQ ID NO: 6) at original amino acid residue 181, 199, 230, 270, 310, 1010, 1016, 1023, 1029, 1041, 1247, 1249, or 1282. Thus, once relocated to the N-terminus, original amino acid 181, 199, 230, 270, 310, 1010, 1016, 1023, 1029, 1041, 1247, 1249, or 1282 would become the new N-terminal amino acid. Nomenclature of these CP-Cas9 proteins may be referred to as Cas9-CP^{181}, Cas9-CP^{199}, Cas9-CP^{230}, Cas9-CP^{270}, Cas9-CP^{310}, Cas9-CP^{1010}, Cas9-CP^{1016}, Cas9-CP^{1023}, Cas9-CP^{1029}, Cas9-CP^{1041}, Cas9-CP^{1247}, Cas9-CP^{1249}, and Cas9-CP^{1282}, respectively. This description is not meant to be limited to making CP variants from SEQ ID NO: 6, but may be implemented to make CP variants in any Cas9 sequence, either at CP sites that correspond to these positions, or at other CP sites entirely. This description is not meant to limit the specific CP sites in any way. Virtually any CP site may be used to form a CP-Cas9 variant.

[00140] Exemplary CP-Cas9 amino acid sequences, based on the Cas9 of SEQ ID NO: 6, are provided below in which linker sequences are indicated by underlining and optional methionine (M) residues are indicated in bold. It should be appreciated that the disclosure provides CP-Cas9 sequences that do not include a linker sequence or that include different linker sequences. It should be appreciated that CP-Cas9 sequences may be based on Cas9 sequences other than that of SEQ ID NO: 6 and any examples provided herein are not meant to be limiting.

[00141] CP1012

DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD
FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA
SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS
ITGLYETRIDLSQLGGD<u>GGSGGSGGSGGSGGSGGSGG</u>DKKYSIGLAIGTNSVGWAVITDEYKVPSKKF
KVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR
LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHF
LIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNG
LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSD
ILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKI
EKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD
SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFD
DKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ
VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGF
IKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHA
HDAYLNAVVGTALIKKYPKLESEFVYG (SEQ ID NO: 12)

[00142] CP1028

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV
KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK
YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD
KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG

DGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNL
IGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHER
HPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDK
LFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPN
FKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTE
ELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL
GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGW
GRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL
AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKN
RGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV
AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK
KYPKLESEFVYGDYKVYDVRKMIAKSEQ (SEQ ID NO: 13)

[00143] CP1041

NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK
NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHL
FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGSGGS
GGSGGSGGDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEA
TRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYH
EKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE
ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLT
LLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK
QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKS
EETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKD
FLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK
QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV
KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNE
KLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVK
KMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE
NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYS (SEQ ID NO: 14)

[00144] CP1249

PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT
NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGSGGSGGSG
GSGGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK
DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK
ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT
FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET
ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL
SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD
NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV
DELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMK
NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK
LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVY
DVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVR
KVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG
KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGEL
QKGNELALPSKYVNFLYLASHYEKLKGS (SEQ ID NO: 15)

[00145] CP1300

KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG
DGGSGGSGGSGGSGGSGGSGGDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI
GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERH
PIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKL
FIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNF
KSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS
MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE
LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLA
RGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLG
TYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA
GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL
KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR

GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA
QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKK
YPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGE
TGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS
PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF
ELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRD (SEQ ID NO: 16)

[00146] Exemplary C-terminal fragments of Cas9, based on the Cas9 of SEQ ID NO: 6, which may be rearranged to an N-terminus of Cas9, are provided below. It should be appreciated that such C-terminal fragments of Cas9 are exemplary and are not meant to be limiting.

[00147] CP1012

DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD
FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVA
KVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLA
SAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQS
ITGLYETRIDLSQLGGD (SEQ ID NO: 17).

[00148] CP1028

EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV
KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKEL
LGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK
YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD
KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG
D (SEQ ID NO: 18).

[00149] CP1041

NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKE
SILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK
NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHL
FTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO:
19).

[00150] CP1249

PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLT
NLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO:
20).

[00151] CP1300

KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG
D (SEQ ID NO: 21).

Cas9 domains

[00152] Some aspects of this disclosure provide Cas9 proteins that exhibit activity on a target sequence that does not comprise the canonical PAM (5′-NGG-3′, where N is A, C, G, or T) at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NGG-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NNG-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NNA-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NNC-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NNT-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NGT-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NGA-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NGC-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NAA-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NAC-3′ PAM sequence at its 3′-end. In some embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NAT-3′ PAM sequence at its 3′-end. In still other embodiments, the Cas9 protein exhibits activity on a target sequence comprising a 5′-NAG-3′ PAM sequence at its 3′-end.

[00153] It should be appreciated that any of the amino acid mutations described herein, (e.g., A262T) from a first amino acid residue (e.g., A) to a second amino acid residue (e.g., T) may also include mutations from the first amino acid residue to an amino acid residue that is similar to (e.g., conserved) the second amino acid residue. For example, mutation of an amino acid with a hydrophobic side chain

(e.g., alanine, valine, isoleucine, leucine, methionine, phenylalanine, tyrosine, or tryptophan) may be a mutation to a second amino acid with a different hydrophobic side chain (e.g., alanine, valine, isoleucine, leucine, methionine, phenylalanine, tyrosine, or tryptophan). For example, a mutation of an alanine to a threonine (e.g., a A262T mutation) may also be a mutation from an alanine to an amino acid that is similar in size and chemical properties to a threonine, for example, serine. As another example, mutation of an amino acid with a positively charged side chain (e.g., arginine, histidine, or lysine) may be a mutation to a second amino acid with a different positively charged side chain (e.g., arginine, histidine, or lysine). As another example, mutation of an amino acid with a polar side chain (e.g., serine, threonine, asparagine, or glutamine) may be a mutation to a second amino acid with a different polar side chain (e.g., serine, threonine, asparagine, or glutamine). Additional similar amino acid pairs include, but are not limited to, the following: phenylalanine and tyrosine; asparagine and glutamine; methionine and cysteine; aspartic acid and glutamic acid; and arginine and lysine. The skilled artisan would recognize that such conservative amino acid substitutions will likely have minor effects on protein structure and are likely to be well tolerated without compromising function. In some embodiments, any amino of the amino acid mutations provided herein from one amino acid to a threonine may be an amino acid mutation to a serine. In some embodiments, any amino of the amino acid mutations provided herein from one amino acid to an arginine may be an amino acid mutation to a lysine. In some embodiments, any amino of the amino acid mutations provided herein from one amino acid to an isoleucine, may be an amino acid mutation to an alanine, valine, methionine, or leucine. In some embodiments, any amino of the amino acid mutations provided herein from one amino acid to a lysine may be an amino acid mutation to an arginine. In some embodiments, any amino of the amino acid mutations provided herein from one amino acid to an aspartic acid may be an amino acid mutation to a glutamic acid or asparagine. In some embodiments, any amino of the amino acid mutations provided herein from one amino acid to a valine may be an amino acid mutation to an alanine, isoleucine, methionine, or leucine. In some embodiments, any amino of the amino acid

mutations provided herein from one amino acid to a glycine may be an amino acid mutation to an alanine. It should be appreciated, however, that additional conserved amino acid residues would be recognized by the skilled artisan and any of the amino acid mutations to other conserved amino acid residues are also within the scope of this disclosure.

Mutations in Wild-Type SpCas9

[00154] Some aspects of this disclosure provide a Cas9 protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of Cas9 as provided by any of the sequences set forth in SEQ ID NO: 2, 4, or 6-11, wherein the amino acid sequence of the Cas9 protein comprises at least one mutation in an amino acid residue selected from the group consisting of amino acid residues 10, 177, 218, 322, 367, 409, 427, 589, 599, 614, 630, 631, 654, 673, 693, 710, 715, 727, 743, 753, 757, 758, 762, 763, 768, 803, 859, 861, 865, 869, 921, 946, 1016, 1021, 1028, 1054, 1077, 1080, 1114, 1134, 1135, 1137, 1139, 1151, 1180, 1188, 1211, 1219, 1221, 1223, 1256, 1264, 1274, 1290, 1318, 1317, 1320, 1323, and 1333 of *S. pyogenes* having the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding amino acid residue in any of the amino acid sequences provided in SEQ ID NO: 2. In some embodiments, the Cas9 protein comprises a RuvC and an HNH domain. In some embodiments, the amino acid sequence of the Cas9 protein is not identical to the amino acid sequence of a naturally occurring Cas9 domain. In some embodiments, the Cas9 protein is a nuclease-inactive Cas9 protein. In some embodiments, the Cas9 domain is a Cas9 nickase. In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one mutation selected from the group consisting of X10T, X177N, X218R, X322V, X367T, X409I, X427G, X589S, X599R, X614N, X630K, X631A, X654L, X673E, X693L, X710E, X715C, X727I, X743I, X753G, X757K, X758H, X762G, X763I, X768H, X803S, X859S, X861N, X865G, X869S, X921P, X946D, X1016D, X1021T, X1028D, X1054D, X1077D, X1080S, X1114G, X1134L, X1135N, X1137S, X1139A, X1151E, X1180G, X1188R, X1211R, X1219V, X1221H, X1223S, X1256R, X1264Y, X1274R, X1290G, X1318S, X1317T, X1320V, X1323D, and X1333K of the amino acid sequence provided in

SEQ ID NO: 2, or in a corresponding mutation, or mutations, in any of the amino acid sequences provided in SEQ ID NO: 2, wherein X represents any amino acid. In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one mutation selected from group consisting of A10T, D177N, K218R, I322V, A367T, S409I, E427G, A589S, K599R, D614N, E630K, M631A, R654L, K673E, F693L, K710E, G715C, L727I, V743I, R753G, E757K, N758H, E762G, M763I, Q768H, N803S, R859S, D861N, G865G, N869S, L921P, N946D, Y1016D, M1021T, E1028D, V1139A, N1054D, G1077D, F1080S, R1114G, F1134L, D1135N, P1137S, K1151E, D1180G, K1188R, K1211R, E1219V, Q1221H, G1223S, Q1256R, H1264Y, S1274R, V1290G, L1318S, N1317T, A1320V, A1323D, and R1333K of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in any of the amino acid sequences provided in SEQ ID NO: 2, 4, or 6-11, wherein X represents any amino acid.

[00155] Some aspects of this disclosure provide a Cas9 protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of Cas9 as provided by SEQ ID NO: 2, wherein the amino acid sequence of the Cas9 protein comprises at least one mutation in an amino acid residue selected from the group consisting of amino acid residues 472, 562, 565, 570, 589, 608, 625, 627, 629, 630, 631, 638, 647, 652, 653, 654, 670, 673, 676, 687, 703, 710, 711, 716, 740, 742, 752, 753, 767, 771, 775, 789, 790, 795, 797, 803, 804, 808, 848, 866, 875, 890, 922, 928, 948, 959, 990, 995, 1014, 1015, 1016, 1021, 1030, 1036, 1055, 1057, 1114, 1127, 1135, 1156, 1177, 1180, 1184, 1207, 1219, 1234, 1246, 1251, 1252, 1286, 1301, 1332, 1335, 1337, 1338, 1348, 1349, 1365, 1367, and 1368 of *S. pyogenes* having the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding amino acid residue in any of the amino acid sequences provided in SEQ ID NO: 2, 4, or 6-11. In some embodiments, the Cas9 protein comprises a RuvC and an HNH domain.

[00156] In some embodiments, the amino acid sequence of the Cas9 protein is not identical to the amino acid sequence of a naturally occurring Cas9 protein. In some embodiments, the Cas9 protein is a nuclease-inactive Cas9 protein. In some embodiments, the Cas9 protein is a Cas9 nickase. In some

embodiments, the amino acid sequence of the Cas9 protein comprises at least one mutation selected from the group consisting of X472I, X562F, X565D, X570T, X570S, X589V, X608R, X625S, X627K, X629G, X630G, X631I, X638P, X647A, X647I, X652T, X653K, X654L, X654I, X654H, X670T, X673E, X676G, X687R, X703P, X710E, X711T, X716R, X740A, X742E, X752R, X753G, X767D, X771H, X775R, X789E, X790A, X795L, X797N, X803S, X804A, X808D, X848N, X866R, X875I, X890E, X890N, X922A, X928T, X948E, X959N, X990S, X995S, X1014N, X1015A, X1016C, X1016S, X1021L, X1030R, X1036H, X1036D, X1055E, X1057S, X1057T, X1114G, X1127A, X1135N, X1156E, X1156N, X1177S, X1180E, X1184T, X1207G, X1219V, X1234D, X1246E, X1251G, X1252D, X1286H, X1301S, X1332N, X1332G, X1335Q, X1337N, X1338T, X1348V, X1349R, X1365L, X1367E, X1367T, X1367fs?, and X1368D of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in any of the amino acid sequences provided in SEQ ID NOs: 2, 4, or 6-11, wherein X represents any amino acid. In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one mutation selected from group consisting of T472I, I562F, V565D, I570T, I570S, A589V, K608R, L625S, E627K, R629G, E630G, M631I, T638P, V647A, V647I, K652T, R653K, R654L, R654I, R654H, I670T, K673E, G676G, G687R, T703P, K710E, A711T, Q716R, T740A, K742E, G752R, R753G, N767D, Q771H, K775R, K789E, E790A, I795L, K797N, N803S, T804A, N808D, K848N, K866R, V875I, K890E, K890N, V922A, K948E, K959N, N990S, T995S, K1014N, V1015A, Y1016C, Y1016S, M1021L, G1030R, Y1036H, Y1036D, I1055E, I1057S, I1057T, R1114G, D1127A, D1135N, K1156E, K1156N, N1177S, D1180E, A1184T, E1207G, E1219V, N1234D, K1246E, D1251G, N1252D, N1286H, P1301S, D1332N, D1332G, R1335Q, T1337N, S1338T, I1348V, H1349R, L1365L, G1367E, G1367T, and D1368D of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in SEQ ID NO: 2, 4, or 6-11, wherein X represents any amino acid.

[00157] Some aspects of this disclosure provide a Cas9 protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of Cas9 as provided by SEQ ID NOs: 2, wherein the amino acid sequence of the Cas9 protein comprises at least one mutation in an amino acid residue selected from the group consisting of amino acid residues 575, 596, 631, 649, 654, 664, 710, 740, 743, 748, 750, 753, 765, 790, 797, 853, 922, 955, 961, 985, 1012, 1049, 1057, 1114, 1131, 1135, 1150, 1156, 1162, 1180, 1191, 1218, 1219, 1221, 1227, 1249, 1253, 1256, 1286, 1293, 1308, 1317, 1320, 1321, 1332, 1335, and 1339 of *S. pyogenes* having the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding amino acid residue in another Cas9 sequence (e.g., any of the sequences of 2, 4, or 6-11). In some embodiments, the Cas9 protein comprises a RuvC and an HNH domain. In some embodiments, the amino acid sequence of the Cas9 protein is not identical to the amino acid sequence of a naturally occurring Cas9 protein. In some embodiments, the Cas9 protein is a nuclease-inactive Cas9 domain. In some embodiments, the Cas9 protein is a Cas9 nickase. In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one mutation selected from the group consisting of X575S, X596Y, X631L, X649R, X654L, X664K, X710E, X740A, X743I, X748I, X750A, X753G, X765X, X790A, X797E, X853E, X922A, X955L, X961E, X985Y, X1012A, X1049G, X1057V, X1114G, X1131C, X1135N, X1150V, X1156E, X1162A, X1180G, X1180A, X1191N, X1218S, X1219V, X1221H, X1227V, X1249S, X1253K, X1256R, X1286K, X1293T, X1308D, X1317K, X1320V, X1321S, X1332G, X1335L, and X1339I of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid. In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one mutation selected from group consisting of F575S, D596Y, M631L, K649R, R654L, R664K, K710E, T740A, V743I, V748I, V750A, R753G, R765X, E790A, K797E, D853E, V922A, V955L, K961E, H985Y, D1012A, E1049G, I1057V, R1114G, Y1131C, D1135N, E1150V, K1156E, E1162A, D1180G, D1180A, K1191N, G1218S, E1219V, Q1221H, A1227V, P1249S, E1253K, Q1256R, N1286K,

A1293T, N1308D, N1317K, A1320V, P1321S, D1332G, R1335L, and T1339I of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00158] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, at least forty-three, at least forty-four, at least forty-five, at least forty-six, at least forty-seven, at least forty-eight, at least forty-nine, at least fifty, at least fifty-one, at least fifty-two, at least fifty-three, at least fifty-four, at least fifty-five, at least fifty-six, at least fifty-seven, at least fifty-eight, or at least fifty-nine mutations in amino acid residues selected from the group consisting of 10, 177, 218, 322, 367, 409, 427, 589, 599, 614, 630, 631, 654, 673, 693, 710, 715, 727, 743, 753, 757, 758, 762, 763, 768, 803, 859, 861, 865, 869, 921, 946, 1016, 1021, 1028, 1054, 1077, 1080, 1114, 1134, 1135, 1137, 1139, 1151, 1180, 1188, 1211, 1219, 1221, 1223, 1256, 1264, 1274, 1290, 1318, 1317, 1320, 1323, and 1333 of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00159] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at

least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, at least forty-three, at least forty-four, at least forty-five, at least forty-six, at least forty-seven, at least forty-eight, at least forty-nine, at least fifty, at least fifty-one, at least fifty-two, at least fifty-three, at least fifty-four, at least fifty-five, at least fifty-six, at least fifty-seven, at least fifty-eight, or at least fifty-nine mutations selected from the group consisting of X10T, X177N, X218R, X322V, X367T, X409I, X427G, X589S, X599R, X614N, X630K, X631A, X654L, X673E, X693L, X710E, X715C, X727I, X743I, X753G, X757K, X758H, X762G, X763I, X768H, X803S, X859S, X861N, X865G, X869S, X921P, X946D, X1016D, X1021T, X1028D, X1054D, X1077D, X1080S, X1114G, X1134L, X1135N, X1137S, X1139A, X1151E, X1180G, X1188R, X1211R, X1219V, X1221H, X1223S, X1256R, X1264Y, X1274R, X1290G, X1318S, X1317T, X1320V, X1323D, and X1333K of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00160] In some embodiments, the present disclosure may utilize any of the Cas9 variants disclosed in the SEQUENCES section herein.

[00161] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at

least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, at least forty-three, at least forty-four, at least forty-five, at least forty-six, at least forty-seven, at least forty-eight, at least forty-nine, at least fifty, at least fifty-one, at least fifty-two, at least fifty-three, at least fifty-four, at least fifty-five, at least fifty-six, at least fifty-seven, at least fifty-eight, or at least fifty-nine mutations selected from the group consisting of A10T, D177N, K218R, I322V, A367T, S409I, E427G, A589S, K599R, D614N, E630K, M631A, R654L, K673E, F693L, K710E, G715C, L727I, V743I, R753G, E757K, N758H, E762G, M763I, Q768H, N803S, R859S, D861N, G865G, N869S, L921P, N946D, Y1016D, M1021T, E1028D, N1054D, G1077D, F1080S, R1114G, F1134L, D1135N, P1137S, V1139A, K1151E, D1180G, K1188R, K1211R, E1219V, Q1221H, G1223S, Q1256R, H1264Y, S1274R, V1290G, L1318S, N1317T, A1320V, A1323D, and R1333K of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00162] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, at least forty-three, at least forty-four, at least forty-five, at least forty-six, at least forty-seven, or at least forty-eight mutations selected from the group consisting of 10, 177, 218, 322, 367, 427, 589, 599, 614, 630, 631, 693, 710, 743, 753, 757, 758, 762, 768, 803, 859, 861, 865, 869, 921, 946, 1016,

1021, 1028, 1054, 1077, 1080, 1114, 1134, 1135, 1137, 1151, 1180, 1188, 1211, 1221, 1223, 1274, 1290, 1317, 1320, 1323, and 1333 of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00163] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, at least forty-three, at least forty-four, at least forty-five, at least forty-six, at least forty-seven, or at least forty-eight mutations selected from the group consisting of X10T, X177N, X218R, X322V, X367T, X427G, X589S, X599R, X614N, X630K, X631A, X693L, X710E, X743I, X753G, X757K, X758H, X762G, X768H, X803S, X859S, X861N, X865G, X869S, X921P, X946D, X1016D, X1021T, X1028D, X1054D, X1077D, X1080S, X1114G, X1134L, X1135N, X1137S, X1151E, X1180G, X1188R, X1211R, X1221H, X1223S, X1274R, X1290G, X1317T, X1320V, X1323D, and X1333K of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00164] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one,

at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, at least forty-three, at least forty-four, at least forty-five, at least forty-six, at least forty-seven, or at least forty-eight mutations selected from the group consisting of A10T, D177N, K218R, I322V, A367T, E427G, A589S, K599R, D614N, E630K, M631A, F693L, K710E, V743I, R753G, E757K, N758H, E762G, Q768H, N803S, R859S, D861N, N869S, L921P, N946D, Y1016D, M1021T, Q1028D, N1054D, G1077D, F1080S, R1114G, F1134L, D1135N, P1137S, K1151E, D1180G, K1188R, K1211R, Q1221H, G1223S, S1274R, V1290G, N1317T, A1320V, A1323D, and R1333K of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00165] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, at least forty-three, at least forty-four, at least forty-five, at least forty-six, at least forty-seven, at least forty-eight, at least forty-nine, at least fifty, at least fifty-one, at least fifty-two, at least fifty-three, at least fifty-four, at least fifty-five, at least fifty-six, at least fifty-seven, at least fifty-eight, at least fifty-nine, at least sixty, at least sixty-one, at least sixty-two, at least sixty-three, at least sixty-four, at least

sixty-five, at least sixty-six, at least sixty-seven, at least sixty-eight, at least sixty-nine, at least seventy, at least seventy-one, at least seventy-two, at least seventy-three, at least seventy-four, at least seventy-five, at least seventy-six, at least seventy-seven, at least seventy-eight, at least seventy-nine, or at least eighty mutations selected from the group consisting of 472, 562, 565, 570, 589, 608, 625, 627, 629, 630, 631, 638, 647, 647, 652, 653, 654, 670, 673, 676, 687, 703, 710, 711, 716, 740, 742, 752, 753, 767, 771, 775, 789, 790, 795, 797, 803, 804, 808, 848, 866, 875, 890, 922, 928, 948, 959, 990, 995, 1014, 1015, 1016, 1021, 1030, 1036, 1055, 1057, 1114, 1127, 1135, 1156, 1177, 1180, 1184, 1207, 1219, 1234, 1246, 1251, 1252, 1286, 1301, 1332, 1335, 1337, 1338, 1348, 1349, 1365, 1367, and 1368 of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00166] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, at least forty-three, at least forty-four, at least forty-five, at least forty-six, at least forty-seven, at least forty-eight, at least forty-nine, at least fifty, at least fifty-one, at least fifty-two, at least fifty-three, at least fifty-four, at least fifty-five, at least fifty-six, at least fifty-seven, at least fifty-eight, at least fifty-nine, at least sixty, at least sixty-one, at least sixty-two, at least sixty-three, at least sixty-four, at least sixty-five, at least sixty-six, at least sixty-seven, at least sixty-eight, at least sixty-nine, at least seventy, at least seventy-one, at least seventy-two, at least seventy-three, at least seventy-four, at least

seventy-five, at least seventy-six, at least seventy-seven, at least seventy-eight, at least seventy-nine, or at least eighty mutations selected from the group consisting of X472I, X562F, X565D, X570T, X570S, X589V, X608R, X625S, X627K, X629G, X630G, X631I, X638P, X647A, X647I, X652T, X653K, X654L, X654I, X654H, X670T, X673E, X676G, X687R, X703P, X710E, X711T, X716R, X740A, X742E, X752R, X753G, X767D, X771H, X775R, X789E, X790A, X795L, X797N, X803S, X804A, X808D, X848N, X866R, X875I, X890E, X890N, X922A, X928T, X948E, X959N, X990S, X995S, X1014N, X1015A, X1016C, X1016S, X1021L, X1030R, X1036H, X1036D, X1055E, X1057S, X1057T, X1114G, X1127A, X1135N, X1156E, X1156N, X1177S, X1180E, X1184T, X1207G, X1219V, X1234D, X1246E, X1251G, X1252D, X1286H, X1301S, X1332N, X1332G, X1335Q, X1337N, X1338T, X1348V, X1349R, X1365L, X1367E, X1367T, X1367fs?, and X1368D of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00167] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, at least forty-three, at least forty-four, at least forty-five, at least forty-six, at least forty-seven, at least forty-eight, at least forty-nine, at least fifty, at least fifty-one, at least fifty-two, at least fifty-three, at least fifty-four, at least fifty-five, at least fifty-six, at least fifty-seven, at least fifty-eight, at least fifty-nine, at least sixty, at least sixty-one, at least sixty-two, at least sixty-three, at least sixty-four, at least

sixty-five, at least sixty-six, at least sixty-seven, at least sixty-eight, at least sixty-nine, at least seventy, at least seventy-one, at least seventy-two, at least seventy-three, at least seventy-four, at least seventy-five, at least seventy-six, at least seventy-seven, at least seventy-eight, at least seventy-nine, at least eighty mutations selected from the group consisting of T472I, I562F, V565D, I570T, I570S, A589V, K608R, L625S, E627K, R629G, E630G, M631I, T638P, V647A, V647I, K652T, R653K, R654L, R654I, R654H, I670T, K673E, G676G, G687R, T703P, K710E, A711T, Q716R, T740A, K742E, G752R, R753G, N767D, Q771H, K775R, K789E, E790A, I795L, K797N, N803S, T804A, N808D, K848N, K866R, V875I, K890E, K890N, V922A, K948E, K959N, N990S, T995S, K1014N, V1015A, Y1016C, Y1016S, M1021L, G1030R, Y1036H, Y1036D, I1055E, I1057S, I1057T, R1114G, D1127A, D1135N, K1156E, K1156N, N1177S, D1180E, A1184T, E1207G, E1219V, N1234D, K1246E, D1251G, N1252D, N1286H, P1301S, D1332N, D1332G, R1335Q, T1337N, S1338T, I1348V, H1349R, L1365L, G1367E, G1367T, and D1368D of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00168] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, at least forty-three, at least forty-four, at least forty-five, at least forty-six, at least forty-seven, at least forty-eight, at least forty-nine, at least fifty, at least fifty-one, at least fifty-two, at least fifty-three, at least fifty-four, at least fifty-five, at least fifty-six, at least fifty-seven, at least fifty-eight, at least fifty-

nine, at least sixty, at least sixty-one, at least sixty-two, at least sixty-three, at least sixty-four, at least sixty-five, at least sixty-six, at least sixty-seven, at least sixty-eight, at least sixty-nine, at least seventy, at least seventy-one, or at least seventy-two mutations in amino acid residues selected from the group consisting of 472, 562, 565, 570, 589, 608, 625, 627, 629, 630, 631, 638, 647, 647, 652, 653, 654, 676, 687, 703, 716, 740, 742, 752, 753, 767, 771, 775, 789, 790, 795, 797, 803, 804, 808, 848, 866, 875, 890, 890, 922, 948, 959, 990, 995, 1014, 1015, 1016, 1021, 1030, 1036, 1055, 1057, 1057, 1114, 1127, 1135, 1156, 1177, 1180, 1184, 1234, 1246, 1251, 1252, 1286, 1301, 1332, 1335, 1348, 1367, and 1368 of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00169] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, at least forty-three, at least forty-four, at least forty-five, at least forty-six, at least forty-seven, at least forty-eight, at least forty-nine, at least fifty, at least fifty-one, at least fifty-two, at least fifty-three, at least fifty-four, at least fifty-five, at least fifty-six, at least fifty-seven, at least fifty-eight, at least fifty-nine, at least sixty, at least sixty-one, at least sixty-two, at least sixty-three, at least sixty-four, at least sixty-five, at least sixty-six, at least sixty-seven, at least sixty-eight, at least sixty-nine, at least seventy, at least seventy-one, or at least seventy-two mutations selected from the group consisting of X472I, X562F, X565D, X570T, X570S, X589V, X608R, X625S, X627K, X629G, X630G, X631I,

X631V, X638P, X647A, X647I, X652T, X653K, X654L, X654I, X654H, X670T, X676G, X687R, X703P, X710E, X716R, X740A, X742E, X752R, X753G, X767D, X771H, X775R, X789E, X790A, X795L, X797N, X803S, X804A, X808D, X848N, X866R, X875I, X890E, X890N, X922A, X948E, X959N, X990S, X995S, X1014N, X1015A, X1016C, X1016S, X1021L, X1030R, X1036H, X1036D, X1055E, X1057S, X1057T, X1114G, X1127A, X1127G, X1135N, X1156E, X1156N, X1177S, X1180E, X1184T, X1234D, X1246E, X1251G, X1252D, X1286H, X1301S, X1332N, X1332G, X1335Q, X1338T, X1348V, X1349R, X1367E, X1367T, X1367fs?, and X1368D of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00170] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, at least forty-three, at least forty-four, at least forty-five, at least forty-six, at least forty-seven, at least forty-eight, at least forty-nine, at least fifty, at least fifty-one, at least fifty-two, at least fifty-three, at least fifty-four, at least fifty-five, at least fifty-six, at least fifty-seven, at least fifty-eight, at least fifty-nine, at least sixty, at least sixty-one, at least sixty-two, at least sixty-three, at least sixty-four, at least sixty-five, at least sixty-six, at least sixty-seven, at least sixty-eight, at least sixty-nine, at least seventy, at least seventy-one, or at least seventy-two mutations selected from the group consisting of T472I, I562F, V565D, I570T, I570S, A589V, K608R, L625S, E627K, R629G, E630G, M631I,

M631V, T638P, V647A, V647I, K652T, R653K, R654L, R654I, R654H, I670T, G676G, G687R, T703P, K710E, Q716R, T740A, K742E, G752R, R753G, N767D, Q771H, K775R, K789E, E790A, I795L, K797N, N803S, T804A, N808D, K848N, K866R, V875I, K890E, K890N, V922A, K948E, K959N, N990S, T995S, K1014N, V1015A, Y1016C, Y1016S, M1021L, G1030R, Y1036H, Y1036D, I1055E, I1057S, I1057T, R1114G, D1127A, D1127G, D1135N, K1156E, K1156N, N1177S, D1180E, A1184T, N1234D, K1246E, D1251G, N1252D, N1286H, P1301S, D1332N, D1332G, R1335Q, S1338T, I1348V, S1349R, G1367E, G1367T, G1367fs?, and D1368D of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00171] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, at least forty-three, at least forty-four, at least forty-five, at least forty-six, at least forty-seven, or at least forty-eight mutations selected from the group consisting of 575, 596, 631, 649, 654, 664, 710, 740, 743, 748, 750, 753, 765, 790, 797, 853, 922, 955, 961, 985, 1012, 1049, 1057, 1114, 1131, 1135, 1150, 1156, 1162, 1180, 1191, 1218, 1219, 1221, 1227, 1249, 1253, 1256, 1286, 1293, 1308, 1317, 1320, 1321, 1332, 1335, and 1339 of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00172] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, at least forty-three, at least forty-four, at least forty-five, at least forty-six, at least forty-seven, or at least forty-eight mutations selected from the group consisting of X575S, X596Y, X631L, X649R, X654L, X664K, X710E, X740A, X743I, X748I, X750A, X753G, X765X, X790A, X797E, X853E, X922A, X955L, X961E, X985Y, X1012A, X1049G, X1057V, X1114G, X1131C, X1135N, X1150V, X1156E, X1162A, X1180G, X1180A, X1191N, X1218S, X1219V, X1221H, X1227V, X1249S, X1253K, X1256R, X1286K, X1293T, X1308D, X1317K, X1320V, X1321S, X1332G, X1335L, and X1339I of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00173] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two,

at least forty-three, at least forty-four, at least forty-five, at least forty-six, at least forty-seven, or at least forty-eight mutations selected from the group consisting of F575S, D596Y, M631L, K649R, R654L, R664K, K710E, T740A, V743I, V748I, V750A, R753G, R765X, E790A, K797E, D853E, V922A, V955L, K961E, H985Y, D1012A, E1049G, I1057V, R1114G, Y1131C, D1135N, E1150V, K1156E, E1162A, D1180G, D1180A, K1191N, G1218S, E1219V, Q1221H, A1227V, P1249S, E1253K, Q1256R, N1286K, A1293T, N1308D, N1317K, A1320V, P1321S, D1332G, R1335L, and T1339I of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00174] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, or at least forty-three mutations selected from the group consisting of 575, 596, 631, 649, 664, 710, 740, 743, 748, 750, 753, 765, 790, 797, 853, 922, 961, 985, 1012, 1049, 1057, 1114, 1131, 1135, 1150, 1156, 1162, 1180, 1191, 1218, 1221, 1249, 1253, 1286, 1293, 1308, 1317, 1320, 1321, 1332, 1335, and 1339 of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00175] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least

nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, or at least forty-three mutations selected from the group consisting of X575S , X596Y, X631L, X649R, X664K, X710E, X740A, X743I, X748I, X750A, X753G, X765X, X790A, X797E, X853E, X922A, X961E, X985Y, X1012A, X1049G, X1057V, X1114G, X1131C, X1135N, X1150V, X1156E, X1162A, X1180G, X1180A, X1191N, X1218S, X1221H, X1249S, X1253K, X1286K, X1293T, X1308D, X1317K, X1320V, X1321S, X1332G, X1335L, and X1339I of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00176] In some embodiments, the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, at least thirteen, at least fourteen, at least fifteen, at least sixteen, at least seventeen, at least eighteen, at least nineteen, at least twenty, at least twenty-one, at least twenty-two, at least twenty-three, at least twenty-four, at least twenty-five, at least twenty-six, at least twenty-seven, at least twenty-eight, at least twenty-nine, at least thirty, at least thirty-one, at least thirty-two, at least thirty-three, at least thirty-four, at least thirty-five, at least thirty-six, at least thirty-seven, at least thirty-eight, at least thirty-nine, at least forty, at least forty-one, at least forty-two, or at least forty-three mutations selected from the group consisting of F575S, D596Y, M631L, K649R, R664K, K710E, T740A, V743I, V748I, V750A, R753G, R765X, E790A, K797E, D853E, V922A, K961E, H985Y, D1012A, E1049G, I1057V, R1114G, Y1131C, D1135N, E1150V, K1156E, E1162A, D1180G, D1180A, K1191N, G1218S, Q1221H, P1249S, E1253K, N1286K, A1293T,

N1308D, N1317K, A1320V, P1321S, D1332G, R1335L, and T1339I of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in another Cas9 sequence (e.g., any of the sequences of SEQ ID NOs: 2, 4, or 6-11), wherein X represents any amino acid.

[00177] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X570T mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X570S.

[00178] In some embodiments, the amino acid sequence of the Cas9 domain comprises an I570T mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is I570S.

[00179] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X589S mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X589V.

[00180] In some embodiments, the amino acid sequence of the Cas9 domain comprises an A589S mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is A589V.

[00181] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X630G mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X630K.

[00182] In some embodiments, the amino acid sequence of the Cas9 domain comprises an E630G mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is E630K.

[00183] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X631A mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence 2, wherein X represents any amino acid. In some embodiments, the mutation is X631I. In some embodiments, the mutation is X631L. In some embodiments, the mutation is X631V.

[00184] In some embodiments, the amino acid sequence of the Cas9 domain comprises an M631A mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is M631I. In some embodiments, the mutation is M631L. In some embodiments, the mutation is M631V.

[00185] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X647A mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X647I.

[00186] In some embodiments, the amino acid sequence of the Cas9 domain comprises an V647A mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is V647I.

[00187] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X654H mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X654I. In some embodiments, the mutation is X654L.

[00188] In some embodiments, the amino acid sequence of the Cas9 domain comprises an R654H mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is R654I. In some embodiments, the mutation is R654L.

[00189] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X890E mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X890N.

[00190] In some embodiments, the amino acid sequence of the Cas9 domain comprises a K890E mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is K890N.

[00191] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X1016C mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X1016D. In some embodiments, the mutation is X1016S.

[00192] In some embodiments, the amino acid sequence of the Cas9 domain comprises an Y1016C mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is Y1016D. In some embodiments, the mutation is Y1016S.

[00193] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X1021L mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X1021T.

[00194] In some embodiments, the amino acid sequence of the Cas9 domain comprises an M1021L mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is M1021T.

[00195] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X1036D mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in

another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X1036H.

[00196] In some embodiments, the amino acid sequence of the Cas9 domain comprises an Y1036D mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is Y1036H.

[00197] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X1057S mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X1057T. In some embodiments, the mutation is X1057V.

[00198] In some embodiments, the amino acid sequence of the Cas9 domain comprises an I1057S mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is I1057T. In some embodiments, the mutation is X1057V.

[00199] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X1127A mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X1121G.

[00200] In some embodiments, the amino acid sequence of the Cas9 domain comprises an D1127A mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is D1127G.

[00201] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X1156E mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X1156N.

[00202] In some embodiments, the amino acid sequence of the Cas9 domain comprises an K1156E mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is K1156N.

[00203] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X1180E mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X1180G.

[00204] In some embodiments, the amino acid sequence of the Cas9 domain comprises an D1180E mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is D1180G.

[00205] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X1286H mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X1286K.

[00206] In some embodiments, the amino acid sequence of the Cas9 domain comprises an N1286H mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is N1286K.

[00207] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X1132G mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X1132N.

[00208] In some embodiments, the amino acid sequence of the Cas9 domain comprises an D1132G mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is D1132N.

[00209] In some embodiments, the amino acid sequence of the Cas9 protein comprises an X1335L mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence, wherein X represents any amino acid. In some embodiments, the mutation is X1335Q.

[00210] In some embodiments, the amino acid sequence of the Cas9 domain comprises an R1335L mutation in the amino acid sequence provided in SEQ ID NO: 2, or a corresponding mutation in another Cas9 sequence. In some embodiments, the mutation is R1335Q.

[00211] In some embodiments, the Cas9 protein comprises a combination of mutations that exhibit activity on a target sequence comprising a 5′-NAA-3′ PAM sequence at its 3'-end. In some embodiments, the combination of mutations are present in any one of the clones listed in Table 1. In some embodiments, the combination of mutations are conservative mutations of the clones listed in Table 1. In some embodiments, the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones listed in Table 1. In some embodiments, the Cas9 protein comprises the combination of mutations present in any one of the Cas9 clones selected from the group consisting of N3.19.4c-3; N3.19.4c-4; P4.2-72-4; P4.2-72-5; P10.6.144.2; P10.5.192.7; P10.5.192.10; P10.6.144.5; P10.6.192.1; P10.6.192.9; P10.6.192.12; P13.2-8; P13.3-3; P13.4-3; P16.2-120-1; P16.2-120-2; P16.2-120-3; P16.2-120-4; P16.2-120-5; P16.2-120-6; P16.1-3; P16.3-2; P16.4-5(es); and P16.6-2. In some embodiments, the Cas9 protein comprises the combination of mutations present in any one of the Cas9 clones selected from the group consisting of N3.19.4c-3; N3.19.4c-4; P4.2-72-4; P4.2-72-5; P10.6.144.2; P10.5.192.7; P10.5.192.10; P10.6.144.5; P10.6.192.1; P10.6.192.9; P10.6.192.12; P13.2-8; P13.3-3; P13.4-3; P16.2-120-1; P16.2-120-2; P16.2-120-3; P16.2-120-4; P16.2-120-5; P16.2-120-6; P16.1-3; P16.3-2; P16.4-5(es); and P16.6-2, or a combination of conservative mutations thereto.

[00212] Table 1: NAA PAM Clones

Clone	Mutations from wild-type SpCas9 (e.g., SEQ ID NO: 2)

N3.19.4-c	D177N, K218R, D614N, D1135N, P1137S, E1219V, A1320V, A1323D, R1333K
N3.19.4-4	D177N, K218R, D614N, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
P4.2-72-4	A10T, I322V, S409I, E427G, G715C, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
P4.2-72-5	A367T, K710E, R1114G, D1135N, P1137S, E1219V, Q1221H, H1264Y, A1320V, R1333K
P10.6.144.2	A10T, I322V, S409I, E427G, R753G, D861N, D1135N, K1188R, E1219V, Q1221H, H1264H, A1320V, R1333K
P10.5.192.7	A10T, I322V, S409I, E427G, R654L, V743I, R753G, M1021T, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, A1320V, R1333K
P10.5.192.10	A10T, I322V, S409I, E427G, V743I, R753G, E762G, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, A1320V, R1333K
P10.6.144.5	A10T, I322V, S409I, E427G, R753G, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, S1274R, A1320V, R1333K
P10.6.192.1	A10T, I322V, S409I, E427G, A589S, R753G, D1135N, E1219V, Q1221H, H1264H, A1320V, R1333K
P10.6.192.9	A10T, I322V, S409I, E427G, R753G, E757K, G865G, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
P10.6.192.12	A10T, I322V, S409I, E427G, R654L, R753G, E757K, D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K
P13.2-8	A10T, I322V, S409I, E427G, K599R, M631A, R654L, K673E, V743I, R753G, N758H, E762G, D1135N, D1180G, E1219V, Q1221H, Q1256R, H1264Y, A1320V, A1323D, R1333K
P13.3-3	A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N869S, N1054D, R1114G, D1135N, D1180G, E1219V, Q1221H, H1264Y, A1320V, A1323D, R1333K
P13.4-3	A10T, I322V, S409I, E427G, R654L, L727I, V743I, R753G, E762G, R859S, N946D, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, N1317T, A1320V, A1323D, R1333K
P16.2-120-1	A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, V1290G, L1318S, A1320V, A1323D, R1333K

P16.2-120-2	A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, K1151E, D1180G, E1219V, Q1221H, H1264Y, V1290G, L1318S, A1320V, R1333K
P16.2-120-3	A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, V1290G, L1318S, A1320V, A1323D, R1333K
P16.2-120-4	A10T, I322V, S409I, E427G, R654L, K673E, F693L, V743I, R753G, E762G, N803S, N869S, L921P, Y1016D, G1077D, F1080S, R1114G, D1135N, D1180G, E1219V, Q1221H, H1264Y, L1318S, A1320V, A1323D, R1333K
P16.2-120-5	A10T, I322V, S409I, E427G, E630K, R654L, K673E, V743I, R753G, E762G, Q768H, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, H1264Y, L1318S, A1320V, R1333K
P16.2-120-6	A10T, I322V, S409I, E427G, R654L, K673E, F693L, V743I, R753G, E762G, Q768H, N803S, N869S, Y1016D, G1077D, R1114G, F1134L, D1135N, D1180G, E1219V, Q1221H, G1223S, H1264Y, L1318S, A1320V, R1333K
P16.1-3	A10T, I322V, S409I, E427G, R654L, K673E, F693L, V743I, R753G, E762G, N803S, N869S, L921P, Y1016D, G1077D, F1801S, R1114G, D1135N, D1180G, E1219V, Q1221H, H1264Y, L1318S, A1320V, A1323D, R1333K
P16.3-2	A10T, I322V, S409I, E427G, R654L, V743I, R753G, M1021T, D1135N, D1180G, K1211R, E1219V, Q1221H, H1264Y, A1320V, R1333K
P16.4-5(es)	A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, M673I, N803S, N869S, G1077D, R1114G, D1135N, V1139A, D1180G, E1219V, Q1221H, A1320V, R1333K
P16.6-2	A10T, I322V, S409I, E427G, R654L, K673E, V743I, R753G, E762G, N803S, N869S, R1114G, D1135N, E1219V, Q1221H, A1320V, R1333K

[00213] In some embodiments, the Cas9 protein comprises an amino acid sequence that is at least 80% identical to the amino acid sequence of a Cas9 protein as provided by any one of the variants of Table 1. In some embodiments, the Cas9 protein comprises an amino acid sequence that is at least

85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence of a Cas9 protein as provided by any one of the variants of Table 1.

[00214] In some embodiments, the Cas9 protein exhibits an increased activity on a target sequence that does not comprise the canonical PAM (5′-NGG-3′) at its 3' end as compared to *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2. In some embodiments, the Cas9 protein exhibits an activity on a target sequence having a 3' end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 5-fold increased as compared to the activity of *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2 on the same target sequence. In some embodiments, the Cas9 protein exhibits an activity on a target sequence that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold increased as compared to the activity of *Streptococcus pyogenes* as provided by SEQ ID NO: 2 on the same target sequence. In some embodiments, the 3' end of the target sequence is directly adjacent to an AAA, GAA, CAA, or TAA sequence.

[00215] In some embodiments, the Cas9 protein comprises a combination of mutations that exhibit activity on a target sequence comprising a 5′-NAC-3′ PAM sequence at its 3'-end. In some embodiments, the combination of mutations are present in any one of the clones listed in Table 2. In some embodiments, the combination of mutations are conservative mutations of the clones listed in Table 2. In some embodiments, the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones listed in Table 2. In some embodiments, the Cas9 protein comprises the combination of mutations present in any one of the Cas9 clones selected from the group consisting of N4.CAC-1; N4.CAC-5; N4.CAC06; SacB.CAC.4h; N3.CAC-1; N3.CAC-5; N3.CAC-6; N3.CAC-8; P15.1.166-3; P15.1.166-8; P15.2.166-2; P15.3.166-4; P15.3.166-5; P15.3.166-7; P15.4.166-4;

P15.4.166-8; P17.1.144-1; P17.1.144-2; P17.1.144-3; P17.1.144-4; P17.1.144-5; P17.1.144-7; P17.1.144-8; P17.2.144-1; P17.2.144-2; P17.2.144-3; P17.2.144-4; P17.2.144-5; P17.2.144-6; P17.2.144-7; P17.2.144-8; P17.1-1; P17.1-5; and P17.1.7-4(fn). In some embodiments, the Cas9 protein comprises the combination of mutations present in any one of the Cas9 clones selected from the group consisting of N4.CAC-1; N4.CAC-5; N4.CAC06; SacB.CAC.4h; N3.CAC-1; N3.CAC-5; N3.CAC-6; N3.CAC-8; P15.1.166-3; P15.1.166-8; P15.2.166-2; P15.3.166-4; P15.3.166-5; P15.3.166-7; P15.4.166-4; P15.4.166-8; P17.1.144-1; P17.1.144-2; P17.1.144-3; P17.1.144-4; P17.1.144-5; P17.1.144-7; P17.1.144-8; P17.2.144-1; P17.2.144-2; P17.2.144-3; P17.2.144-4; P17.2.144-5; P17.2.144-6; P17.2.144-7; P17.2.144-8; P17.1-1; P17.1-5; and P17.1.7-4(fn), or a combination of conservative mutations thereto.

[00216] Table 2: NAC PAM Clones

Clone	Mutations from wild-type SpCas9 (e.g., SEQ ID NO: 2)
N4.CAC-1	T472I, R753G, K890E, D1332N, R1335Q, T1337N
N4.CAC-5	I1057S, D1135N, P1301S, R1335Q, T1337N
N4.CAC06	T472I, R753G, D1332N, R1335Q, T1337N
SacB.CAC.4h	D1135N, E1219V, D1332N, R1335Q, T1337N
N3.CAC-1	T472I, R753G, K890E, D1332N, R1335Q, T1337N
N3.CAC-5	I1057S, D1135N, P1301S, R1335Q, T1337N
N3.CAC-6	T472I, R753G, D1332N, R1335Q, T1337N
N3.CAC-8	T472I, R753G, Q771H, D1332N, R1335Q, T1337N
P15.1.166-3	E627K, T638P, K652T, R753G, N803S, K959N, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
P15.1.166-8	E627K, T638P, K652T, R753G, N803S, K959N, R1114G, D1135N, K1156E, E1219V, D1332N, R1335Q, T1337N
P15.2.166-2	E627K, T638P, V647I, R753G, N803S, K959N, G1030R, I1055E, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
P15.3.166-6	E627K, E630G, T638P, V647A, G687R, N767D, N803S, K959N, R1114G, D1135N, E1219V, D1332G, R1335Q, T1337N
P15.3.166-4	E627K, T638P, R753G, N803S, K959N, R1114G, D1135N, E1219V, N1266H, D1332N, R1335Q, T1337N
P15.3.166-5	E627K, T638P, R753G, N803S, K959N, I1057T, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N

P15.3.166-7	E627K, T638P, R753G, N803S, K959N, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
P15.4.166-4	E627K, M631I, T638P, R753G, N803S, K959N, Y1036H, R1114G, D1135N, E1219V, D1251G, D1332G, R1335Q, T1337N
P15.4.166-8	E627K, T638P, R753G, N803S, V875I, K959N, Y1016C, R1114G, D1135N, E1219V, D1251G, D1332G, R1335Q, T1337N, I1348V
P17.1.144-1	K608R, E627K, T638P, V647I, R654L, R753G, N803S, T804A, K848N, V922A, K959N, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
P17.1.144-2	K608R, E627K, T638P, V647I, R753G, N803S, V922A, K959N, K1014N, V1015A, R1114G, D1135N, K1156N, E1219V, N1252D, D1332N, R1335Q, T1337N
P17.1.144-3	K608R, E627K, R629G, T638P, V647I, A711T, R753G, K775R, K789E, N803S, K959N, V1015A, Y1036H, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
P17.1.144-4	K608R, E627K, T638P, V647I, T740A, R753G, N803S, K948E, K959N, Y1016S, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
P17.1.144-5	K608R, E627K, T638P, V647I, T740A, N803S, K948E, K959N, Y1016S, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
P17.1.144-7	I670S, K608R, E627K, E630G, T638P, V647I, R653K, R753G, I795L, K797N, N803S, K866R, K890N, K959N, Y1016C, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
P17.1.144-8	K608R, E627K, T638P, V647I, T740A, G752R, R753G, K797N, N803S, K948E, K959N, V1015A, Y1016S, R1114G, D1135N, E1219V, N1266H, D1332N, R1335Q, T1337N
P17.2.144-1	I570T, A589V, K608R, E627K, T638P, V647I, R654L, Q716R, R753G, N803S, K948E, K959N, Y1016S, R1114G, D1135N, E1207G, E1219V, N1234D, D1332N, R1335Q, T1337N
P17.2.144-2	K608R, E627K, R629G, T638P, V647I, R654L, Q740R, R753G, N803S, K959N, N990S, T995S, V1015A, Y1036D, R1114G, D1135N, E1207G, E1219V, N1234D, N1266H, D1332N, R1335Q, T1337N
P17.2.144-3	I562F, V565D, I570T, K608R, L625S, E627K, T638P, V647I, R654I, G752R, R753G, N803S, N808D, K959N,

	M1021L, R1114G, D1135N, N1177S, N1234D, D1332N, R1335Q, T1337N
P17.2.144-4	I562F, I570T, K608R, E627K, T638P, V647I, R753G, E790A, N803S, K959N, V1015A, Y1036H, R1114G, D1135N, D1180E, A1184T, E1219V, D1332N, R1335Q, T1337N
P17.2.144-5	I570T, K608R, E627K, T638P, V647I, R654H, R753G, E790A, N803S, K959N, V1015A, R1114G, D1127A, D1135N, E1219V, D1332N, R1335Q, T1337N
P17.2.144-6	I570T, K608R, L625S, E627K, T638P, V647I, R654I, T703P, R753G, N803S, N808D, K959N, M1021L, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
P17.2.144-7	I570S, K608R, E627K, E630G, T638P, V647I, R653K, R753G, I795L, N803S, K866R, K890N, K959N, Y1016C, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N
P17.2.144-8	I570T, K608R, E627K, T638P, V647I, R654H, R753G, E790A, N803S, K959N, V1016A, R1114G, D1135N, E1219V, K1246E, D1332N, R1335Q, T1337N
P17.1-1	K608R, E627K, T638P, V647I, R654L, K673E, R753G, E790A, N803S, K948E, K959N, R1114G, D1127G, D1135N, D1180E, E1219V, N1286H, D1332N, R1335Q, T1337N
P17.1-5	K608R, L625S, E627K, T638P, V647I, R654I, I670T, R753G, N803S, N808D, K959N, M1021L, R1114G, D1135N, E1219V, N1286H, D1332N, R1335Q, T1337N
P17.1.7-4(fn)	E627K, M631V, T638P, V647I, K710E, R753G, N803S, N808D, K948E, M1021L, R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N, S1338T, H1349R

[00217] In some embodiments, the Cas9 protein comprises an amino acid sequence that is at least 80% identical to the amino acid sequence of a Cas9 protein as provided by any one of the variants of Table 2. In some embodiments, the Cas9 protein comprises an amino acid sequence that is at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence of a Cas9 protein as provided by any one of the variants of Table 2.

[00218] In some embodiments, the Cas9 protein exhibits an increased activity on a target sequence that does not comprise the canonical PAM (5′-NGG-3′) at its 3' end as compared to *Streptococcus*

pyogenes Cas9 as provided by SEQ ID NO: 2. In some embodiments, the Cas9 protein exhibits an activity on a target sequence having a 3' end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 5-fold increased as compared to the activity of *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2 on the same target sequence. In some embodiments, the Cas9 protein exhibits an activity on a target sequence that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold increased as compared to the activity of *Streptococcus pyogenes* as provided by SEQ ID NO: 2 on the same target sequence. In some embodiments, the 3' end of the target sequence is directly adjacent to an AAC, GAC, CAC, or TAC sequence.

[00219] In some embodiments, the Cas9 protein comprises a combination of mutations that exhibit activity on a target sequence comprising a 5′-NAT-3′ PAM sequence at its 3'-end. In some embodiments, the combination of mutations are present in any one of the clones listed in Table 3. In some embodiments, the combination of mutations are conservative mutations of the clones listed in Table 3. In some embodiments, the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones listed in Table 3. In some embodiments, the Cas9 protein comprises the combination of mutations present in any one of the Cas9 clones selected from the group consisting of SacB.N4.19.TAT-4h-1; SacB.N4.19.TAT-4h-3; P12.2.b9-8; P12.3.b9-8; P12.3.b9-8 (ax); P12.3.b10-6; SacB.P12a2.AAT.3hr.maj; SacB.P12a2.AAT.3hr.min; P17.4-1; P17.4-2; P17.4-3; P17.4-4; P17.4-5; P17.4-6; P17.4-8; P17-4-1-1; P17-4-3-1; and P17-4-6-1. In some embodiments, the Cas9 protein comprises the combination of mutations present in any one of the Cas9 clones selected from the group consisting of SacB.N4.19.TAT-4h-1; SacB.N4.19.TAT-4h-3; P12.2.b9-8; P12.3.b9-8; P12.3.b9-8 (ax); P12.3.b10-6; SacB.P12a2.AAT.3hr.maj; SacB.P12a2.AAT.3hr.min; P17.4-1; P17.4-

2; P17.4-3; P17.4-4; P17.4-5; P17.4-6; P17.4-8; P17-4-1-1; P17-4-3-1; and P17-4-6-1, or a combination of conservative mutations thereto.

[00220] Table 3: NAT PAM Clones

Clone	Mutations from wild-type SpCas9 (e.g., SEQ ID NO: 2)
SacB.N4.19.TAT-4h-1	K961E, H985Y, D1135N, K1191N, E1219V, Q1221H, A1320A, P1321S, R1335L
SacB.N4.19.TAT-4h-3	D1135N, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
P12.2.b9-8	V743I, R753G, E790A, D1135N, G1218S, E1219V, Q1221H, A1227V, P1249S, N1286K, A1293T, P1321S, D1322G, R1335L, T1339I
P12.3.b9-8 (ax)	F575S, M631L, R654L, V748I, V743I, R753G, D853E, V922A, R1114G D1135N, G1218S, E1219V, Q1221H, A1227V, P1249S, N1286K, A1293T, P1321S, D1322G, R1335L, T1339I
P12.3.b10-6	F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, N1286K, P1321S, D1322G, R1335L
SacB.P12a2.AAT.3hr.maj	M631L, R654L, R753G, K797E, D853E, V922A, D1012A, R1114G D1135N, G1218S, E1219V, Q1221H, P1249S, N1317K, P1321S, D1322G, R1335L
P17.4-1	F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
P17.4-2	F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
P17.4-3	F575S, D596Y, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, Q1256R, P1321S, D1322G, R1335L
P17.4-4	F575S, M631L, R654L, R664K, K710E, V750A, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L

P17.4-5	F575S, M631L, K649R, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, K1156E, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
P17.4-6	F575S, M631L, R654L, R664K, R753G, D853E, V922A, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1322G, R1335L
P17.4-8	F575S, M631L, R654L, R664K, R753G, D853E, V922A, I1057G, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, N1308D, P1321S, D1322G, R1335L
P17-4-1-1	M631L, R654L, R753G, D853E, V922A, R1114G, Y1131C, D1135N, E1150V, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1332G, R1335L
P17-4-3-1	M631L, R654L, R664K, R753G, D853E, I1057V, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1332G, R1335L
P17-4-6-1	M631L, R654L, R664K, R753G, I1057V, R1114G, Y1131C, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, P1321S, D1332G, R1335L

[00221] In some embodiments, the Cas9 protein comprises an amino acid sequence that is at least 80% identical to the amino acid sequence of a Cas9 protein as provided by any one of the variants of Table 3. In some embodiments, the Cas9 protein comprises an amino acid sequence that is at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence of a Cas9 protein as provided by any one of the variants of Table 3.

Cas9 Activity

[00222] In some embodiments, the Cas9 protein exhibits an increased activity on a target sequence that does not comprise the canonical PAM (5′-NGG-3′) at its 3' end as compared to *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2. In some embodiments, the Ca9 protein exhibits an activity on a target sequence having a 3' end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 5-fold increased as compared to the activity of *Streptococcus*

pyogenes Cas9 as provided by SEQ ID NO: 2 on the same target sequence. In some embodiments, the Cas9 protein exhibits an activity on a target sequence that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold increased as compared to the activity of *Streptococcus pyogenes* as provided by SEQ ID NO: 2 on the same target sequence. In some embodiments, the 3' end of the target sequence is directly adjacent to an AAT, GAT, CAT, or TAT sequence.

[00223] In some embodiments, the Cas9 domain exhibits activity on a target sequence having a 3′-end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′), or on a target sequence that does not comprise the canonical PAM sequence (5′-NGG-3′), that is at least 2-fold, at least 3-fold, at least 4-fold, at least 5-fold, at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold increased as compared to the activity of *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2 on the same target sequence. In some embodiments, the Cas9 domain exhibits activity on a target sequence having a 3′-end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′), or on a target sequence that does not comprise the canonical PAM sequence (5′-NGG-3′), that is at least 10%, at least 15%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, or at least 90% greater than the activity of *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2 on the same target sequence. In some embodiments, the 3′-end of the target sequence is directly adjacent to an NGT, NGA, NGC, and NNG sequence, wherein N is A, G, T, or C. In some embodiments, the 3′-end of the target sequence is directly adjacent to an AAA, AAC, AAG, AAT, CAA, CAC, CAG, CAT, GAA, GAC, GAG, GAT, TAA, TAC, TAG, TAT, ACA, ACC, ACG, ACT, CCA, CCC, CCG, CCT, GCA, GCC, GCG, GCT, TCA, TCC, TCG, TCT, AGA, AGC, AGT, CGA, CGC, CGT, GGA, GGC, GGT, TGA, TGC, TGT,

ATA, ATC, ATG, ATT, CTA, CTC, CTG, CTT, GTA, GTC, GTG, GTT, TTA, TTC, TTG, or TTT PAM sequence. In some embodiments, the 3′-end of the target sequence is directly adjacent to an CGG, AGT, TGG, AGT, CGT, GGG, CGT, TGT, GGT, AGC, CGC, TGC, GGC, AGA, CGA, TGA, GGA, GAA, GAT, or CAA sequence. In some embodiments, the Cas9 domain activity is measured by a nuclease assay, a deamination assay, a transcriptional activation assay, a binding assay, or by PCR or sequencing. In some embodiments, the transcriptional activation assay is a reporter activation assay, such as a GFP activation assay. Exemplary methods for measuring binding activity (*e.g.,* of Cas9) using transcriptional activation assays are known in the art and would be apparent to the skilled artisan. For example, methods for measuring Cas9 activity using the tripartite activator VPR have been described in Chavez A., *et al.*, "Highly efficient Cas9-mediated transcriptional programming." *Nature Methods* 12, 326–328 (2015), the entire contents of which are incorporated by reference herein.

[00224] In some embodiments, the Cas9 domain is mutated with respect to a corresponding wild-type protein such that the mutated Cas9 domain lacks the ability to cleave one or both strands of a target polynucleotide containing a target sequence. In particular embodiments, an aspartate-to-alanine substitution (D10A) in the RuvC1 catalytic domain of *S. pyogenes* Cas9 converts Cas9 from a nuclease that cleaves both strands to a nickase that nicks the targeted strand, or the strand that is complementary to the gRNA. A histidine-to-alanine substitution (H840A) in the HNH catalytic domain of S. pyogenes Cas9 generates a nick on the strand that is displaced by the gRNA during strand invasion, also referred to herein as the non-edited strand. The single catalytically active nuclease site of the nCas9 leaves a nick in the non-edited strand, which will direct mismatch repair machinery to read (rather than remove) the modified base during repair (i.e., a substituted guanine or guanine derivative at the target site). Other examples of mutations that render Cas9 a nickase include, without limitation, N854A and N863A in SpCas9, and corresponding mutations in other wild-

type Cas9 proteins or variants thereof. Reference is made to U.S. Patent No. 8,945,839, incorporated herein by reference.

[00225] In some embodiments, the amino acid sequence of the HNH domain is at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence of the HNH domain of any of SEQ ID NO: 2. In some embodiments, the amino acid sequence of the RuvC domain is at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence of the RuvC domain of SEQ ID NO: 2. In some embodiments, the Cas9 domain comprises the RuvC and HNH domains of SEQ ID NO: 2. In some embodiments, the Cas9 domain comprises a D10A and/or a H840A mutation in the amino acid sequence provided in SEQ ID NO: 2, or corresponding mutation(s) in another Cas9 sequence.

[00226] In some embodiments, the disclosure provides SpCas9 mutant proteins that work best on NRRH, NRCH, and NRTH PAMs. The SpCas9 mutant protein that works best on NARH ("es" variant), has an amino acid sequence as presented in SEQ ID NO: 22 (underligned residues are mutated from SpCas9)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
IIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS

MPQVNIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGFNSPTAAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIGFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLHKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGVPAAFKYFDTTIDKKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD (SEQ ID NO: 22).

[00227] The SpCas9 mutant protein that works best on NRCH (“fn” variant), has an amino acid sequence as presented in SEQ ID NO: 23 (underligned residues are mutated from SpCas9)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
IIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTINRKQYNTTKEVLDATLIRQSITGLYETRI
DLSQLGGD (SEQ ID NO: 23)

[00228] The SpCas9 mutant protein that works best on NRTH (“ax” variant), has an amino acid sequence as presented in SEQ ID NO: 24 (underligned residues are mutated from SpCas9)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD

DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
IIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIGFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASASVLHKGN
ELALPSKYVNFLYLASHYEKLKGSSEDNKQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGASAAFKYFDTTIGRKLYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD (SEQ ID NO: 24)

[00229]Some aspects of the disclosure provide high fidelity Cas9 domains. In some embodiments, high fidelity Cas9 domains have decreased electrostatic interactions between the Cas9 domain and a sugar-phosphate backbone of a DNA, as compared to a wild-type Cas9 domain. In some embodiments, any of the Cas9 domains provided herein comprise one or more mutations that decrease the association between the Cas9 domain and a sugar-phosphate backbone of a DNA. In some embodiments, any of the Cas9 domains provided herein comprise one or more mutations that decrease the association between the Cas9 domain and a sugar-phosphate backbone of a DNA by at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95%. In some embodiments, any of the Cas9 domains provided herein comprise one or more of a N497X, a R661X, a Q695X, and/or a Q926X mutation of the amino acid sequence provided in SEQ ID NO: 135, or a corresponding mutation in another Cas9 sequence, wherein X is any amino acid. In some embodiments, any of the Cas9 domains provided herein comprise one or more of a N497A, a R661A, a Q695A, and/or a Q926A mutation of the amino acid sequence provided in SEQ ID NO: 135, or a

corresponding mutation in another Cas9 sequence. In some embodiments, the Cas9 domain comprises a D10A mutation of the amino acid sequence provided in SEQ ID NO: 135, or a corresponding mutation in another Cas9 sequence. In some embodiments, the Cas9 domain comprises the amino acid sequence as set forth in SEQ ID NO: 135. High fidelity Cas9 domains have been described in the art and would be apparent to the skilled artisan. For example, high fidelity Cas9 domains have been described in Kleinstiver, B.P., *et al.* "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." *Nature* 529, 490-495 (2016); and Slaymaker, I.M., et al. "Rationally engineered Cas9 nucleases with improved specificity." Science 351, 84-88 (2015); the entire contents of each are incorporated herein by reference. It should be appreciated that, based on the present disclosure and knowledge in the art, that mutations in any Cas9 domain may be generated to make high fidelity Cas9 domains that have decreased electrostatic interactions between the Cas9 domain and a sugar-phosphate backbone of a DNA, as compared to a wild-type Cas9 domain.

[00230]Cas9 domain where mutations relative to Cas9 of SEQ ID NO: 6 are shown in bold and underlines.

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMT**A**FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG**A**LSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFM**A**LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR**A**ITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS

MPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD (SEQ ID NO: 25)

[00231] In some embodiments, the Cas9 domain of any of the fusion proteins provided herein comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid set forth as SEQ ID NO: 10 (*S. aureus* Cas9), below. In some embodiments, the Cas9 domain of any of the fusion proteins provided herein comprises the amino acid sequence of SEQ ID NO: 10. In some embodiments, the Cas9 domain of any of the fusion proteins provided herein consists of the amino acid sequence of SEQ ID NO: 10.

[00232] An exemplary SaCas9 amino acid sequence is:

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKK
LLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQ
ISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLL
ETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENE
KLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEI
IENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWH
TNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIE
LAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLED
LLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAK
GKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFT
SFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQE
YKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLK
KLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGN
KLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKL
KKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIA
SKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG (SEQ ID NO: 10, *Staph. aureus* Cas9)

[00233] An additional Cas9 domain with altered PAM specificity, such as a domain having at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% sequence identity with wild type *Geobacillus thermodenitrificans* Cas9 (SEQ ID NO: 11, GeoCas9) may be used.

MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRRLRRRKHRLE
RIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARILLHLAKRRGFRSNRKSERT
NKENSTMLKHIEENQSILSSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIFAKQREY
GNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFTVWEHINKLRLVS
PGGIRALTDDERRLIYKQAFHKNKITFHDVRTLLNLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYH
KIRKAIDSVYGKGAAKSFRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLADKVYDEELIE
ELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQKTVLLPNIPPIANPVVMRA
LTQARKVVNAIIKKYGSPVSIHIELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTG
LDIVKFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDHVIPYSRSLDDSYTNKVLVLTKENREKGNRT
PAEYLGLGSERWQQFETFVLTNKQFSKKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFLANFIREH
LKFADSDDKQKVYTVNGRITAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQRREQN
KELSKKTDPQFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMPKRSITGAAH
QETLRRYIGIDERSGKIQTVVKKKLSEIQLDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQE
PLYKPKKNGELGPIIRTIKIIDTTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTIDMMKGI
LPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIDSSN
GGLSLVSHDNNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRGEKRVGVASSSHSKAGETIRPL (SEQ
ID NO: 11).

[00234] In some embodiments, a Cas9 domain refers to a Cas9 or Cas9 homolog from archaea (e.g., nanoarchaea), which constitute a domain and kingdom of single-celled prokaryotic microbes. In some embodiments, a Cas9 domain may comprise a CasX (now referred to as Cas12e) or CasY (now referred to as Cas12d) omain, which have been described in, for example, Burstein *et al.,* "New CRISPR–Cas systems from uncultivated microbes." *Cell Res.* 2017 Feb 21. doi: 10.1038/cr.2017.21, and Liu *et al.,* "CasX enzymes comprise a distinct family of RNA-guided genome editors," *Nature.* 2019; 566(7743):218-223, each of which is incorporated herein by reference. Using genome-resolved metagenomics, a number of CRISPR–Cas systems were identified, including the first reported Cas9 in the archaeal domain of life. This divergent Cas9 protein was found in little-studied nanoarchaea as part of an active CRISPR–Cas system. In bacteria, two previously unknown systems

were discovered, CRISPR–CasX and CRISPR–CasY, which are among the most compact systems yet discovered. In some embodiments, napDNAbp domain refers to CasX, or a variant of CasX. In some embodiments, napDNAbp domain refers to a CasY, or a variant of CasY. It should be appreciated that other RNA-guided DNA binding proteins may be used as a napDNAbp and are within the scope of this disclosure. In some embodiments, the napDNAbp comprises an amino acid sequence that is at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to a naturally-occurring CasX or CasY protein.

Cytidine deaminases

[00235] In some embodiments, the deaminase domain is a cytidine deaminase domain. A cytidine deaminase domain may also be referred to interchangeably as a cytosine deaminase domain. In some embodiments, the cytidine deaminase catalyzes the hydrolytic deamination of cytidine (C) or deoxycytidine (dC) to uridine (U) or deoxyuridine (dU), respectively. In some embodiments, the cytidine deaminase domain catalyzes the hydrolytic deamination of cytosine (C) to uracil (U). In some embodiments, the cytidine deaminase catalyzes the hydrolytic deamination of cytidine or cytosine in deoxyribonucleic acid (DNA). Without wishing to be bound by any particular theory, fusion proteins comprising a cytidine deaminase are useful *inter alia* for targeted editing, referred to herein as "base editing," of nucleic acid sequences *in vitro* and *in vivo*.

[00236] One exemplary suitable type of cytidine deaminase is a cytidine deaminase, for example, of the APOBEC family. The apolipoprotein B mRNA-editing complex (APOBEC) family of cytidine deaminase enzymes encompasses eleven proteins that serve to initiate mutagenesis in a controlled and beneficial manner (see, *e.g.*, Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* 2008; 9(6):229). One family member, activation-induced cytidine deaminase (AID), is responsible for the maturation of antibodies by converting cytosines in ssDNA to uracils in a transcription-dependent, strand-biased fashion (see, *e.g.*, Reynaud CA, *et al.* What role for AID:

mutator, or assembler of the immunoglobulin mutasome, *Nat Immunol.* 2003; 4(7):631-638). The apolipoprotein B editing complex 3 (APOBEC3) enzyme provides protection to human cells against a certain HIV-1 strain via the deamination of cytosines in reverse-transcribed viral ssDNA (see, *e.g.*, Bhagwat AS. DNA-cytosine deaminases: from antibody maturation to antiviral defense. *DNA Repair (Amst).* 2004; 3(1):85-89). These proteins all require a Zn^{2+}-coordinating motif (His-X-Glu-X_{23-26}-Pro-Cys-X_{2-4}-Cys; SEQ ID NO: 405) and bound water molecule for catalytic activity. The Glu residue acts to activate the water molecule to a zinc hydroxide for nucleophilic attack in the deamination reaction. Each family member preferentially deaminates at its own particular "hotspot", ranging from WRC (W is A or T, R is A or G) for hAID, to TTC for hAPOBEC3F (see, *e.g.*, Navaratnam N and Sarwar R. An overview of cytidine deaminases. *Int J Hematol.* 2006; 83(3):195-200). A recent crystal structure of the catalytic domain of APOBEC3G revealed a secondary structure comprised of a five-stranded β-sheet core flanked by six α-helices, which is believed to be conserved across the entire family (see, *e.g.*, Holden LG, *et al.* Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. *Nature.* 2008; 456(7218):121-4). The active center loops have been shown to be responsible for both ssDNA binding and in determining "hotspot" identity (see, *e.g.*, Chelico L, *et al.* Biochemical basis of immunological and retroviral responses to DNA-targeted cytosine deamination by activation-induced cytidine deaminase and APOBEC3G. *J Biol Chem.* 2009; 284(41). 27761-5). Overexpression of these enzymes has been linked to genomic instability and cancer, thus highlighting the importance of sequence-specific targeting (see, *e.g.*, Pham P, *et al.* Reward versus risk: DNA cytidine deaminases triggering immunity and disease. *Biochemistry.* 2005; 44(8):2703-15).

[00237] Some aspects of this disclosure relate to the recognition that the activity of cytidine deaminase enzymes such as APOBEC enzymes can be directed to a specific site in genomic DNA. Without wishing to be bound by any particular theory, advantages of using a nucleic acid programmable binding protein (*e.g.*, a Cas9 domain) as a recognition agent include (1) the sequence specificity of nucleic acid programmable binding protein (*e.g.*, a Cas9 domain) can be easily altered

by simply changing the sgRNA sequence; and (2) the nucleic acid programmable binding protein (*e.g.*, a Cas9 domain) may bind to its target sequence by denaturing the dsDNA, resulting in a stretch of DNA that is single-stranded and therefore a viable substrate for the deaminase. It should be understood that other catalytic domains of napDNAbps, or catalytic domains from other nucleic acid editing proteins, can also be used to generate fusion proteins with Cas9, and that the disclosure is not limited in this regard.

[00238] In view of the results provided herein regarding the nucleotides that can be targeted by Cas9:deaminase fusion proteins, a person of ordinary skill in the art will be able to design suitable guide RNAs to target the fusion proteins to a target sequence that comprises a nucleotide to be deaminated.

[00239] In some embodiments, the cytidine deaminase is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. In some embodiments, the cytidine deaminase is an APOBEC1 deaminase. In some embodiments, the cytidine deaminase is an APOBEC2 deaminase. In some embodiments, the cytidine deaminase is an APOBEC3 deaminase. In some embodiments, the cytidine deaminase is an APOBEC3A deaminase. In some embodiments, the cytidine deaminase is an APOBEC3B deaminase. In some embodiments, the cytidine deaminase is an APOBEC3C deaminase. In some embodiments, the cytidine deaminase is an APOBEC3D deaminase. In some embodiments, the cytidine deaminase is an APOBEC3E deaminase. In some embodiments, the cytidine deaminase is an APOBEC3F deaminase. In some embodiments, the cytidine deaminase is an APOBEC3G deaminase. In some embodiments, the cytidine deaminase is an APOBEC3H deaminase. In some embodiments, the cytidine deaminase is an APOBEC4 deaminase. In some embodiments, the cytidine deaminase is an activation-induced deaminase (AID). In some embodiments, the cytidine deaminase is a vertebrate cytidine deaminase. In some embodiments, the cytidine deaminase is an invertebrate cytidine deaminase. In some embodiments, the cytidine deaminase is a human, chimpanzee, gorilla, monkey, cow, dog, rat, or mouse deaminase. In some embodiments, the cytidine deaminase is a

human cytidine deaminase. In some embodiments, the cytidine deaminase is a rat cytidine deaminase, *e.g.,* rAPOBEC1. In some embodiments, the cytidine deaminase is a *Petromyzon marinus* cytidine deaminase 1 (pmCDA1) (SEQ ID NO: 58). In some embodiments, the cytidine deaminase is a human APOBEC3G (SEQ ID NO: 60). In some embodiments, the cytidine deaminase is a fragment of the human APOBEC3G. In some embodiments, the deaminase is a human APOBEC3G variant comprising a D316R and D317R mutation. In some embodiments, the deaminase is a fragment of the human APOBEC3G and comprising mutations corresponding to the D316R and D317R mutations in SEQ ID NO: 61.

[00240] In some embodiments, the nucleic acid editing domain is at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the cytidine deaminase domain of any one of SEQ ID NOs: 27-61. In some embodiments, the nucleic acid editing domain comprises the amino acid sequence of any one of SEQ ID NOs: 27-61.

[00241] Some exemplary suitable nucleic-acid editing domains, *e.g.*, cytidine deaminases and cytidine deaminase domains, that can be fused to napDNAbps (*e.g.*, Cas9 domains) according to aspects of this disclosure are provided below. It should be understood that, in some embodiments, the active domain of the respective sequence can be used, *e.g.*, the domain without a localizing signal (nuclear localization sequence, without nuclear export signal, cytoplasmic localizing signal).

[00242] Human AID:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDW
DLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQI
AIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 27)

(underline: nuclear localization sequence; double underline: nuclear export signal)

[00243] Mouse AID:

MDSLLMKQKKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDW
DLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDRKAEPEGLRRLHRAGVQI

GIMTFKDYFYCWNTFVENRERTFKAWEGLHENSVRLTRQLRRILLPLYEVDDLRDAFRMLGF (SEQ ID NO: 28)

(underline: nuclear localization sequence; double underline: nuclear export signal)

[00244] Dog AID:

MDSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDW DLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDRKAEPEGLRRLHRAGVQI AIMTFKDYFYCWNTFVENREKTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 29)

(underline: nuclear localization sequence; double underline: nuclear export signal)

[00245] Bovine AID:

MDSLLKKQRQFLYQFKNVRWAKGRHETYLCYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDW DLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKERKAEPEGLRRLHRAGVQ IAIMTFKDYFYCWNTFVENHERTFKAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 30)

(underline: nuclear localization sequence; double underline: nuclear export signal)

[00246] Rat AID:

MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRSLLMKQRKFLYHF KNVRWAKGRHETYLCYVVKRRDSATSFSLDFGYLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTS WSPCYDCARHVADFLRGNPNLSLRIFTARLTGWGALPAGLMSPARPSDYFYCWNTFVENHERTFKAWE GLHENSVRLSRRLRRILLPLYEVDDLRDAFRTLGL (SEQ ID NO: 31)

(underline: nuclear localization sequence; double underline: nuclear export signal)

[00247] Mouse APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKD NI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQIVRFLATHHNLSLDIFSSRLYNVQD PETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPV PSSSSSTLSNICLTKGLPETRFCVEGRRMDPLSEEEFYSQFYNQRVKHLCYYHRMKPYLCYQLEQFNG QAPLKGCLLSEKGKQ*HAEILFLDKIRSMELSQVTITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTS RLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKE SWGLQDLVNDFGNLQLGPPMS (SEQ ID NO: 32)

(italic: nucleic acid editing domain)

[00248] Rat APOBEC-3:

MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLRYAIDRKDTFLCYEVTRKDCDSPVSLHHGVFKNKD NI*HAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFEC*AEQVLRFLATHHNLSLDIFSSRLYNIRD

PENQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKKLLTNFRYQDSKLQEILRPCYIPV
PSSSSSTLSNICLTKGLPETRFCVERRRVHLLSEEEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNG
QAPLKGCLLSEKGKQ*HAEILFLDKIRSMELSQVIITCYLTWSPCPNC*AWQLAAFKRDRPDLILHIYTS
RLYFHWKRPFQKGLCSLWQSGILVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHRIKE
SWGLQDLVNDFGNLQLGPPMS (SEQ ID NO: 33)

(italic: nucleic acid editing domain)

[00249] Rhesus macaque APOBEC-3G:

MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKY*HPEMRFLRWFH*
*KWRQLHHDQEYKVTWYVSWSPCTRC*ANSVATFLAKDPKVTLTIFVARLYYFWKPDYQQALRILCQKRG
GPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNNLPKHYTLLQATLGELLRHLMDPGTFTSNFNNKPW
VSGQHETYLCYKVERLHNDTWVPLNQHRGFLRNQAPNIHGFPKGR*HAELCFLDLIPFWKLDGQQYRVT*
*CFTSWSPCFSC*AQEMAKFISNNEHVSLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWD
TFVDRQGRPFQPWDGLDEHSQALSGRLRAI (SEQ ID NO: 34)

(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

[00250] Chimpanzee APOBEC-3G:

MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSKLKY*HPEM*
*RFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDVATFLAEDPKVTLTIFVARLYYFWDPDYQEALR
SLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTS
NFNNELWVRGRHETYLCYEVERLHNDTWVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLD*
*LHQDYRVTCFTSWSPCFSC*AQEMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTY
SEFKHCWDTFVDHQGCPFQPWDGLEEHSQALSGRLRAILQNQGN (SEQ ID NO: 35)

(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

[00251] Green monkey APOBEC-3G:

MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKLYPEAKD*HPEM*
*KFLHWFRKWRQLHRDQEYEVTWYVSWSPCTRC*ANSVATFLAEDPKVTLTIFVARLYYFWKPDYQQALR
ILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPFKPRKNLPKHYTLLHATLGELLRHVMDPGTFTS
NFNNKPWVSGQRETYLCYKVERSHNDTWVLLNQHRGFLRNQAPDRHGFPKGR*HAELCFLDLIPFWKLD*
*DQQYRVTCFTSWSPCFSC*AQKMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLHRDGAKIAVMNYS
EFEYCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAI

(SEQ ID NO: 36)

(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

[00252] Human APOBEC-3G:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKY*HPEM*
*RFFHWFSKWRKLHRDQEYEVTWYISWSPCTKC*TRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALR

SLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTF
NFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGR*HAELCFLDVIPFWKLD*
*LDQDYRVTCFTSWSPCFSC*AQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTY
SEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 37)

(italic: nucleic acid editing domain; underline: cytoplasmic localization signal)

[00253] Human APOBEC-3F:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEH*HAEM*
*CFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLAEHPNVTLTISAARLYYYWERDYRRALCR
LSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHF
KNLRKAYGRNESWLCFTMEVVKHHSPVSWKRGVFRNQVDPETHC*HAERCFLSWFCDDILSPNTNYEVT*
*WYTSWSPCPEC*AGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCW
ENFVYNDDEPFKPWKGLKYNFLFLDSKLQEILE

(SEQ ID NO: 38)

(italic: nucleic acid editing domain)

[00254] Human APOBEC-3B:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFKPQY*HAE*
*MCFLSWFCGNQLPAYKCFQITWFVSWTPCPDC*VAKLAEFLSEHPNVTLTISAARLYYYWERDYRRALC
RLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKFDENYAFLHRTLKEILRYLMDPDTFTFNFNN
DPLVLRRRQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFY*GRHAELRFLDLVPSLQLDPAQI*
*YRVTWFISWSPCFSWGC*AGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDE
FEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQNQGN (SEQ ID NO: 39)

(italic: nucleic acid editing domain)

[00255] Rat APOBEC-3B:

MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNNFLCYEVNGMDCA
LPVPLRQGVFRKQGHIHAELCFIYWFHDKVLRVLSPMEEFKVTWYMSWSPCSKCAEQVARFLAAHRNL
SLAIFSSRLYYYLRNPNYQQKLCRLIQEGVHVAAMDLPEFKKCWNKFVDNDGQPFRPWMRLRINFSFY
DCKLQEIFSRMNLLREDVFYLQFNNSHRVKPVQNRYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQH
VEILFLEKMRSMELSQVRITCYLTWSPCPNCARQLAAFKKDHPDLILRIYTSRLYFYWRKKFQKGLCT
LWRSGIHVDVMDLPQFADCWTNFVNPQRPFRPWNELEKNSWRIQRRLRRIKESWGL (SEQ ID NO:
40)

[00256] Bovine APOBEC-3B:

DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVLFKQQFGNQPRVPAP
YYRRKTYLCYQLKQRNDLTLDRGCFRNKKQRHAEIRFIDKINSLDLNPSQSYKIICYITWSPCPNCAN

ELVNFITRNNHLKLEIFASRLYFHWIKSFKMGLQDLQNAGISVAVMTHTEFEDCWEQFVDNQSRPFQP
WDKLEQYSASIRRRLQRILTAPI (SEQ ID NO: 41)

[00257] Chimpanzee APOBEC-3B:

MNPQIRNPMEWMYQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAE
MCFLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHPNVTLTISAARLYYYWERDYRRALC
RLSQAGARVKIMDDEEFAYCWENFVYNEGQPFMPWYKFDDNYAFLHRTLKEIIRHLMDPDTFTFNFNN
DPLVLRRHQTYLCYEVERLDNGTWVLMDQHMGFLCNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQI
YRVTWFISWSPCFSWGCAGQVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDE
FEYCWDTFVYRQGCPFQPWDGLEEHSQALSGRLRAILQVRASSLCMVPHRPPPPPQSPGPCLPLCSEP
PLGSLLPTGRPAPSLPFLLTASFSFPPPASLPPLPSLSLSPGHLPVPSFHSLTSCSIQPPCSSRIRET
EGWASVSKEGRDLG (SEQ ID NO: 42)

[00258] Human APOBEC-3C:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETH*CHAE*
*RCFLSWFCDDILSPNTKYQVTWYTSWSPCPDC*AGEVAEFLARHSNVNLTIFTARLYYFQYPCYQEGLR
SLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPWKGLKTNFRLLKRRLRESLQ (SEQ ID NO: 43)

(italic: nucleic acid editing domain)

[00259] Gorilla APOBEC3C:

MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETHCHAE

RCFLSWFCDDILSPNTNYQVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTARLYYFQDTDYQEGLR

SLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPWKGLKYNFRFLKRRLQEILE (SEQ ID NO: 44)

(italic: nucleic acid editing domain)

[00260] Human APOBEC-3A:

MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYG
R*HAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGC*AGEVRAFLQENTHVRLRIFAARIYDYDPLY
KEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN
(SEQ ID NO: 45)

(italic: nucleic acid editing domain)

[00261] Rhesus macaque APOBEC-3A:

MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRGFLCNKAKNVPCG

DYGCHVELRFLCEVPSWQLDPAQTYRVTWFISWSPCFRRGCAGQVRVFLQENKHVRLRIFAARIYDYD

PLYQEALRTLRDAGAQVSIMTYEEFKHCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAILQNQGN

(SEQ ID NO: 46)

(italic: nucleic acid editing domain)

[00262] Bovine APOBEC-3A:

MDEYTFTENFNNQGWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPCHAELYFLGKIHSW

NLDRNQHYRLTCFISWSPCYDCAQKLTTFLKENHHISLHILASRIYTHNRFGCHQSGLCELQAAGARI

TIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCTELQAILKTQQN (SEQ ID NO: 47)

(italic: nucleic acid editing domain)

[00263] Human APOBEC-3H:

MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKC*HAEICFINEIKSMGL*
*DETQCYQVTCYLTWSPCSSC*AWELVDFIKAHDHLNLGIFASRLYYHWCKPQQKGLRLLCGSQVPVEVM
GFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV
(SEQ ID NO: 48) (italic: nucleic acid editing domain)

[00264] Rhesus macaque APOBEC-3H:

MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIRFINKIKSMGL
DETQCYQVTCYLTWSPCPSCAGELVDFIKAHRHLNLRIFASRLYYHWRPNYQEGLLLLCGSQVPVEVM
GLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAIKRRLERIKSRSVDVLENGLRSLQLGPVTPSS
SIRNSR

(SEQ ID NO: 49)

[00265] Human APOBEC-3D:

MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHR
QEVYFRFEN*HAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPC*VVKVTKFLAEHPNVTLTISAARLY
YYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNP
MEAMYPHIFYFHFKNLLKACGRNESWLCFTMEVTKHHSAVFRKRGVFRNQVDPETHC*HAERCFLSWFC*
*DDILSPNTNYEVTWYTSWSPCPEC*AGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGAS
VKIMGYKDFVSCWKNFVYSDDEPFKPWKGLQTNFRLLKRRLREILQ (SEQ ID NO: 50)

(italic: nucleic acid editing domain)

[00266] Human APOBEC-1:

MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIK
KFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARLFWHMDQQNRQGLRDLVN

SGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLT
FFRLHLQNCHYQTIPPHILLATGLIHPSVAWR (SEQ ID NO: 51)

[00267] Mouse APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSNHVEVNFLE
KFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARLYHHTDQRNRQGLRDLIS
SGVTIQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYVLELYCIILGLPPCLKILRRKQPQLT
FFTITLQTCHYQRIPPHLLWATGLK (SEQ ID NO: 52)

[00268] Rat APOBEC-1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLIS
SGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLT
FFTIALQSCHYQRLPPHILWATGLK (SEQ ID NO: 53)

[00269] Human APOBEC-2:

MAQKEEAAVATEAASQNGEDLENLDDPEKLKELIELPPFEIVTGERLPANFFKFQFRNVE
YSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNVTWYVSSSPC
AACADRIIKTLSKTKNLRLLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEE
GESKAFQPWEDIQENFLYYEEKLADILK (SEQ ID NO: 54)

[00270] Mouse APOBEC-2:

MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKT
FLCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRIL
KTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEP
WEDIQENFLYYEEKLADILK

(SEQ ID NO: 55)

[00271] Rat APOBEC-2:

MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKT
FLCYVVEAQSKGGQVQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNVTWYVSSSPCAACADRIL
KTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYLWQNFVEQEEGESKAFEP
WEDIQENFLYYEEKLADILK

(SEQ ID NO: 56)

[00272] Bovine APOBEC-2:

MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVE
YSSGRNKTFLCYVVEAQSKGGQVQASRGYLEDEHATNHAEEAFFNSIMPTFDPALRYMVTWYVSSSPC
AACADRIVKTLNKTKNLRLLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEE
GESKAFEPWEDIQENFLYYEEKLADILK (SEQ ID NO: 57)

[00273] Petromyzon marinus CDA1 (pmCDA1):

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAE
IFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTLKIWACKLYYEKNARNQI
GLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTK
SPAV (SEQ ID NO: 58)

[00274] Human APOBEC3G D316R_D317R:

MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEM
RFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIFVARLYYFWDPDYQEALR
SLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTF
NFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLD
LDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISIMTY
SEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 59)

[00275] Human APOBEC3G chain A:

MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLD
VIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAG
AKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ (SEQ ID NO: 60)

[00276] Human APOBEC3G chain A D120R_D121R:

MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLD
VIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAG
AKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ (SEQ ID NO: 61)

Adenosine deaminases

[00277] The disclosure provides fusion proteins that comprise one or more adenosine deaminases. In some aspects, such fusion proteins are capable of deaminating adenosine in a nucleic acid sequence (*e.g.,* DNA or RNA). As one example, any of the fusion proteins provided herein may be base editors, (*e.g.,* adenine base editors). Without wishing to be bound by any particular theory, dimerization of adenosine deaminases (*e.g.,* in cis or in trans) may improve the ability (*e.g.,* efficiency) of the fusion protein to modify a nucleic acid base, for example to deaminate adenine. In some embodiments, any of the fusion proteins may comprise 2, 3, 4 or 5 adenosine deaminases. In some embodiments, any of the fusion proteins provided herein comprise two adenosine deaminases. Exemplary, non-limiting, embodiments of adenosine deaminases are provided herein. It should be appreciated that the mutations provided herein (*e.g.*, mutations in ecTadA) may be applied to adenosine deaminases in

other adenosine base editors, for example those provided in U.S. Patent Publication No. 2018/0073012, published March 15, 2018, which issued as U.S. Patent No. 10,113,163, on October 30, 2018; U.S. Patent Publication No. 2017/0121693, published May 4, 2017, which issued as U.S. Patent No. 10,167,457 on January 1, 2019; International Publication No. WO 2017/070633, published April 27, 2017; U.S. Patent Publication No. 2015/0166980, published June 18, 2015; U.S. Patent No. 9,840,699, issued December 12, 2017; and U.S. Patent No. 10,077,453, issued September 18, 2018, all of which are incorporated herein by reference in their entireties.

[00278] In some embodiments, any of the adenosine deaminases provided herein is capable of deaminating adenine. In some embodiments, the adenosine deaminases provided herein are capable of deaminating adenine in a deoxyadenosine residue of DNA. The adenosine deaminase may be derived from any suitable organism (*e.g., E. coli*). In some embodiments, the adenosine deaminase is a naturally-occurring adenosine deaminase that includes one or more mutations corresponding to any of the mutations provided herein (*e.g.,* mutations in ecTadA). One of skill in the art will be able to identify the corresponding residue in any homologous protein and in the respective encoding nucleic acid by methods well known in the art, *e.g.*, by sequence alignment and determination of homologous residues. Accordingly, one of skill in the art would be able to generate mutations in any naturally-occurring adenosine deaminase (*e.g.,* having homology to ecTadA) that corresponds to any of the mutations described herein, *e.g.*, any of the mutations identified in ecTadA. In some embodiments, the adenosine deaminase is from a prokaryote. In some embodiments, the adenosine deaminase is from a bacterium. In some embodiments, the adenosine deaminase is from *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi, Shewanella putrefaciens, Haemophilus influenzae, Caulobacter crescentus,* or *Bacillus subtilis.* In some embodiments, the adenosine deaminase is from *E. coli*.

[00279] In some embodiments, the adenosine deaminase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of

the amino acid sequences set forth in any one of SEQ ID NOs: 62-84, or to any of the adenosine deaminases provided herein. It should be appreciated that adenosine deaminases provided herein may include one or more mutations (*e.g.*, any of the mutations provided herein). The disclosure provides adenosine deaminases with a certain percent identity plus any of the mutations or combinations thereof described herein. In some embodiments, the adenosine deaminase comprises an amino acid sequence that has 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more mutations compared to any one of the amino acid sequences set forth in SEQ ID NOs: 62-84, or any of the adenosine deaminases provided herein. In some embodiments, the adenosine deaminase comprises an amino acid sequence that has at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, or at least 170 identical contiguous amino acid residues as compared to any one of the amino acid sequences set forth in SEQ ID NOs: 62-84, or any of the adenosine deaminases provided herein.

[00280] In some embodiments, the adenosine deaminase comprises an E59X mutation in ecTadA SEQ ID NO: 64, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In particular embodiments, the adenosine deaminase comprises a E59A mutation in SEQ ID NO: 64, or a corresponding mutation in another adenosine deaminase.

[00281] In some embodiments, the adenosine deaminase comprises a D108X mutation in ecTadA SEQ ID NO: 64, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises a D108W, D108Q, D108F, D108K, or D108M mutation in SEQ ID NO: 64, or a corresponding mutation in another adenosine deaminase. In particular embodiments, the adenosine deaminase comprises a D108W mutation in SEQ ID NO: 64,

or a corresponding mutation in another adenosine deaminase. It should be appreciated, however, that additional deaminases may similarly be aligned to identify homologous amino acid residues that may be mutated as provided herein.

[00282] In some embodiments, the adenosine deaminase comprises TadA 7.10, whose sequence is provided as SEQ ID NO: 65, or a variant thereof. TadA7.10 comprises the following mutations in ecTadA: W23R, H36L, P48A, R51L, L84F, A106V, D108N, H123Y, S146C, D147Y, R152P, E155V, I156F, K157N.

[00283] In particular embodiments, the adenosine deaminase comprises an N108W mutation in SEQ ID NO: 65, an embodiment also referred to as TadA 7.10(N108W). Its sequence is provided as SEQ ID NO: 67.

[00284] In some embodiments, the adenosine deaminase comprises an A106X mutation in ecTadA SEQ ID NO: 64, or a corresponding mutation in another adenosine deaminase, where X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A106V mutation in SEQ ID NO: 64, or a corresponding mutation in another adenosine deaminase. In some embodiments, the adenosine deaminase comprises an A106Q, A106F, A106W, or A106M mutation in SEQ ID NO: 64, or a corresponding mutation in another adenosine deaminase.

[00285] In particular embodiments, the adenosine deaminase comprises a V106W mutation in SEQ ID NO: 65, an embodiment also referred to as TadA 7.10(V106W). Its sequence is provided as SEQ ID NO: 66.

[00286] In some embodiments, the adenosine deaminase comprises a R47X mutation in SEQ ID NO: 65, or a corresponding mutation in another adenosine deaminase, where the presence of X indicates any amino acid other than the corresponding amino acid in the wild-type adenosine

deaminase. In some embodiments, the adenosine deaminase comprises a R47Q, R47F, R47W, or R47M mutation in SEQ ID NO: 65, or a corresponding mutation in another adenosine deaminase.

[00287] In particular embodiments, the adenosine deaminase comprises a R47Q, R47F, R47W, or R47M mutation in SEQ ID NO: 65.

[00288] In particular embodiments, the adenosine deaminase comprises a V106Q mutation and an N108W mutation in SEQ ID NO: 65. In particular embodiments, the adenosine deaminase comprises a V106W mutation, an N108W mutation and an R47Z mutation, wherein Z is selected from the residues consisting of Q, F, W and M, in SEQ ID NO: 65.

[00289] It should be appreciated that any of the mutations provided herein (*e.g.,* based on the ecTadA amino acid sequence of SEQ ID NO: 64) may be introduced into other adenosine deaminases, such as *S. aureus* TadA (saTadA), or other adenosine deaminases (*e.g.,* bacterial adenosine deaminases), such as those sequences provided below. It would be apparent to the skilled artisan how to identify amino acid residues from other adenosine deaminases that are homologous to the mutated residues in ecTadA. Thus, any of the mutations identified in ecTadA may be made in other adenosine deaminases that have homologous amino acid residues. It should also be appreciated that any of the mutations provided herein may be made individually or in any combination in ecTadA or another adenosine deaminase. For example, an adenosine deaminase may contain a D108N, an A106V, and/or a R47Q mutation in ecTadA SEQ ID NO: 64, or a corresponding mutation in another adenosine deaminase.

[00290] In some embodiments, the adenosine deaminase comprises one, two, or three mutations selected from the group consisting of D108, A106, and R47 in SEQ ID NO: 64, or a corresponding mutation or mutations in another adenosine deaminase.

[00291] In other aspects, the disclosure provides adenine base editors with broadened target sequence compatibility. In general, native ecTadA deaminates the adenine in the sequence UAC (*e.g.*, the target sequence) of the anticodon loop of $tRNA^{Arg}$. Without wishing to be bound by any particular

theory, in order to expand the utility of ABEs comprising one or more ecTadA deaminases, such as any of the adenosine deaminases provided herein, the adenosine deaminase proteins were optimized to recognize a wide variety of target sequences within the protospacer sequence without compromising the editing efficiency of the adenosine nucleobase editor complex. In some embodiments, the target sequence is an A in the middle of a 5'-NAN-3' sequence, wherein N is T, C, G, or A. In some embodiments, the target sequence comprises 5'-TAC-3'. In some embodiments, the target sequence comprises 5'-GAA-3'.

[00292] In some embodiments, the adenosine deaminase is an N-terminal truncated *E. coli* TadA. In certain embodiments, the adenosine deaminase comprises the amino acid sequence:

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGI LADECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 77).

[00293] In some embodiments, the TadA deaminase is a full-length *E. coli* TadA deaminase (ecTadA). For example, in certain embodiments, the adenosine deaminase comprises the amino acid sequence:

MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAH AEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPG MNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 78)

[00294] It should be appreciated, however, that additional adenosine deaminases useful in the present application would be apparent to the skilled artisan and are within the scope of this disclosure. For example, the adenosine deaminase may be a homolog of an ADAT. Exemplary ADAT homologs include, without limitation:

[00295] *Staphylococcus aureus* TadA:

MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAK VLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMNLLQQSNFNHRAIVDKGV LKEACSTLLTTFFKNLRANKKSTN (SEQ ID NO: 79)

[00296] *Bacillus subtilis* TadA:

MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACKALGT
WRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNHQAEVVSGVLEEE
CGGMLSAFFRELRKKKKAARKNLSE (SEQ ID NO: 63)

[00297] *Salmonella typhimurium* (*S. typhimurium*) TadA:

MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDPTAH
AEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARDAKTGAAGSLIDVLHHPG
MNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV (SEQ ID NO: 80)

[00298] *Shewanella putrefaciens* (*S. putrefaciens*) TadA:

MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGKKLENYR
LLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPAFNHQVEVTSGVLAEACS
AQLSRFFKRRRDEKKALKLAQRAQQGIE (SEQ ID NO: 81)

[00299] *Haemophilus influenzae* F3031 (*H. influenzae*) TadA:

MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHAEIIALR
NGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIGSRFHFFDDYKMNHTLEI
TSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK (SEQ ID NO: 82)

[00300] *Caulobacter crescentus* (*C. crescentus*) TadA:

MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIAAMR
AAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVVHGPKFFAQPTCHWRPEV
TGGVLADESADLLRGFFRARRKAKI (SEQ ID NO: 83)

[00301] *Geobacter sulfurreducens* (*G. sulfurreducens*) TadA:

MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHAEMIAIR
QAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAGSLYDLSADPRLNHQVRL
SPGVCQEECGTMLSDFFRDLRRRKKAKATPALFIDERKVPPEP (SEQ ID NO: 84)

[00302] Exemplary adenosine deaminase variants of the disclosure are described below. In certain embodiments, the adenosine deaminase has a sequence with at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or at least 99.5% sequence identity to one of the following:

[00303] *(Ec)TadA, catalytically inactive (E59A)*

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAAIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGI
LADECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 64)

[00304] *TadA 7.10*

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGI
LADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 65)

[00305] *TadA 7.10 (V106W)*

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGWRNAKTGAAGSLMDVLHYPGMNHRVEITEGI
LADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 66)

[00306] *TadA 7.10 (N108W)*

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRWAKTGAAGSLMDVLHYPGMNHRVEITEGI
LADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 67)

[00307] *TadA 7.10 (N108Q)*

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRQAKTGAAGSLMDVLHYPGMNHRVEITEGI
LADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 68)

[00308] *TadA 7.10 (V106F)*

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGFRNAKTGAAGSLMDVLHYPGMNHRVEITEGI
LADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 69)

[00309] *TadA 7.10 (V106Q)*

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGQRNAKTGAAGSLMDVLHYPGMNHRVEITEGI
LADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 70)

[00310] *TadA 7.10 (V106M)*

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGMRNAKTGAAGSLMDVLHYPGMNHRVEITEGI
LADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 71)

[00311] *TadA 7.10 (R47F)*

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNFAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGI
LADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 72)

[00312] *TadA 7.10 (R47W)*

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNWAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGI
LADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 73)

[00313] *TadA 7.10 (R47Q)*

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNQAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGI
LADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 74)

[00314] *TadA 7.10 (R47M)*

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNMAIGLHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGI
LADECAALLCYFFRMPRQVFNAQKKAQSSTD (SEQ ID NO: 75)

[00315] *TadA (E59Q)*

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAQIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGI
LADECAALLSDFFRMRRQEIKAQKKAQSSTD (SEQ ID NO: 76)

[00316] Any two or more of the adenosine deaminases described herein may be connected to one another (*e.g.* by a linker) within an adenosine deaminase domain of the fusion proteins provided herein. For instance, the fusion proteins provided herein may contain only two adenosine deaminases. In some embodiments, the adenosine deaminases are the same. In some embodiments, the adenosine deaminases are any of the adenosine deaminases provided herein. In some embodiments, the adenosine deaminases are different. In some embodiments, the first adenosine deaminase is any of the adenosine deaminases provided herein, and the second adenosine is any of the adenosine deaminases provided herein, but is not identical to the first adenosine deaminase. In some embodiments, the fusion protein comprises two adenosine deaminases (*e.g.,* a first adenosine deaminase and a second adenosine deaminase). In some embodiments, the fusion protein comprises a first adenosine deaminase and a second adenosine deaminase. In some embodiments, the first adenosine deaminase is N-terminal to the second adenosine deaminase in the fusion protein. In some embodiments, the first adenosine deaminase is C-terminal to the second adenosine deaminase in the fusion protein. In some

embodiments, the first adenosine deaminase and the second deaminase are fused directly or via a linker.

[00317] In particular embodiments, the base editors disclosed herein comprise a heterodimer of a first adenosine deaminase that is N-terminal to a second adenosine deaminase, wherein the first adenosine deaminase comprises a sequence with at least 80%, 85%, 90%, 95%, 98%, 99%, or 99.5% sequence identity to SEQ ID NO: 62-84; and the second adenosine deaminase comprises a sequence with at least 80%, 85%, 90%, 95%, 98%, 99%, or 99.5% sequence identity to SEQ ID NO: 62-84.

[00318] In other embodiments, the second adenosine deaminase of the base editors provided herein comprises a sequence with at least 80%, 85%, 90%, 95%, 98%, 99%, or 99.5% sequence identity to SEQ ID NO: 65 (TadA 7.10), wherein any sequence variation may only occur in amino acid positions other than R47, V106 or N108 of SEQ ID NO: 65. In other words, these embodiments must contain amino acid substitutions at R47, V106 or N108 of SEQ ID NO: 65.

[00319] In other embodiments, the second adenosine deaminase of the heterodimer comprises a sequence with at least 80%, 85%, 90%, 95%, 98%, 99%, or 99.5% sequence identity to SEQ ID NO: 62-84.

Base editor constructs

[00320] Any of the Cas9 domains (e.g., Cas9 domains that recognize a non-canonical PAM sequence) disclosed herein may be fused to a second protein, thus providing fusion proteins that comprise a Cas9 domain as provided herein and a second protein, or a "fusion partner." In some embodiments, the second protein is an effector domain. As used herein, an "effector domain" refers to a molecule (e.g., a protein) that regulates a biological activity and/or is capable of modifying a biological molecule (e.g., a protein, or a nucleic acid such as DNA or RNA). In some embodiments, the effector domain is a protein. In some embodiments, the effector domain is capable of modifying a protein (e.g., a histone). In some embodiments, the effector domain is capable of modifying DNA (e.g., genomic DNA). In some embodiments, the effector domain is capable of modifying RNA

(e.g., mRNA). In some embodiments, the effector molecule is a nucleic acid editing domain. In some embodiments, the effector molecule is capable of regulating an activity of a nucleic acid (e.g., transcription, and/or translation). Exemplary effector domains include, without limitation, a deaminase, a nuclease, a nickase, a recombinase, a methyltransferase, a methylase, an acetylase, an acetyltransferase, a transcriptional activator, or a transcriptional repressor domain. In some embodiments, the effector domain is a nucleic acid editing domain. Some aspects of the disclosure provide fusion proteins comprising a Cas9 domain and a nucleic acid editing domain.

[00321] In some embodiments, the fusion proteins provided herein exhibit increased activity on a target sequence that does not comprise the canonical PAM (5′-NGG-3′) at its 3′ end as compared to a fusion protein comprising *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2. In some embodiments, the fusion protein exhibits an activity on a target sequence having a 3′ end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 2-fold, at least 3-fold, at least 4-fold, at least 5-fold, at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold increased as compared to the activity of a fusion protein comprising *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2 on the same target sequence. In some embodiments, the 3′ end of the target sequence is directly adjacent to a sequence selected from the group consisting of NGT, NGA, NGC, and NNG, wherein N is an A, G, T, or C. In some embodiments, the 3′ end of the target sequence is directly adjacent to a sequence selected from the group consisting of CGG, AGT, TGG, AGT, CGT, GGG, CGT, TGT, GGT, AGC, CGC, TGC, GGC, AGA, CGA, TGA, GGA, GAA, GAT, and CAA. In some embodiments, the fusion protein activity is measured by a nuclease assay, a deamination assay, a transcriptional activation assay, a binding assay, PCR, or sequencing. In some embodiments, the transcriptional activation assay is a GFP activation assay. In some embodiments, sequencing is used to measure indel formation. In some

embodiments, the increased activity is increased binding. In some embodiments, the increased activity is increased deamination of a nucleobase in the target sequence.

[00322] Some aspects of the disclosure provide a fusion protein comprising a Cas9 domain fused to a nucleic acid editing domain, wherein the nucleic acid editing domain is fused to the N-terminus of the Cas9 domain. In some embodiments, the nucleic acid editing domain is fused to the C-terminus of the Cas9 domain. In some embodiments, the Cas9 domain and the nucleic acid editing-editing domain are fused via a linker. In some embodiments, the linker comprises a `(GGGS)n` (SEQ ID NO: 93), a `(GGGGS)n` (SEQ ID NO: 95), a `(G)n` (SEQ ID NO: 97), an `(EAAAK)n` (SEQ ID NO: 99), a `(GGS)n` (SEQ ID NO: 101), $(SGGS)_n$ (SEQ ID NO: 91), an `SGSETPGTSESATPES` (SEQ ID NO: 89) motif (see, *e.g.*, Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 2014**;** 32(6): 577-82; the entire contents are incorporated herein by reference), or a combination of any of these, wherein n is independently an integer between 1 and 30. In some embodiments, n is independently 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30, or, if more than one linker or more than one linker motif is present, any combination thereof. In some embodiments, the linker comprises a `(GGS)n` motif (SEQ ID NO: 101), wherein n is 1,2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15. Additional suitable linker motifs and linker configurations will be apparent to those of ordinary skill in the art (e.g., SEQ ID NOs: 89-112). In some embodiments, suitable linker motifs and configurations include those described in Chen *et al.*, Fusion protein linkers: property, design and functionality. *Adv. Drug Deliv. Rev.* 2013; 65(10):1357-69, the entire contents of which are incorporated herein by reference. Additional suitable linker sequences will be apparent to those of ordinary skill in the art based on the instant disclosure. In some embodiments, the general architecture of exemplary Cas9 fusion proteins provided herein comprises the structure:

$[NH_2]$-[nucleic acid editing domain]-[Cas9 domain]-[COOH];

$[NH_2]$-[nucleic acid editing domain]-[linker]-[Cas9 domain]-[COOH];

[NH2]-[Cas9 domain]-[nucleic acid editing domain]-[COOH]; or

[NH2]-[Cas9 domain]-[linker]-[nucleic acid editing domain]-[COOH],

wherein NH_2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. In some embodiments, the "]-[" used in the general architecture above indicates the presence of an optional linker sequence.

[00323] The fusion proteins of the present disclosure may comprise one or more additional features. For example, in some embodiments, the fusion protein comprises a nuclear localization sequence (NLS). In some embodiments, the NLS of the fusion protein is localized between the nucleic acid editing domain and the Cas9 domain. In some embodiments, the NLS of the fusion protein is localized C-terminal to the Cas9 domain. In some embodiments, the NLS of the fusion protein is localized N-terminal to the Cas9 domain. In some embodiments, the NLS comprises the amino acid sequence of SEQ ID NO: 113 or 114. In some embodiments, the NLS comprises the amino acid sequence of SEQ ID NO: 113.

[00324] Other exemplary features that may be present are localization sequences, such as cytoplasmic localization sequences, export sequences, such as nuclear export sequences, or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable protein tags provided herein include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin (HA)-tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags (e.g., Softag 1, Softag 3), strep-tags , biotin ligase tags, FlAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of ordinary skill in the art. In some embodiments, the fusion protein comprises one or more His tags.

[00325] In some embodiments, the nucleic acid editing domain is a deaminase. In some embodiments, the deaminase is a cytidine deaminase. For example, in some embodiments, the general

architecture of exemplary Cas9 fusion proteins with a cytidine deaminase domain comprises the structure:

[NH_2]-[NLS]-[cytidine deaminase]-[Cas9]-[COOH];

[NH_2]-[Cas9]-[cytidine deaminase]-[COOH];

[NH_2]-[cytidine deaminase]-[Cas9]-[COOH]; or

[NH_2]-[cytidine deaminase]-[Cas9]-[NLS]-[COOH],

wherein NLS is a nuclear localization sequence, NH_2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. Nuclear localization sequences are known in the art and would be apparent to the skilled artisan. For example, NLS sequences are described in Plank *et al.*, International PCT Application, PCT/EP2000/011690, the contents of which are incorporated herein by reference for their disclosure of exemplary nuclear localization sequences. In some embodiments, a NLS comprises the amino acid sequence PKKKRKV (SEQ ID NO: 113) or MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 114). In some embodiments, a linker is inserted between the Cas9 and the cytidine deaminase. In some embodiments, the NLS is located C-terminal of the Cas9 domain. In some embodiments, the NLS is located N-terminal of the Cas9 domain. In some embodiments, the NLS is located between the cytidine deaminase and the Cas9 domain. In some embodiments, the NLS is located N-terminal of the cytidine deaminase domain. In some embodiments, the NLS is located C-terminal of the cytidine deaminase domain. In some embodiments, the "]-[" used in the general architecture above indicates the presence of an optional linker sequence.

[00326] In some embodiments, the fusion protein comprises any one of nucleic acid editing domains provided herein. In some embodiments, the nucleic acid editing domain is a cytidine or adenosine deaminase domain provided herein.

[00327] In some embodiments, the cytidine deaminase domain and the Cas9 domain are fused to each other via a linker. Various linker lengths and flexibilities between the deaminase domain (*e.g.*,

AID, APOBEC family deaminase) and the Cas9 domain can be employed, for example, ranging from very flexible linkers of the form (GGGS)n (SEQ ID NO: 93), (GGGGS)n (SEQ ID NO: 95), (GGS)n (SEQ ID NO: 101), and (G)n (SEQ ID NO: 97), to more rigid linkers of the form (EAAAK)n (SEQ ID NO: 99), (SGGS)n (SEQ ID NO: 91), SGGS(GGS)n (SEQ ID NO: 103), SGSETPGTSESATPES (SEQ ID NO: 89) (see, *e.g.*, Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 2014; 32(6): 577-82; the entire contents are incorporated herein by reference), (SGGS)n-SGSETPGTSESATPES-(SGGS)n (SEQ ID NO: 98), and $(XP)_n$, wherein n is an integer between 1 and 30, inclusive, in order to achieve the optimal length for deaminase activity for the specific application. In some embodiments, the linker comprises a (GGS)n motif, wherein n is 1, 3, or 7. In some embodiments, the linker comprises a SGSETPGTSESATPES (SEQ ID NO: 89) motif. In some embodiments, the linker comprises a (SGGS)2-SGSETPGTSESATPES-(SGGS)2 (SEQ ID NO: 96) motif.

[00328] In some embodiments, the fusion protein comprises a Cas9 domain (*e.g.*, a Cas9 domain comprising one or more mutations that recognizes a non-canonical PAM sequence) fused to a cytidine deaminase domain, wherein the fusion protein comprises the amino acid sequence of SEQ ID NO: 2. In some embodiments, the fusion protein comprises any one of the amino acid sequences of SEQ ID NOs: 122-132.

[00329] Some aspects of the disclosure relate to fusion proteins that comprise a uracil glycosylase inhibitor (UGI) domain. In some embodiments, any of the fusion proteins provided herein that comprise a Cas9 domain (*e.g.,* a Cas9 domain comprising one or more mutations that recognizes a non-canonical PAM sequence) may be further fused to a UGI domain either directly or via a linker. Some aspects of this disclosure provide deaminase-dCas9 fusion proteins, deaminase-nuclease active Cas9 fusion proteins and deaminase-Cas9 nickase fusion proteins with increased nucleobase editing efficiency. Without wishing to be bound by any particular theory, cellular DNA-repair response to the

presence of U:G heteroduplex DNA may be responsible for the decrease in nucleobase editing efficiency in cells. For example, uracil DNA glycosylase (UDG) catalyzes removal of U from DNA in cells, which may initiate base excision repair, with reversion of the U:G pair to a C:G pair as the most common outcome. A Uracil DNA Glycosylase Inhibitor (UGI) may inhibit human UDG activity. Thus, this disclosure contemplates a fusion protein comprising a Cas9 domain and a nucleic acid editing domain (*e.g.*, a deaminase) further fused to a UGI domain. In some embodiments, the fusion protein comprising a Cas9 nickase-nucleic acid editing domain further fused to a UGI domain. In some embodiments, the fusion protein comprising a dCas9-nucleic acid editing domain further fused to a UGI domain. It should be understood that the use of a UGI domain may increase the editing efficiency of a nucleic acid editing domain that is capable of catalyzing, for example, a C to U change. For example, fusion proteins comprising a UGI domain may be more efficient in deaminating C residues.

[00330] In some embodiments, the fusion protein comprises the structure:

[nucleic acid editing domain]-[optional linker sequence]-[Cas9]-[optional linker sequence]-[UGI];

[nucleic acid editing domain]-[optional linker sequence]-[UGI]-[optional linker sequence]-[Cas9];

[UGI]-[optional linker sequence]-[nucleic acid editing domain]-[optional linker sequence]-[Cas9];

[UGI]-[optional linker sequence]-[Cas9]-[optional linker sequence]-[nucleic acid editing domain];

[Cas9]-[optional linker sequence]-[nucleic acid editing domain]-[optional linker sequence]-[UGI]; or

[Cas9]-[optional linker sequence]-[UGI]-[optional linker sequence]-[nucleic acid editing domain].

[00331] In some embodiments, the fusion protein comprises the structure:

[deaminase]-[optional linker sequence]-[Cas9]-[optional linker sequence]-[UGI];

[deaminase]-[optional linker sequence]-[UGI]-[optional linker sequence]-[Cas9];

[UGI]-[optional linker sequence]-[deaminase]-[optional linker sequence]-[Cas9];

[UGI]-[optional linker sequence]-[Cas9]-[optional linker sequence]-[deaminase];

[Cas9]-[optional linker sequence]-[deaminase]-[optional linker sequence]-[UGI]; or

[Cas9]-[optional linker sequence]-[UGI]-[optional linker sequence]-[deaminase].

[00332] In some embodiments, the fusion protein comprises the structure:

[cytidine deaminase]-[optional linker sequence]-[Cas9]-[optional linker sequence]-[UGI];

[cytidine deaminase]-[optional linker sequence]-[UGI]-[optional linker sequence]-[Cas9];

[UGI]-[optional linker sequence]-[cytidine deaminase]-[optional linker sequence]-[Cas9];

[UGI]-[optional linker sequence]-[Cas9]-[optional linker sequence]-[cytidine deaminase];

[Cas9]-[optional linker sequence]-[cytidine deaminase]-[optional linker sequence]-[UGI]; or

[Cas9]-[optional linker sequence]-[UGI]-[optional linker sequence]-[cytidine deaminase].

[00333] In some embodiments, the fusion proteins provided herein do not comprise a linker sequence. In some embodiments, one or both of the optional linker sequences are present.

[00334] In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker sequence. In some embodiments, the fusion proteins comprising a UGI domain further comprise a nuclear targeting sequence, for example, a nuclear localization sequence. In some embodiments, fusion proteins provided herein further comprise a nuclear localization sequence (NLS). In some embodiments, the NLS is fused to the N-terminus of the fusion protein. In some embodiments, the NLS is fused to the C-terminus of the fusion protein. In some embodiments, the NLS is fused to the N-terminus of the UGI protein. In some embodiments, the NLS is fused to the C-terminus of the UGI protein. In some embodiments, the NLS is fused to the N-terminus of the Cas9 domain. In some embodiments, the NLS is fused to the C-terminus of the Cas9 domain. In some embodiments, the NLS is fused to the N-terminus of the deaminase. In some embodiments, the NLS is

fused to the C-terminus of the deaminase. In some embodiments, the NLS is fused to the N-terminus of the second Cas9. In some embodiments, the NLS is fused to the C-terminus of the second Cas9. In some embodiments, the NLS is fused to the fusion protein via one or more linkers. In some embodiments, the NLS is fused to the fusion protein without a linker. In some embodiments, the NLS comprises an amino acid sequence of any one of the NLS sequences provided or referenced herein. In some embodiments, the NLS comprises an amino acid sequence as set forth in SEQ ID NO: 113 or SEQ ID NO: 114.

[00335] In some embodiments, a UGI domain comprises a wild-type UGI or a UGI as set forth in any of SEQ ID NOs: 115-120. In some embodiments, the UGI proteins provided herein include fragments of UGI and proteins homologous to a UGI or a UGI fragment. For example, in some embodiments, a UGI domain comprises a fragment of the amino acid sequence set forth in SEQ ID NO: 115. In some embodiments, a UGI fragment comprises an amino acid sequence that comprises at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid sequence as set forth in SEQ ID NO: 115. In some embodiments, a UGI comprises an amino acid sequence homologous to the amino acid sequence set forth in SEQ ID NO: 115 or an amino acid sequence homologous to a fragment of the amino acid sequence set forth in SEQ ID NO: 115. In some embodiments, proteins comprising UGI or fragments of UGI or homologs of UGI or UGI fragments are referred to as "UGI variants." A UGI variant shares homology to UGI, or a fragment thereof. For example a UGI variant is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, at least 99% identical, at least 99.5% identical, or at least 99.9% identical to a wild type UGI or a UGI as set forth in SEQ ID NO: 115. In some embodiments, the UGI variant comprises a fragment of UGI, such that the fragment is at least 70% identical, at least 80% identical, at least 90% identical, at least 95% identical, at least 96% identical, at

least 97% identical, at least 98% identical, at least 99% identical, at least 99.5% identical, or at least 99.9% to the corresponding fragment of wild-type UGI or a UGI as set forth in SEQ ID NO: 115.

[00336] Suitable UGI protein and nucleotide sequences are provided herein and additional suitable UGI sequences are known to those in the art, and include, for example, those published in Wang *et al.*, Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. *J. Biol. Chem.* 264:1163-1171(1989); Lundquist *et al.*, Site-directed mutagenesis and characterization of uracil-DNA glycosylase inhibitor protein. Role of specific carboxylic amino acids in complex formation with Escherichia coli uracil-DNA glycosylase. *J. Biol. Chem.* 272:21408-21419(1997); Ravishankar *et al.*, X-ray analysis of a complex of Escherichia coli uracil DNA glycosylase (EcUDG) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic UDG. *Nucleic Acids Res.* 26:4880-4887(1998); and Putnam *et al.*, Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with Escherichia coli uracil-DNA glycosylase. *J. Mol. Biol.* 287:331-346(1999), the entire contents of each of which are incorporated herein by reference.

[00337] It should be appreciated that additional proteins may be uracil glycosylase inhibitors. For example, other proteins that are capable of inhibiting (*e.g.*, sterically blocking) a uracil-DNA glycosylase base-excision repair enzyme are within the scope of this disclosure. Additionally, any proteins that block or inhibit base-excision repair as also within the scope of this disclosure. In some embodiments, a protein that binds DNA is used. In another embodiment, a substitute for UGI is used. In some embodiments, a uracil glycosylase inhibitor is a protein that binds single-stranded DNA. For example, a uracil glycosylase inhibitor may be a *Erwinia tasmaniensis* single-stranded binding protein. In some embodiments, the single-stranded binding protein comprises the amino acid sequence (SEQ ID NO: 118). In some embodiments, a uracil glycosylase inhibitor is a protein that binds uracil. In some embodiments, a uracil glycosylase inhibitor is a protein that binds uracil in DNA. In some embodiments, a uracil glycosylase inhibitor is a catalytically inactive uracil DNA-

glycosylase protein. In some embodiments, a uracil glycosylase inhibitor is a catalytically inactive uracil DNA-glycosylase protein that does not excise uracil from the DNA. For example, a uracil glycosylase inhibitor is a UdgX. In some embodiments, the UdgX comprises the amino acid sequence (SEQ ID NO: 119). As another example, a uracil glycosylase inhibitor is a catalytically inactive UDG. In some embodiments, a catalytically inactive UDG comprises the amino acid sequence (SEQ ID NO: 55). It should be appreciated that other uracil glycosylase inhibitors would be apparent to the skilled artisan and are within the scope of this disclosure. In some embodiments, a uracil glycosylase inhibitor is a protein that is homologous to any one of SEQ ID NOs: 115-120. In some embodiments, a uracil glycosylase inhibitor is a protein that is at least 50% identical, at least 55% identical, at least 60% identical, at least 65% identical, at least 70% identical, at least 75% identical, at least 80% identical at least 85% identical, at least 90% identical, at least 95% identical, at least 96% identical, at least 98% identical, at least 99% identical, or at least 99.5% identical to any one of SEQ ID NOs: 115-120.

[00338] *Erwinia tasmaniensis* SSB (themostable single-stranded DNA binding protein)

MASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRDKQTGETKEKTEWHRVVLFGKLAEVA
GEYLRKGSQVYIEGALQTRKWTDQAGVEKYTTEVVVNVGGTMQMLGGRSQGGGASAGGQNGGSNNGWG
QPQQPQGGNQFSGGAQQQARPQQQPQQNNAPANNEPPIDFDDDIP (SEQ ID NO: 118)

[00339] UdgX (binds to Uracil in DNA but does not excise)

MAGAQDFVPHTADLAELAAAAGECRGCGLYRDATQAVFGAGGRSARIMMIGEQPGDKEDLAGLPFVGP
AGRLLDRALEAADIDRDALYVTNAVKHFKFTRAAGGKRRIHKTPSRTEVVACRPWLIAEMTSVEPDVV
VLLGATAAKALLGNDFRVTQHRGEVLHVDDVPGDPALVATVHPSSLLRGPKEERESAFAGLVDDLRVA
ADVRP (SEQ ID NO: 119)

[00340] UDG (catalytically inactive human UDG, binds to Uracil in DNA but does not excise)

MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAAIPAKKAPAGQEEPGTPPSSPLSAEQL
DRIQRNKAAALLRLAARNVPVGFGESWKKHLSGEFGKPYFIKLMGFVAEERKHYTVYPPPHQVFTWTQMCD
IKDVKVVILGQEPYHGPNQAHGLCFSVQRPVPPPPSLENIYKELSTDIEDFVHPGHGDLSGWAKQGVLLLN
AVLTVRAHQANSHKERGWEQFTDAVVSWLNQNSNGLVFLLWGSYAQKKGSAIDRKRHHVLQTAHPSPLSVY
RGFFGCRHFSKTNELLQKSGKKPIDWKEL (SEQ ID NO: 120)

[00341] In various embodiments, the fusion protein is:

[00342]xCas9(3.7)–BE3 (***APOBEC–linker(16aa)–*****xCas9(3.7)n*****–*****linker(4aa)–*****UGI–linker(4aa)–***

NLS):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLT FFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS FFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE KKNGLFGNLIALSLGLTPNFKSNFDLAED**T**KLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIK**L**YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG**I**IPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE**K**VVDKGASAQSFIERMTNFDKNLPN EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG**D**QKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF**I**QLIHDDSLTFKEDI QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKN SRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQ SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG**V**LQKGNELALPSKYVNFLYLASHYEKLKGSPEDN EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD***SGGS****TNLSDIIEKETGKQLV IQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML**S GGS**PKKKRKV (SEQ ID NO: 122)

[00343]xCas9(3.6)–BE3 (***APOBEC–linker(16aa)–*****xCas9(3.6)n*****–*****linker(4aa)–*****UGI–linker(4aa)–***

NLS):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLIS SGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLT FFTIALQSCHYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS

FFHRLEESFLV**G**EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF
RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL**A**KSRRLENLIAQLPGE
KKNGLFGNLIALSLGLTPNFKSNFDLAED**T**KLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI
LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGAS
QEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG**I**IPHQIHLGELHAILRRQEDFYPFLKDN
REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE**K**VVDKGASAQSFIERMTNFDKNLPN
EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG**D**QKKAIVDLLFKTNRKVTVKQLKEDYFKKI
ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA
HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF**I**QLIHDDSLTFKEDI
QKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKN
SRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQ
SFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELD
KAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT
EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEA
KGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG**V**LQKGNELALPSKYVNFLYLASHYEKLKGSPEDN
EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA
PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR**I**D**I**SQLGGD***SGGS****TNLSDIIEKETGKQLV*
*IQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML****S***
GGSPKKKRKV (SEQ ID NO: 123)

[00344]BE3 (rAPOBEC1-XTEN-Cas9n-UGI-NLS)

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKF
TTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQ
IMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSC
HYQRLPPHILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDR
HSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK
HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF
IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDL
AEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ
DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ
RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
WNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE
DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVI
EMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI
NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG
YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQL
FVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIG
NKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRKV

(SEQ ID NO: 124)

[00345]In some embodiments, any of the fusion proteins provided herein comprise a second UGI domain. In some embodiments, the second UGI domain comprises a wild-type UGI or a UGI as set forth in SEQ ID NO: 115-120. In some embodiments, the UGI proteins provided herein include fragments of UGI and proteins homologous to a UGI or a UGI fragment. For example, in some embodiments, the second UGI domain comprises a fragment of the amino acid sequence set forth in SEQ ID NO: 115. In some embodiments, a UGI fragment comprises an amino acid sequence that comprises at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid sequence as set forth in SEQ ID NO: 115. In some embodiments, the second UGI domain comprises an amino acid sequence homologous to the amino acid sequence set forth in SEQ ID NO: 115 or an amino acid sequence homologous to a fragment of the amino acid sequence set forth in SEQ ID NO: 115. In some embodiments, proteins comprising UGI or fragments of UGI or homologs of UGI or UGI fragments are referred to as "UGI variants." A UGI variant shares homology to UGI, or a fragment thereof. For example a UGI variant is at least 70% identical, at least 75% identical, at least 80% identical, at least 85% identical, at least 90% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, at least 99% identical, at least 99.5% identical, or at least 99.9% identical to a wild type UGI or a UGI as set forth in SEQ ID NO: 39. In some embodiments, the UGI variant comprises a fragment of UGI, such that the fragment is at least 70% identical, at least 80% identical, at least 90% identical, at least 95% identical, at least 96% identical, at least 97% identical, at least 98% identical, at least 99% identical, at least 99.5% identical, or at least 99.9% to the corresponding fragment of wild-type UGI or a UGI as set forth in SEQ ID NO: 115.

[00346] In some embodiments, the fusion protein comprises the amino acid sequence of any one of SEQ ID NOs: 122-132. In some embodiments, the fusion protein comprises the amino acid sequence of SEQ ID NO: 122. In some embodiments, the fusion protein consists of the amino acid sequence of

SEQ ID NO: 123. In some embodiments, the fusion protein comprises the amino acid sequence of SEQ ID NO: 124. In some embodiments, the fusion protein consists of the amino acid sequence of SEQ ID NO: 125. In some embodiments, the fusion protein comprises the amino acid sequence of SEQ ID NO: 126. In some embodiments, the fusion protein consists of the amino acid sequence of SEQ ID NO: 127. In some embodiments, the fusion protein comprises an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence as set forth in SEQ ID NOs: 56-61. In some embodiments, the Cas9 domain is replaced with any of the Cas9 domains comprising one or more mutations provided herein.

[00347] xCas9 3.6-BE4 (APOBEC1-linker(32aa)-xCas9(3.6)n-linker(9aa)-UGI-linker(9aa)-UGI):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLIS
SGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLT
FFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIG
TNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFFHRLEESFLV**G**EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK
ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL
AKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED**T**KLQLSKDTYDDDLDNLLAQIGD
QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG**I**IPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE**K**VVDKGASA
QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG**D**QKKAIVDLLFKTN
RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
F**I**QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV
IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI
MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG**V**LQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIRE

QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETR**I**D**I**SQLGGDSGGS
GGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAP
EYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPE
SDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRK (SEQ ID NO:
133)

[00348] xCas9 3.7-BE4 (APOBEC1-linker(32aa)-xCas9(3.7)n-linker(9aa)-UGI-linker(9aa)-UGI):

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLIS
SGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLT
FFTIALQSCHYQRLPPHILWATGLK<u>SGGSSGGSSGSETPGTSESATPESSGGSSGGS</u>DKKYSIGLAIG
TNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK
ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL
SKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED**T**KLQLSKDTYDDDLDNLLAQIGD
QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIK**L**YDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG**I**IPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE**K**VVDKGASA
QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG**D**QKKAIVDLLFKTN
RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
F**I**QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV
IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI
MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG**V**LQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIRE
QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGS
GGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAP
EYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPE
SDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRK (SEQ ID NO:
125)

[00349] BE4:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIE
KFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLIS

SGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLT
FFTIALQSCHYQRLPPHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIG
TNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK
ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL
SKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD
QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASA
QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV
IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI
MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIRE
QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGS
GGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAP
EYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPE
SDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRK (SEQ ID NO: 130)

[00350]In some embodiments, any of the fusion proteins provided herein may further comprise a Gam protein. The term "Gam protein," as used herein, refers generally to proteins capable of binding to one or more ends of a double strand break of a double stranded nucleic acid (e.g., double stranded DNA). In some embodiments, the Gam protein prevents or inhibits degradation of one or more strands of a nucleic acid at the site of the double strand break. In some embodiments, a Gam protein is a naturally-occurring Gam protein from bacteriophage Mu, or a non-naturally occurring variant thereof. Fusion proteins comprising Gam proteins are described in Komor *et al*. (2017) Improved Base Excision Repair Inhibition and Bateriophage Mu Gam Protein Yields C:G-to-T:A

base editors with higher efficiency and product purity. *Sci Adv, 3*: eaao4774; the entire contents of which is incorporated by reference herein. In some embodiments, the Gam protein comprises an amino acid sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence provided by SEQ ID NO: 121. In some embodiments, the Gam protein comprises the amino acid sequence of SEQ ID NO: 121. In some embodiments, the fusion protein (*e.g.*, BE4-Gam of SEQ ID NO: 126) comprises a Gam protein, wherein the Cas9 domain of BE4 is replaced with any of the Cas9 domains provided herein.

[00351] Gam from bacteriophage Mu:

AKPAKRIKSAAAAYVPQNRDAVITDIKRIGDLQREASRLETEMNDAIAEITEKFAARIAPIKTDIETL
SKGVQGWCEANRDELTNGGKVKTANLVTGDVSWRVRPPSVSIRGMDAVMETLERLGLQRFIRTKQEIN
KEAILLEPKAVAGVAGITVKSGIEDFSIIPFEQEAGI (SEQ ID NO: 121)

[00352] BE4-Gam:

**MAKPAKRIKSAAAAYVPQNRDAVITDIKRIGDLQREASRLETEMNDAIAEITEKFAARIAPIKTDIET
LSKGVQGWCEANRDELTNGGKVKTANLVTGDVSWRVRPPSVSIRGMDAVMETLERLGLQRFIRTKQEI
NKEAILLEPKAVAGVAGITVKSGIEDFSIIPFEQEAGI**SGSETPGTSESATPESSSETGPVAVDPTLR
RRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCS
ITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGY
CWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLP
PHILWATGLKSGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKV
PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDD
SFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIK
FRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG
EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDA
ILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA
SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD
NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLP
NEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKK
IECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY
AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKED
IQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK
NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP
QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL

DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN
NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRN
SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED
NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLG
APAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEK
ETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGE
NKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDE
NVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRK (SEQ ID NO: 126)

[00353]Some aspects of the disclosure provide fusion proteins comprising a nucleic acid Cas9 domain (e.g.,) and an adenosine deaminase. In some embodiments, any of the fusion proteins provided herein are base editors. Some aspects of the disclosure provide fusion proteins comprising a Cas9 domain and an adenosine deaminase. The Cas9 domain may be any of the Cas9 domains (*e.g.*, a Cas9 domain) provided herein. In some embodiments, any of the Cas9 domains (*e.g.*, a Cas9 domain) provided herein may be fused with any of the adenosine deaminases provided herein. In some embodiments, the fusion protein comprises the structure:

[NH_2]-[adenosine deaminase]-[Cas9]-[COOH]; or

[NH_2]-[Cas9]-[adenosine deaminase]-[COOH].

[00354] In some embodiments, the fusion proteins comprising an adenosine deaminase and a Cas9 domain do not include a linker sequence. In some embodiments, a linker is present between the adenosine deaminase domain and the Cas9 domain. In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker. In some embodiments, the adenosine deaminase and the Cas9 domain are fused via any of the linkers provided herein. For example, in some embodiments the adenosine deaminase and the Cas9 domain are fused via any of the linkers provided below. In some embodiments, the linker comprises the amino acid sequence of any one of SEQ ID NOs: 89-112. In some embodiments, the adenosine deaminase and the Cas9 domain are fused via a linker that comprises between 1 and 200 amino acids. In some embodiments,

the adenosine deaminase and the Cas9 domain are fused via a linker that comprises from 1 to 5, 1 to 10, 1 to 20, 1 to 30, 1 to 40, 1 to 50, 1 to 60, 1 to 80, 1 to 100, 1 to 150, 1 to 200, 5 to 10, 5 to 20, 5 to 30, 5 to 40, 5 to 60, 5 to 80, 5 to 100, 5 to 150, 5 to 200, 10 to 20, 10 to 30, 10 to 40, 10 to 50, 10 to 60, 10 to 80, 10 to 100, 10 to 150, 10 to 200, 20 to 30, 20 to 40, 20 to 50, 20 to 60, 20 to 80, 20 to 100, 20 to 150, 20 to 200, 30 to 40, 30 to 50, 30 to 60, 30 to 80, 30 to 100, 30 to 150, 30 to 200, 40 to 50, 40 to 60, 40 to 80, 40 to 100, 40 to 150, 40 to 200, 50 to 60 50 to 80, 50 to 100, 50 to 150, 50 to 200, 60 to 80, 60 to 100, 60 to 150, 60 to 200, 80 to 100, 80 to 150, 80 to 200, 100 to 150, 100 to 200, or 150 to 200 amino acids in length. In some embodiments, the adenosine deaminase and the Cas9 domain are fused via a linker that comprises 3, 4, 16, 24, 32, 64, 100, or 104 amino acids in length. In some embodiments, the adenosine deaminase and the Cas9 domain are fused via a linker that comprises the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 89), SGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 106), or GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS (SEQ ID NO: 110). In some embodiments, the adenosine deaminase and the Cas9 domain are fused via a linker comprising the amino acid sequence SGSETPGTSESATPES (SEQ ID NO: 89), which may also be referred to as the XTEN linker. In some embodiments, the linker is 24 amino acids in length. In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPES (SEQ ID NO: 111). In some embodiments, the linker is 32 amino acids in length. In some embodiments, the linker comprises the amino acid sequence (SGGS)2-SGSETPGTSESATPES-(SGGS)2 (SEQ ID NO: 96), which may also be referred to as (SGGS)2-XTEN-(SGGS)2. In some embodiments, the linker comprises the amino acid sequence (SGGS)n-SGSETPGTSESATPES-(SGGS)n (SEQ ID NO: 98), wherein n is 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10. In some embodiments, the linker is 40 amino acids in length. In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGS (SEQ ID NO: 106). In some

embodiments, the linker is 64 amino acids in length. In some embodiments, the linker comprises the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGS (SEQ ID NO: 107). In some embodiments, the linker is 92 amino acids in length. In some embodiments, the linker comprises the amino acid sequence PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEP SEGSAPGTSESATPESGPGSEPATS (SEQ ID NO: 108).

[00355] In some embodiments, the fusion proteins comprise one or more adenosine deaminases defined herein, or to any amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the amino acid sequences set forth herein.

[00356] In some embodiments, the fusion proteins comprising an adenosine deaminase provided herein further comprise one or more nuclear targeting sequences, for example, a nuclear localization sequence (NLS). In some embodiments, a NLS comprises an amino acid sequence that facilitates the importation of a protein, that comprises an NLS, into the cell nucleus (*e.g.,* by nuclear transport). In some embodiments, any of the fusion proteins provided herein further comprise a nuclear localization sequence (NLS). In some embodiments, the NLS is fused to the N-terminus of the fusion protein. In some embodiments, the NLS is fused to the C-terminus of the fusion protein. In some embodiments, the NLS is fused to the N-terminus of the IBR (*e.g.*, dISN). In some embodiments, the NLS is fused to the C-terminus of the IBR (*e.g.,* dISN). In some embodiments, the NLS is fused to the N-terminus of the Cas9 domain. In some embodiments, the NLS is fused to the C-terminus of the Cas9 domain. In some embodiments, the NLS is fused to the N-terminus of the adenosine deaminase. In some embodiments, the NLS is fused to the C-terminus of the adenosine deaminase. In some embodiments, the NLS is fused to the fusion protein via one or more linkers. In

some embodiments, the NLS is fused to the fusion protein without a linker. In some embodiments, the NLS comprises an amino acid sequence of any one of the NLS sequences provided or referenced herein. In some embodiments, the NLS comprises an amino acid sequence as set forth in SEQ ID NO: 37 or SEQ ID NO: 38. Additional nuclear localization sequences are known in the art and would be apparent to the skilled artisan. For example, NLS sequences are described in Plank *et al.*, PCT/EP2000/011690, the contents of which are incorporated herein by reference for their disclosure of exemplary nuclear localization sequences. In some embodiments, a NLS comprises the amino acid sequence PKKKRKV (SEQ ID NO: 113). In some embodiments, a NLS comprises the amino acid sequence MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 114).

[00357] In some embodiments, the general architecture of exemplary fusion proteins with an adenosine deaminase and a Cas9 domain comprises any one of the following structures, where NLS is a nuclear localization sequence (*e.g.,* any NLS provided herein), NH_2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein. Fusion proteins comprising an adenosine deaminase, a napDNAbp, and a NLS:

NH_2-[NLS]-[adenosine deaminase]-[Cas9]-COOH;

NH_2-[adenosine deaminase]-[NLS]-[Cas9]-COOH;

NH_2-[adenosine deaminase]-[Cas9]-[NLS]-COOH;

NH_2-[NLS]-[Cas9]-[adenosine deaminase]-COOH;

NH_2-[Cas9]-[NLS]-[adenosine deaminase]-COOH; and

NH_2-[Cas9]-[adenosine deaminase]-[NLS]-COOH.

[00358] In some embodiments, the fusion proteins comprising an adenosine deaminase domain provided herein do not comprise a linker. In some embodiments, a linker is present between one or more of the domains or proteins (*e.g.,* adenosine deaminase, Cas9 domain, and/or NLS). In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker.

[00359] Some aspects of the disclosure provide fusion proteins that comprise a Cas9 domain (e.g. a Cas9 domain) and at least two adenosine deaminase domains. Without wishing to be bound by any particular theory, dimerization of adenosine deaminases (*e.g.,* in cis or in trans) may improve the ability (*e.g.,* efficiency) of the fusion protein to modify a nucleic acid base, for example to deaminate adenine. In some embodiments, any of the fusion proteins may comprise 2, 3, 4 or 5 adenosine deaminase domains. In some embodiments, any of the fusion proteins provided herein comprise two adenosine deaminases. In some embodiments, any of the fusion proteins provided herein contain only two adenosine deaminases. In some embodiments, the adenosine deaminases are the same. In some embodiments, the adenosine deaminases are any of the adenosine deaminases provided herein. In some embodiments, the adenosine deaminases are different. In some embodiments, the first adenosine deaminase is any of the adenosine deaminases provided herein, and the second adenosine is any of the adenosine deaminases provided herein, but is not identical to the first adenosine deaminase. Additional fusion protein constructs comprising two adenosine deaminase domains suitable for use herein are illustrated in Gaudelli *et al.* (2017) Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage, *Nature, 551*(23); 464-471; the entire contents of which is incorporated herein by reference.

[00360] In some embodiments, the first adenosine deaminase and the second deaminase are fused directly or via a linker. In some embodiments, the linker is any of the linkers provided herein. In some embodiments, the linker comprises the amino acid sequence of any one of the linker sequences disclosed herein (e.g., linkers of SEQ ID NOs: 21-36, 64, 65, 66, or 67). In some embodiments, the first adenosine deaminase is the same as the second adenosine deaminase. In some embodiments, the first adenosine deaminase and the second adenosine deaminase are any of the adenosine deaminases described herein. In some embodiments, the first adenosine deaminase and the second adenosine deaminase are different. In some embodiments, the first adenosine deaminase is any of the adenosine deaminases provided herein. In some embodiments, the second adenosine deaminase is any of the

adenosine deaminases provided herein but is not identical to the first adenosine deaminase. In some embodiments, the first adenosine deaminase is an ecTadA adenosine deaminase. In some embodiments, the first adenosine deaminase comprises an amino acid sequence that is at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to any one of the amino acid sequences set forth herein.

[00361] In some embodiments, the general architecture of exemplary fusion proteins with a first adenosine deaminase, a second adenosine deaminase, and a Cas9 domain (e.g.) comprises any one of the following structures, where NLS is a nuclear localization sequence (*e.g.,* any NLS provided herein), NH_2 is the N-terminus of the fusion protein, and COOH is the C-terminus of the fusion protein:

NH_2-[first adenosine deaminase]-[second adenosine deaminase]-[Cas9]-COOH;

NH_2-[first adenosine deaminase]-[Cas9]-[second adenosine deaminase]-COOH;

NH_2-[Cas9]-[first adenosine deaminase]-[second adenosine deaminase]-COOH;

NH_2-[second adenosine deaminase]-[first adenosine deaminase]-[Cas9]-COOH;

NH_2-[second adenosine deaminase]-[Cas9]-[first adenosine deaminase]-COOH;

NH_2-[Cas9]-[second adenosine deaminase]-[first adenosine deaminase]-COOH;

[00362] In some embodiments, the fusion proteins provided herein do not comprise a linker. In some embodiments, a linker is present between one or more of the domains or proteins (*e.g.,* first adenosine deaminase, second adenosine deaminase, and/or napDNAbp). In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker.

[00363] In some embodiments, a fusion protein comprising a first adenosine deaminase, a second adenosine deaminase, and a Cas9 domain further comprise a NLS. Exemplary fusion proteins comprising a first adenosine deaminase, a second adenosine deaminase, a napDNAbp, and an NLS are shown as follows:

NH_2-[NLS]-[first adenosine deaminase]-[second adenosine deaminase]-[Cas9]-COOH;

NH_2-[first adenosine deaminase]-[NLS]-[second adenosine deaminase]-[Cas9]-COOH;

NH_2-[first adenosine deaminase]-[second adenosine deaminase]-[NLS]-[Cas9]-COOH;

NH_2-[first adenosine deaminase]-[second adenosine deaminase]-[Cas9]-[NLS]-COOH;

NH_2-[NLS]-[first adenosine deaminase]-[Cas9]-[second adenosine deaminase]-COOH;

NH_2-[first adenosine deaminase]-[NLS]-[Cas9]-[second adenosine deaminase]-COOH;

NH_2-[first adenosine deaminase]-[Cas9]-[NLS]-[second adenosine deaminase]-COOH;

NH_2-[first adenosine deaminase]-[Cas9]-[second adenosine deaminase]-[NLS]-COOH;

NH_2-[NLS]-[Cas9]-[first adenosine deaminase]-[second adenosine deaminase]-COOH;

NH_2-[Cas9]-[NLS]-[first adenosine deaminase]-[second adenosine deaminase]-COOH;

NH_2-[Cas9]-[first adenosine deaminase]-[NLS]-[second adenosine deaminase]-COOH;

NH_2-[Cas9]-[first adenosine deaminase]-[second adenosine deaminase]-[NLS]-COOH;

NH_2-[NLS]-[second adenosine deaminase]-[first adenosine deaminase]-[Cas9]-COOH;

NH_2-[second adenosine deaminase]-[NLS]-[first adenosine deaminase]-[Cas9]-COOH;

NH_2-[second adenosine deaminase]-[first adenosine deaminase]-[NLS]-[Cas9]-COOH;

NH_2-[second adenosine deaminase]-[first adenosine deaminase]-[Cas9]-[NLS]-COOH;

NH_2-[NLS]-[second adenosine deaminase]-[Cas9]-[first adenosine deaminase]-COOH;

NH_2-[second adenosine deaminase]-[NLS]-[Cas9]-[first adenosine deaminase]-COOH;

NH_2-[second adenosine deaminase]-[Cas9]-[NLS]-[first adenosine deaminase]-COOH;

NH_2-[second adenosine deaminase]-[Cas9]-[first adenosine deaminase]-[NLS]-COOH;

NH_2-[NLS]-[Cas9]-[second adenosine deaminase]-[first adenosine deaminase]-COOH;

NH_2-[Cas9]-[NLS]-[second adenosine deaminase]-[first adenosine deaminase]-COOH;

NH_2-[Cas9]-[second adenosine deaminase]-[NLS]-[first adenosine deaminase]-COOH;

NH_2-[Cas9]-[second adenosine deaminase]-[first adenosine deaminase]-[NLS]-COOH;

[00364] In some embodiments, the fusion proteins provided herein do not comprise a

linker. In some embodiments, a linker is present between one or more of the domains or proteins (*e.g.,* first adenosine deaminase, second adenosine deaminase, Cas9 domain, and/or NLS). In some embodiments, the "-" used in the general architecture above indicates the presence of an optional linker.

[00365] In some embodiments, the fusion protein comprises a Cas9 domain fused to one or more adenosine deaminase domains (*e.g.*, a first adenosine deaminase and a second adenosine deaminase), wherein the fusion protein comprises or consists of the amino acid sequence of SEQ ID NO: 127. In some embodiments, the fusion protein comprises the amino acid sequence of SEQ ID NO: 128. In some embodiments, the fusion protein is the amino acid sequence of SEQ ID NO: 129. In some embodiments, the Cas9 domain of SEQ ID NOs: 127-129 is replaced with any of the Cas9 domains provided herein.

[00366]xCas9(3.7)–ABE: (***ecTadA(wt)*–*linker(32 aa)*–*ecTadA*(7.10)*–*linker(32 aa)*–nxCas9(3.7)**–NLS):

***MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGI LADECAALLSDFFRMRRQEIKAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGS*SEVEF SHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYR LIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECA ALLCYFFRMPRQVFNAQKKAQSSTD*SGGSSGGSSGSETPGTSESATPESSGGSSGGSDKKYSIGLAIG TNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL SKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFF DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEKVVDKGASA QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTN RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK

FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI
MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG**V**LQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIRE
QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD***EGAD***
KRTADGSEFESPKKKRKV (SEQ ID NO: 127)

[00367]xCas9(3.6)–ABE: (***ecTadA(wt)–linker(32 aa)–****ecTadA*(7.10)–****linker(32 aa)–*****nxCas9(3.6)–***

NLS):

MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGL
VMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMNHRVEITEGI
LADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGS*SEVEF*
SHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYR
LIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECA
*ALLCYFFRMPRQVFNAQKKAQSSTD****SGGSSGGSSGSETPGTSESATPESSGGSSGGS***DKKYSIGLAIG
TNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFFHRLEESFLV**G**EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK
ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL
AKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED**T**KLQLSKDTYDDDLDNLLAQIGD
QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG**I**IPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE**K**VVDKGASA
QSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSG**D**QKKAIVDLLFKTN
RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLT
LFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN
F**I**QLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV
IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKA
TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV
QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI
MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAG**V**LQKGNELALPSKYVNFL
YLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIRE
QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITG**I**YETRIDLSQLGGD***EGAD***
KRTADGSEFESPKKKRKV (SEQ ID NO: 128)

[00368] ABE7.10: **ecTadA**$_{(wild\text{-}type)}$-(SGGS)$_2$-*XTEN*-(SGGS)$_2$-**ecTadA**$_{(W23R_H36L_P48A_R51L_L84F_A106V_D108N_H123Y_S146C_D147Y_R152P_E155V_I156F_K157N)}$-(SGGS)$_2$-*XTEN*-(SGGS)$_2$_nCas9_SGGS_*NLS*

[00369] **MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEI MALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHHPGMN HRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD**SGGSSGGS*SGSETPGTSESATPES*SGG SSGGS**SEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMAL RQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRV EITEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD**SGGSSGGS*SGSETPGTSESATPES*SGGSSG GSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYP TIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTL LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRK QRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRK SEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLING IRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDS RMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKL ESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGDSGGS*PKKKRKV* (SEQ ID NO: 129)

[00370] In some embodiments, the fusion proteins provided herein comprising one or more adenosine deaminase domains and a Cas9 domain exhibit an increased activity on a target sequence that does not comprise the canonical PAM (5′-NGG-3′) at its 3′ end as compared to a fusion protein

comprising *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2. In some embodiments, the fusion protein exhibits an activity on a target sequence having a 3′ end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 2-fold, at least 3-fold, at least 4-fold, at least 5-fold, at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold increased as compared to the activity of a fusion protein comprising *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2 on the same target sequence. In some embodiments, the 3′ end of the target sequence is directly adjacent to a sequence selected from the group consisting of NGT, NGA, NGC, and NNG, wherein N is an A, G, T, or C. In some embodiments, the 3′ end of the target sequence is directly adjacent to a sequence selected from the group consisting of CGG, AGT, TGG, AGT, CGT, GGG, CGT, TGT, GGT, AGC, CGC, TGC, GGC, AGA, CGA, TGA, GGA, GAA, GAT, and CAA. In some embodiments, the fusion protein activity is measured by a nuclease assay, a deamination assay, a transcriptional activation assay, or high-throughput sequencing. In some embodiments, the transcriptional activation assay is a GFP activation assay. In some embodiments, high-throughput sequencing is used to measure indel formation.

[00371] It should be appreciated that the fusion proteins of the present disclosure may comprise one or more additional features. For example, in some embodiments, the fusion protein may comprise cytoplasmic localization sequences, export sequences, such as nuclear export sequences, or other localization sequences, as well as sequence tags that are useful for solubilization, purification, or detection of the fusion proteins. Suitable protein tags provided herein include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin (HA)-tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags (*e.g.*, Softag 1, Softag 3), strep-tags, biotin ligase tags, FlAsH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of

ordinary skill in the art. In some embodiments, the fusion protein comprises one or more His tags.

[00372] Additional suitable strategies for generating fusion proteins comprising a napDNAbp (*e.g.*, a Cas9 domain) and a nucleic acid editing domain (*e.g.*, a deaminase domain) will be apparent to those of ordinary skill in the art based on this disclosure in combination with the general knowledge in the art. Suitable strategies for generating fusion proteins according to aspects of this disclosure using linkers or without the use of linkers will also be apparent to those of ordinary skill in the art in view of the instant disclosure and the knowledge in the art. For example, Gilbert *et al.*, CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.* **2013**; 154(2):442-51, showed that C-terminal fusions of Cas9 with VP64 using 2 NLS's as a linker, can be employed for transcriptional activation. Mali *et al.* (CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol.* 2013; 31(9):833-8), reported that C-terminal fusions with VP64 without linker can be employed for transcriptional activation. And Maeder *et al.* (CRISPR RNA-guided activation of endogenous human genes. *Nat Methods.* **2013**; 10: 977-979), reported that C-terminal fusions with VP64 using a GGGGS (SEQ ID NO: 94) linker can be used as transcriptional activators. Recently, dCas9- FokI nuclease fusions have successfully been generated and exhibit improved enzymatic specificity as compared to the parental Cas9 enzyme (In Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 2014; 32(6): 577-82; and in Tsai SQ, *et al.* Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol.* 2014; 32(6):569-76. PMID: 24770325 a SGSETPGTSESATPES (SEQ ID NO: 89) or a GGGGS (SEQ ID NO: 94) linker was used in FokI-dCas9 fusion proteins, respectively).

[00373]In some embodiments, the Cas9 fusion protein comprises: (i) Cas9 domain; and (ii) a transcriptional activator domain. In some embodiments, the transcriptional activator domain

comprises a VPR. VPR is a VP64-SV40-P65-RTA tripartite activator. In some embodiments, VPR comprises a VP64 amino acid sequence encoded by the nucleic acid sequence of SEQ ID NO: 85:

GAGGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGACGC
CCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCG
GCAGTGACGCCCTTGATGATTTCGACCTGGACATGCTGATTAACTCTAGATAG (SEQ ID NO: 73).

[00374]In some embodiments, VPR comprises a VP64 amino acid sequence as set forth in SEQ ID NO: 86:

EASGSGRADALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLINSR (SEQ ID NO: 86).

[00375]In some embodiments, VPR compises a VP64-SV40-P65-RTA amino acid sequence encoded by the nucleic acid sequence of SEQ ID NO: 87:

TCGCCAGGGATCCGTCGACTTGACGCGTTGATATCAACAAGTTTGTACAAAAAAGCAGGCTACAAAGA
GGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGACGCCC
TCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGC
AGTGACGCCCTTGATGATTTCGACCTGGACATGCTGATTAACTCTAGAAGTTCCGGATCTCCGAAAAA
GAAACGCAAAGTTGGTAGCCAGTACCTGCCCGACACCGACGACCGGCACCGGATCGAGGAAAAGCGGA
AGCGGACCTACGAGACATTCAAGAGCATCATGAAGAAGTCCCCCTTCAGCGGCCCCACCGACCCTAGA
CCTCCACCTAGAAGAATCGCCGTGCCCAGCAGATCCAGCGCCAGCGTGCCAAAACCTGCCCCCCAGCC
TTACCCCTTCACCAGCAGCCTGAGCACCATCAACTACGACGAGTTCCCTACCATGGTGTTCCCCAGCG
GCCAGATCTCTCAGGCCTCTGCTCTGGCTCCAGCCCCTCCTCAGGTGCTGCCTCAGGCTCCTGCTCCT
GCACCAGCTCCAGCCATGGTGTCTGCACTGGCTCAGGCACCAGCACCCGTGCCTGTGCTGGCTCCTGG
ACCTCCACAGGCTGTGGCTCCACCAGCCCCTAAACCTACACAGGCCGGCGAGGGCACACTGTCTGAAG
CTCTGCTGCAGCTGCAGTTCGACGACGAGGATCTGGGAGCCCTGCTGGGAAACAGCACCGATCCTGCC
GTGTTCACCGACCTGGCCAGCGTGGACAACAGCGAGTTCCAGCAGCTGCTGAACCAGGGCATCCCTGT
GGCCCCTCACACCACCGAGCCCATGCTGATGGAATACCCCGAGGCCATCACCCGGCTCGTGACAGGCG
CTCAGAGGCCTCCTGATCCAGCTCCTGCCCCTCTGGGAGCACCAGGCCTGCCTAATGGACTGCTGTCT
GGCGACGAGGACTTCAGCTCTATCGCCGATATGGATTTCTCAGCCTTGCTGGGCTCTGGCAGCGGCAG
CCGGGATTCCAGGGAAGGGATGTTTTTGCCGAAGCCTGAGGCCGGCTCCGCTATTAGTGACGTGTTTG
AGGGCCGCGAGGTGTGCCAGCCAAAACGAATCCGGCCATTTCATCCTCCAGGAAGTCCATGGGCCAAC
CGCCCACTCCCCGCCAGCCTCGCACCAACACCAACCGGTCCAGTACATGAGCCAGTCGGGTCACTGAC
CCCGGCACCAGTCCCTCAGCCACTGGATCCAGCGCCCGCAGTGACTCCCGAGGCCAGTCACCTGTTGG
AGGATCCCGATGAAGAGACGAGCCAGGCTGTCAAAGCCCTTCGGGAGATGGCCGATACTGTGATTCCC
CAGAAGGAAGAGGCTGCAATCTGTGGCCAAATGGACCTTTCCCATCCGCCCCCAAGGGGCCATCTGGA
TGAGCTGACAACCACACTTGAGTCCATGACCGAGGATCTGAACCTGGACTCACCCCTGACCCCGGAAT

TGAACGAGATTCTGGATACCTTCCTGAACGACGAGTGCCTCTTGCATGCCATGCATATCAGCACAGGA
CTGTCCATCTTCGACACATCTCTGTTTTGA

(SEQ ID NO: 87).

[00376]In some embodiments, VPR comprises a VP64-SV40-P65-RTA amino acid sequence as set forth in SEQ ID NO: 88:

SPGIRRLDALISTSLYKKAGYKEASGSGRADALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLG
SDALDDFDLDMLINSRSSGSPKKKRKVGSQYLPDTDDRHRIEEKRKRTYETFKSIMKKSPFSGPTDPR
PPPRRIAVPSRSSASVPKPAPQPYPFTSSLSTINYDEFPTMVFPSGQISQASALAPAPPQVLPQAPAP
APAPAMVSALAQAPAPVPVLAPGPPQAVAPPAPKPTQAGEGTLSEALLQLQFDDEDLGALLGNSTDPA
VFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGAPGLPNGLLS
GDEDFSSIADMDFSALLGSGSGSRDSREGMFLPKPEAGSAISDVFEGREVCQPKRIRPFHPPGSPWAN
RPLPASLAPTPTGPVHEPVGSLTPAPVPQPLDPAPAVTPEASHLLEDPDEETSQAVKALREMADTVIP
QKEEAAICGQMDLSHPPPRGHLDELTTTLESMTEDLNLDSPLTPELNEILDTFLNDECLLHAMHISTG
LSIFDTSLF

(SEQ ID NO: 88).

[00377]Some aspects of this disclosure provide fusion proteins comprising a transcription activator. In some embodiments, the transcriptional activator is VPR. In some embodiments, the VPR comprises a wild type VPR or a VPR as set forth in SEQ ID NO: 88. In some embodiments, the VPR proteins provided herein include fragments of VPR and proteins homologous to a VPR or a VPR fragment. For example, in some embodiments, a VPR comprises a fragment of the amino acid sequence set forth in SEQ ID NO: 88. In some embodiments, a VPR comprises an amino acid sequence homologous to the amino acid sequence set forth in SEQ ID NO: 88 or an amino acid sequence homologous to a fragment of the amino acid sequence set forth in SEQ ID NO: 8. In some embodiments, proteins comprising VPR or fragments of VPR or homologs of VPR or VPR fragments are referred to as "VPR variants." A VPR variant shares homology to VPR, or a fragment thereof. For example a VPR variant is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to a wild type VPR or a VPR as set forth in SEQ ID NO: 88. In some embodiments, the VPR variant comprises a

fragment of VPR, such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to the corresponding fragment of wild type VPR or a VPR as set forth in SEQ ID NO: 88. In some embodiments, the VPR comprises the amino acid sequence set forth in SEQ ID NO: 88. In some embodiments, the VPR comprises an amino acid sequence encoded by the nucleotide sequence set forth in SEQ ID NO: 88.

[00378]In some embodiments, a VPR is a VP64-SV40-P65-RTA triple activator. In some embodiments, the VP64-SV40-P65-RTA comprises a VP64-SV40-P65-RTA as set forth in SEQ ID NO: 88. In some embodiments, the VP64-SV40-P65-RTA proteins provided herein include fragments of VP64-SV40-P65-RTA and proteins homologous to a VP64-SV40-P65-RTA or a VP64-SV40-P65-RTA fragment. For example, in some embodiments, a VP64-SV40-P65-RTA comprises a fragment of the amino acid sequence set forth in SEQ ID NO: 88. In some embodiments, a VP64-SV40-P65-RTA comprises an amino acid sequence homologous to the amino acid sequence set forth in SEQ ID NO: 88 or an amino acid sequence homologous to a fragment of the amino acid sequence set forth in SEQ ID NO: 88. In some embodiments, proteins comprising VP64-SV40-P65-RTA or fragments of VP64-SV40-P65-RTA or homologs of VP64-SV40-P65-RTA or VP64-SV40-P65-RTA fragments are referred to as "VP64-SV40-P65-RTA variants." A VP64-SV40-P65-RTA variant shares homology to VP64-SV40-P65-RTA, or a fragment thereof. For example a VP64-SV40-P65-RTA variant is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to a VP64-SV40-P65-RTA as set forth in SEQ ID NO: 88. In some embodiments, the VP64-SV40-P65-RTA variant comprises a fragment of VP64-SV40-P65-RTA, such that the fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about

96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% to a fragment of a VP64-SV40-P65-RTA as set forth in SEQ ID NO: 88. In some embodiments, the VP64-SV40-P65-RTA comprises the amino acid sequence set forth in SEQ ID NO: 88. In some embodiments, the VP64-SV40-P65-RTA comprises an amino acid sequence encoded by the nucleotide sequence set forth in SEQ ID NO: 87.

[00379]In some embodiments, the fusion protein comprises the nucleic acid sequence of SEQ ID NO: 87.

[00380] dCas9–VPR (dCas9(3.7)–NLS–linker(22aa)–VP64–linker(4aa)–NLS–p65AD–linker(6aa)-RtaAD)

ATGGACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCGGCTGGGCCGTCATTACGGA
CGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCGCCACAGCATAAAGAAGA
ACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCACGCGGCTCAAAAGAACAGCA
CGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAGGAGATCTTTAGTAATGAGATGGC
TAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGGTGGAGGAGGATAAAAAGCACG
AGCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGTACCCAACCATATAT
CATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTGCGGTTGATCTATCTCGCGCTGGC
GCATATGATCAAATTTCGGGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCG
ACAAACTCTTTATCCAACTGGTTCAGACTTACAATCAGCTTTTCGAAGAGAACCCGATCAACGCATCC
GGAGTTGACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGC
ACAGCTCCCTGGGGAGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCCCTCGGGCTGACCC
CCAACTTTAAATCTAACTTCGACCTGGCCGAAGATACCAAGCTTCAACTGAGCAAAGACACCTACGAT
GATGATCTCGACAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGACCTTTTTTTGGCGGCAAAGAA
CCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAAGCTCCGCTGA
GCGCTAGTATGATCAAGCTCTATGATGAGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGA
CAGCAACTGCCTGAGAAGTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACAT
TGACGGCGGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGCA
CCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACAATGGA
ATCATCCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGATTTCTACCC
CTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCCTCACATTTCGGATACCCTACTATGTAGGCC
CCCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAAATCAGAAGAGACCATCACTCCCTGG
AACTTCGAGAAAGTCGTGGATAAGGGGGCCTCTGCCCAGTCCTTCATCGAAAGGATGACTAACTTTGA
TAAAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTTATA
ACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAGCCAGCATTCCTGTCTGGAGATCAG
AAGAAAGCTATTGTGGACCTCCTCTTCAAGACGAACCGGAAAGTTACCGTGAAACAGCTCAAAGAAGA

CTATTTCAAAAAGATTGAATGTTTCGACTCTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCAT
CCCTGGGAACGTATCACGATCTCCTGAAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAAC
GAGGACATTCTTGAGGACATTGTCCTCACCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACG
CTTGAAAACTTACGCTCATCTCTTCGACGACAAAGTCATGAAGCAGCTCAAGAGGCGCCGATATACAG
GATGGGGGCGGCTGTCAAGAAAACTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCCTG
GATTTTCTTAAGTCCGATGGATTTGCCAACCGGAACTTCATTCAGTTGATCCATGATGACTCTCTCAC
CTTTAAGGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTA
ATCTTGCAGGTAGCCCAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGTC
AAAGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACCACCCA
GAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAGAACTGGGGTCCC
AAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTACTACCTG
CAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGA
TCATATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTTGACAAGATCCGATA
AAAATAGAGGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATTATTGGCGG
CAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGG
CCTGTCTGAGTTGGATAAAGCCGGTTTCATCAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGC
ACGTGGCCCAAATTCTCGATTCACGCATGAACACCAAGTACGATGAAAATGACAAACTGATTCGAGAG
GTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTCAGAAAGGACTTTCAGTTTTATAAGGT
GAGAGAGATCAACAATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTA
TCAAAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAA
ATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTAT
GAATTTTTTCAAGACCGAGATTACACTGGCCAATGGAGAGATTCGGAAGCGACCACTTATCGAAACAA
ACGGAGAAACAGGAGAAATCGTGTGGGACAAGGGTAGGGATTTCGCGACAGTCCGGAAGGTCCTGTCC
ATGCCGCAGGTGAACATCGTTAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCT
CCCGAAAAGGAACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGGAT
TCGATTCTCCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAAA
CTCAAAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCAT
CGACTTTCTCGAGGCGAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAGCTTCCCAAGTACT
CTCTCTTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGCGGGCGTGCTGCAGAAAGGTAAC
GAGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGG
GTCTCCCGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACACAAACACTACCTTGATGAGATCA
TCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGATAAGGTGCTTTCT
GCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTTACTCT
GACCAACTTGGGCGCGCCTGCAGCCTTCAAGTACTTCGACACCACCATAGACAGAAAGCGGTACACCT
CTACAAAGGAGGTCCTGGACGCCACACTGATTCATCAGTCAATTACGGGGCTCTATGAAACAAGAATC
GACCTCTCTCAGCTCGGTGGAGACCCCCAAGAAGAAGAGGAAGGTGTCGCCAGGGATCCGTCGACTTG
ACGCGTTGATATCAACAAGTTTGTACAAAAAAGCAGGCTACAAAGAGGCCAGCGGTTCCGGACGGGCT
GACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACAT

GCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATTTCG
ACCTGGACATGCTGATTAACTCTAGAAGTTCCGGATCTCCGAAAAAGAAACGCAAAGTTGGTAGCCAG
TACCTGCCCGACACCGACGACCGGCACCGGATCGAGGAAAAGCGGAAGCGGACCTACGAGACATTCAA
GAGCATCATGAAGAAGTCCCCCTTCAGCGGCCCCACCGACCCTAGACCTCCACCTAGAAGAATCGCCG
TGCCCAGCAGATCCAGCGCCAGCGTGCCAAAACCTGCCCCCCAGCCTTACCCCTTCACCAGCAGCCTG
AGCACCATCAACTACGACGAGTTCCCTACCATGGTGTTCCCCAGCGGCCAGATCTCTCAGGCCTCTGC
TCTGGCTCCAGCCCCTCCTCAGGTGCTGCCTCAGGCTCCTGCTCCTGCACCAGCTCCAGCCATGGTGT
CTGCACTGGCTCAGGCACCAGCACCCGTGCCTGTGCTGGCTCCTGGACCTCCACAGGCTGTGGCTCCA
CCAGCCCCTAAACCTACACAGGCCGGCGAGGGCACACTGTCTGAAGCTCTGCTGCAGCTGCAGTTCGA
CGACGAGGATCTGGGAGCCCTGCTGGGAAACAGCACCGATCCTGCCGTGTTCACCGACCTGGCCAGCG
TGGACAACAGCGAGTTCCAGCAGCTGCTGAACCAGGGCATCCCTGTGGCCCCTCACACCACCGAGCCC
ATGCTGATGGAATACCCCGAGGCCATCACCCGGCTCGTGACAGGCGCTCAGAGGCCTCCTGATCCAGC
TCCTGCCCCTCTGGGAGCACCAGGCCTGCCTAATGGACTGCTGTCTGGCGACGAGGACTTCAGCTCTA
TCGCCGATATGGATTTCTCAGCCTTGCTGGGCTCTGGCAGCGGCAGCCGGGATTCCAGGGAAGGGATG
TTTTTGCCGAAGCCTGAGGCCGGCTCCGCTATTAGTGACGTGTTTGAGGGCCGCGAGGTGTGCCAGCC
AAAACGAATCCGGCCATTTCATCCTCCAGGAAGTCCATGGGCCAACCGCCCACTCCCCGCCAGCCTCG
CACCAACACCAACCGGTCCAGTACATGAGCCAGTCGGGTCACTGACCCCGGCACCAGTCCCTCAGCCA
CTGGATCCAGCGCCCGCAGTGACTCCCGAGGCCAGTCACCTGTTGGAGGATCCCGATGAAGAGACGAG
CCAGGCTGTCAAAGCCCTTCGGGAGATGGCCGATACTGTGATTCCCCAGAAGGAAGAGGCTGCAATCT
GTGGCCAAATGGACCTTTCCCATCCGCCCCCAAGGGGCCATCTGGATGAGCTGACAACCACACTTGAG
TCCATGACCGAGGATCTGAACCTGGACTCACCCCTGACCCCGGAATTGAACGAGATTCTGGATACCTT
CCTGAACGACGAGTGCCTCTTGCATGCCATGCATATCAGCACAGGACTGTCCATCTTCGACACATCTC
TGTTT

(SEQ ID NO: 131).

[00381]Some aspects of this disclosure provide fusion proteins comprising a Cas9 domain as provided herein that is fused to a second protein, or a "fusion partner", such as a nucleic acid editing domain, thus forming a fusion protein. In some embodiments, the nucleic acid editing domain is fused to the N-terminus of the Cas9 domain. In some embodiments, the nucleic acid editing domain is fused to the C-terminus of the Cas9 domain. In some embodiments, the Cas9 domain and the nucleic acid editing domain are fused to each other via a linker. Suitable strategies for generating fusion proteins according to aspects of this disclosure using linkers or without the use of linkers will also be apparent to those of skill in the art in view of the instant disclosure and the knowledge in the art. For example, Gilbert *et al.*, CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell.*

2013; 154(2):442-51, showed that C-terminal fusions of Cas9 with VP64 using 2 NLS's as a linker (SPKKKRKVEAS), can be employed for transcriptional activation. Mali *et al.*, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol.* **2013**; 31(9):833-8, reported that C-terminal fusions with VP64 without linker can be employed for transcriptional activation. Maeder *et al.*, CRISPR RNA-guided activation of endogenous human genes. *Nat Methods.* 2013; 10: 977-979, reported that C-terminal fusions with VP64 using a GGGGS (SEQ ID NO: 94) linker can be used as transcriptional activators. Recently, dCas9- FokI nuclease fusions have successfully been generated and exhibit improved enzymatic specificity as compared to the parental Cas9 enzyme (In Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* 2014**;** 32(6): 577-82, and in Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ, Aryee MJ, Joung JK. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol.* 2014**;** 32(6):569-76. PMID: 24770325 a SGSETPGTSESATPES (SEQ ID NO: 89) or a $GGGGS_n$ (SEQ ID NO: 95) linker was used in FokI-dCas9 fusion proteins, respectively).

[00382]In some embodiments, the second protein in the fusion protein (*i.e.*, the fusion partner) comprises a nucleic acid editing domain. Such a nucleic acid editing domain may be, without limitation, a nuclease, a nickase, a recombinase, a deaminase, a methyltransferase, a methylase, an acetylase, or an acetyltransferase. Non-limiting exemplary nucleic acid editing domains that may be used in accordance with this disclosure include cytidine deaminases and adenosine deaminases. In some embodiments, the nucleic acid editing domain is a deaminase domain. In some embodiments, the nucleic acid editing domain is a nuclease domain. In some embodiments, the nuclease domain is a FokI DNA cleavage domain. In some embodiments, this disclosure provides dimers of the fusion proteins provided herein, *e.g.*, dimers of fusion proteins may include a dimerizing nuclease domain. In some embodiments, the nucleic acid editing domain is a nickase domain. In some embodiments, the

nucleic acid editing domain is a recombinase domain. In some embodiments, the nucleic acid editing domain is a methyltransferase domain. In some embodiments, the nucleic acid editing domain is a methylase domain. In some embodiments, the nucleic acid editing domain is an acetylase domain. In some embodiments, the nucleic acid editing domain is an acetyltransferase domain. Additional nucleic acid editing domains would be apparent to a person of ordinary skill in the art based on this disclosure and knowledge in the field and are within the scope of this disclosure. In other embodiments, the second protein comprises a domain that modulates transcriptional activity. Such transcriptional modulating domains may be, without limitation, a transcriptional activator or transcriptional repressor domain.

Guide RNA

[00383]In various embodiments, the base editors described herein may be complexed, bound, or otherwise associated with (e.g., via any type of covalent or non-covalent bond) one or more guide sequences, i.e., the sequence which becomes associated or bound to the base editor and directs its localization to a specific target sequence having complementarity to the guide sequence or a portion thereof. The particular design embodiments of a guide sequence will depend upon the nucleotide sequence of a genomic target site of interest (i.e., the desired site to be edited) and the type of napDNAbp (e.g., type of Cas protein) present in the base editor, among other factors, such as PAM sequence locations, percent G/C content in the target sequence, the degree of microhomology regions, secondary structures, etc.

[00384] In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a napDNAbp (e.g., a Cas9, Cas9 homolog, or Cas9 variant) to the target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal

alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g. the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies, ELAND (Illumina, San Diego, Calif.), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence is about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length.

[00385] In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. The ability of a guide sequence to direct sequence-specific binding of a base editor to a target sequence may be assessed by any suitable assay. For example, the components of a base editor, including the guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of a base editor disclosed herein, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a base editor, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

[00386] A guide sequence may be selected to target any target sequence. In some embodiments, the target sequence is a sequence within a genome of a cell. Exemplary target sequences include those that are unique in the target genome. For example, for the *S. pyogenes* Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXGG (SEQ ID NO: 134) where NNNNNNNNNNNNXGG (N is A, G, T, or C; and X can be anything) (SEQ ID NO: 135) has a single occurrence in the genome. A unique target sequence in a genome may include an

S. pyogenes Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNXGG (SEQ ID NO: 134) where NNNNNNNNNNNXGG (N is A, G, T, or C; and X can be anything) (SEQ ID NO: 135) has a single occurrence in the genome. For the *S. thermophilus* CRISPR1Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXXAGAAW (SEQ ID NO: 138) where NNNNNNNNNNNNXXAGAAW (N is A, G, T, or C; X can be anything; and W is A or T) (SEQ ID NO: 139) has a single occurrence in the genome. A unique target sequence in a genome may include an *S. thermophilus* CRISPR 1 Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXXAGAAW (SEQ ID NO: 140) where NNNNNNNNNNNNXXAGAAW (N is A, G, T, or C; X can be anything; and W is A or T) (SEQ ID NO: 140) has a single occurrence in the genome. For the *S. pyogenes* Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXGGXG (SEQ ID NO: 142) where NNNNNNNNNNNNXGGXG (N is A, G, T, or C; and X can be anything) (SEQ ID NO: 142) has a single occurrence in the genome. A unique target sequence in a genome may include an *S. pyogenes* Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXGGXG (SEQ ID NO: 144) where NNNNNNNNNNNNXGGXG (N is A, G, T, or C; and X can be anything) (SEQ ID NO: 145) has a single occurrence in the genome. In each of these sequences "M" may be A, G, T, or C, and need not be considered in identifying a sequence as unique.

[00387] In some embodiments, a guide sequence is selected to reduce the degree of secondary structure within the guide sequence. Secondary structure may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker & Stiegler (*Nucleic Acids Res.* 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see, e.g., A. R. Gruber *et al.*, 2008, *Cell* 106(1): 23-24; and PA Carr & GM Church, 2009, *Nature Biotechnology* 27(12): 1151-62). Additional algorithms may be found in

Chuai, G. *et al.*, *DeepCRISPR: optimized CRISPR guide RNA design by deep learning*, *Genome Biol.* 19:80 (2018), and U.S. application Ser. No. 61/836,080, the entireties of each of which are incorporated herein by reference.

[00388] The guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. A tracr mate sequence includes any sequence that has sufficient complementarity with a tracr sequence to promote one or more of: (1) excision of a guide sequence flanked by tracr mate sequences in a cell containing the corresponding tracr sequence; and (2) formation of a complex at a target sequence, wherein the complex comprises the tracr mate sequence hybridized to the tracr sequence. In general, degree of complementarity is with reference to the optimal alignment of the tracr mate sequence and tracr sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm, and may further account for secondary structures, such as self-complementarity within either the tracr sequence or tracr mate sequence. In some embodiments, the degree of complementarity between the tracr sequence and tracr mate sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In some embodiments, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and tracr mate sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin. Preferred loop forming sequences for use in hairpin structures are four nucleotides in length, and most preferably have the sequence GAAA. However, longer or shorter loop sequences may be used, as may alternative sequences. The sequences preferably include a nucleotide triplet (for example, AAA), and an additional nucleotide (for example C or G). Examples of loop forming sequences include CAAA and AAAG. In an embodiment of the disclosure, the transcript or transcribed polynucleotide sequence has at least two or more hairpins. In certain embodiments, the transcript has two, three,

four or five hairpins. In a further embodiment of the disclosure, the transcript has at most five hairpins. In some embodiments, the single transcript further includes a transcription termination sequence; preferably this is a polyT sequence, for example six T nucleotides. Further non-limiting examples of single polynucleotides comprising a guide sequence, a tracr mate sequence, and a tracr sequence are as follows (listed 5′ to 3′), where "N" represents a base of a guide sequence, the first block of lower case letters represent the tracr mate sequence, and the second block of lower case letters represent the tracr sequence, and the final poly-T sequence represents the transcription terminator:

(1) NNNNNNNNgttttgtactctcaagatttaGAAAtaaatcttgcagaagctacaaagataaggctt catgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (SEQ ID NO: 146);

(2) NNNNNNNNNNNNNNNNNNgttttgtactctcaGAAAtgcagaagctacaaagataaggcttcat gccgaaatca acaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT (SEQ ID NO: 147);

(3) NNNNNNNNNNNNNNNNNNNNgttttgtactctcaGAAAtgcagaagctacaaagataaggcttc atgccgaaatca acaccctgtcattttatggcagggtgtTTTTT (SEQ ID NO: 148);

(4) NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAAtagcaagttaaaataaggctagtccgtta tcaacttgaaaa agtggcaccgagtcggtgcTTTTTT (SEQ ID NO: 149);

(5) NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAATAGcaagttaaaataaggctagtccgtta tcaacttgaa aaagtgTTTTTTT (SEQ ID NO: 150); and

(6) NNNNNNNNNNNNNNNNNNNNgttttagagctagAAATAGcaagttaaaataaggctagtccgtta tcaTTTTT TTT (SEQ ID NO: 151).

[00389] In some embodiments, sequences (1) to (3) are used in combination with Cas9 from *S. thermophilus* CRISPR1. In some embodiments, sequences (4) to (6) are used in combination with Cas9 from *S. pyogenes*. In some embodiments, the tracr sequence is a separate transcript from a transcript comprising the tracr mate sequence.

It will be apparent to those of skill in the art that in order to target any of the fusion proteins comprising a Cas9 domain and a thymine alkyltransferase, as disclosed herein, to a target site, e.g., a site comprising a point mutation to be edited, it is typically necessary to co-express the fusion protein

together with a guide RNA, e.g., an sgRNA. As explained in more detail elsewhere herein, a guide RNA typically comprises a tracrRNA framework allowing for Cas9 binding, and a guide sequence, which confers sequence specificity to the Cas9:nucleic acid editing enzyme/domain fusion protein. In some embodiments, the guide RNA comprises a structure 5′-[guide sequence]-guuuuagagcuagaaauagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcuuuuu-3′ (SEQ ID NO: 152), wherein the guide sequence comprises a sequence that is complementary to the target sequence. See U.S. Publication No. 2015/0166981, published June 18, 2015, the disclosure of which is incorporated by reference herein in its entirety. The guide sequence is typically 20 nucleotides long.

[00390] The sequences of suitable guide RNAs for targeting Cas9:nucleic acid editing enzyme/domain fusion proteins to specific genomic target sites will be apparent to those of skill in the art based on the instant disclosure. Such suitable guide RNA sequences typically comprise guide sequences that are complementary to a nucleic sequence within 50 nucleotides upstream or downstream of the target nucleotide to be edited. Some exemplary guide RNA sequences suitable for targeting any of the provided fusion proteins to specific target sequences are provided herein. Additional guide sequences are are well known in the art and can be used with the base editors described herein.Additional exemplary guide sequences are disclosed in, for example, Jinek M., *et al.*, *Science* 337:816-821(2012); Mali P, Esvelt KM & Church GM (2013) Cas9 as a versatile tool for engineering biology, *Nature Methods*, 10, 957-963; Li JF *et al.*, (2013) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9, *Nature Biotechnology*, 31, 688-691; Hwang, W.Y. *et al.*, Efficient genome editing in zebrafish using a CRISPR-Cas system, *Nature Biotechnology* 31, 227-229 (2013); Cong L *et al.*, (2013) Multiplex genome engineering using CRIPSR/Cas systems, *Science*, 339, 819-823; Cho SW *et al.*, (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease, *Nature Biotechnology*, 31, 230-232; Jinek, M. *et al.*, RNA-programmed genome editing in human cells, *eLife*

2, e00471 (2013); Dicarlo, J.E. *et al.*, Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acid Res.* (2013); Briner AE *et al.*, (2014) Guide RNA functional modules direct Cas9 activity and orthogonality, *Mol Cell*, 56, 333-339, the entire contents of each of which are herein incorporated by reference.

Base editor complexes

[00391] Further provided herein are complexes comprising (i) any of the fusion proteins provided herein, and (ii) a guide RNA bound to the Cas9 domain of the fusion protein. Without wishing to be bound by any particular theory, these fusion proteins can be directed by designing a suitable guide RNA to specifically and efficiently target single point mutations in a genome without introducing double-stranded DNA breaks or requiring homology directed repair (HDR). However, the suitability of a target site for base editing (*e.g.*, a point mutation in the genome) is dependent on the presence of a suitably positioned PAM. The broaden PAM compatibility of the Cas9 domains provided herein has the potential to expand the targeting scope of base editors to those target sites that do not lie within approximately 15 nucleotides of a canonical 5′-NGG-3′ PAM sequence. A person of ordinary skill in the art will be able to design a suitable guide RNA (gRNA) sequence to target a desired point mutation based on this disclosure and knowledge in the field. In addition, these fusion proteins comprising a Cas9 domain generate fewer insertions and deletions (indels) and exhibit reduced off-target activity compared to fusion proteins (*e.g.*, base editors) comprising a Cas9 domain that can only recognize the canonical 5′-NGG-3′ PAM sequence.

[00392] In some embodiments, the guide RNA is about 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence. In some embodiments, the guide RNA is 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides long. In some embodiments, the guide RNA comprises a sequence of 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 contiguous nucleotides that is complementary

to a target sequence. In some embodiments, the target sequence is a DNA sequence. In some embodiments, the target sequence is in the genome of an organism. In some embodiments, the organism is a prokaryote. In some embodiments, the prokaryote is a bacterium. In some embodiments, the bacterium is *E. coli*. In some embodiments, the organism is a eukaryote. In some embodiments, the organism is a plant or fungus. In some embodiments, the organism is a vertebrate. In some embodiments, the vertebrate is a mammal. In some embodiments, the mammal is a human. In some embodiments, the organism is a cell. In some embodiments, the cell is a human cell. In some embodiments, the cell is a HEK293T or U2OS cell.

[00393] In some embodiments, the target sequence comprises a sequence associated with a disease or disorder. In some embodiments, the target sequence comprises a point mutation associated with a disease or disorder. In some embodiments, the target sequence comprises a T→C point mutation. In some embodiments, the complex deaminates the target C point mutation, wherein the deamination results in a sequence that is not associated with a disease or disorder. In some embodiments, the target C point mutation is present in the DNA strand that is not complementary to the guide RNA. In some embodiments, the target sequence comprises a T→A point mutation. In some embodiments, the complex deaminates the target A point mutation, and wherein the deamination results in a sequence that is not associated with a disease or disorder. In some embodiments, the target A point mutation is present in the DNA strand that is not complementary to the guide RNA.

[00394] In some embodiments, the complex edits a point mutation in the target sequence. In some embodiments, the point mutation is located between about 10 to about 20 nucleotides upstream of the PAM in the target sequence. In some embodiments, the point mutation is located between about 13 to about 17 nucleotides upstream of the PAM in the target sequence. In some embodiments, the point mutation is about 13 nucleotides upstream of the PAM. In some embodiments, the point mutation is about 14 nucleotides upstream of the PAM. In some embodiments, the point mutation is about 15 nucleotides upstream of the PAM. In some embodiments, the point mutation is about 16 nucleotides

upstream of the PAM. In some embodiments, the point mutation is about 17 nucleotides upstream of the PAM.

[00395] In some embodiments, the complex exhibits increased deamination efficiency of a point mutation in a target sequence that does not comprise the canonical PAM (5′-NGG-3′) at its 3′ end as compared to the deamination efficiency of a complex comprising *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2. In some embodiments, the complex exhibits increased deamination efficiency of a point mutation in a target sequence having a 3′ end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 2-fold, at least 3-fold, at least 4-fold, at least 5-fold, at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold increased as compared to the deamination efficiency of complex comprising the *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2 on the same target sequence. In some embodiments, the 3′ end of the target sequence is directly adjacent to a sequence selected from the group consisting of NGT, NGA, NGC, and NNG, wherein N is an A, G, T, or C. In some embodiments, the 3′ end of the target sequence is directly adjacent to a sequence selected from the group consisting of CGG, AGT, TGG, AGT, CGT, GGG, CGT, TGT, GGT, AGC, CGC, TGC, GGC, AGA, CGA, TGA, GGA, GAA, GAT, and CAA. In some embodiments, deamination activity is measured using high-throughput sequencing.

[00396] In some embodiments, the complex produces fewer indels in a target sequence that does not comprise the canonical PAM (5′-NGG-3′) at its 3′ end as compared to the amount of indels produced by a complex comprising *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2. In some embodiments, the complex produces fewer indels in a target sequence having a 3′ end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 2-fold, at least 3-fold, at least 4-fold, at least 5-fold, at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least

500,000-fold, or at least 1,000,000-fold lower as compared to the amount of indels produced by a complex comprising *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2 on the same target sequence. In some embodiments, the 3′ end of the target sequence is directly adjacent to a sequence selected from the group consisting of NGT, NGA, NGC, and NNG, wherein N is an A, G, T, or C. In some embodiments, the 3′ end of the target sequence is directly adjacent to a sequence selected from the group consisting of CGG, AGT, TGG, AGT, CGT, GGG, CGT, TGT, GGT, AGC, CGC, TGC, GGC, AGA, CGA, TGA, GGA, GAA, GAT, and CAA. In some embodiments, indels are measured using high-throughput sequencing.

[00397] In some embodiments, the complex exhibits a decreased off-target activity as compared to the off-target activity of a complex comprising *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2. In some embodiments, the off-target activity of the complex is at least 2-fold, at least 3-fold, at least 4-fold, at least 5-fold, at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold decreased as compared to the off-target activity of a complex comprising *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2. In some embodiments, the off-target activity is determined using a genome-wide off-target analysis. In some embodiments, the off-target activity is determined using GUIDE-seq.

Methods of using base editors

[00398]Some aspects of this disclosure provide methods of using the Cas9 domains, fusion proteins, or complexes provided herein.

[00399]In one aspect, provided herein are methods comprising contacting a nucleic acid molecule (a) with any of the Cas9 domains or fusion proteins provided herein, and with at least one guide RNA, wherein the guide RNA is about 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence in the nucleic acid molecule; or (b) with a Cas9 domain, a fusion protein comprising a Cas9 domain, or a complex comprising a Cas9

domain, wherein the Cas9 domain is associated with at least one gRNA as provided herein. In some embodiments, the nucleic acid is present in a cell. In some embodiments, the nucleic acid is present in a subject. In some embodiments, the contacting is *in vitro*. In some embodiments, the contacting is *in vivo* in a subject.

[00400]In another aspect, provided herein are methods comprising contacting a cell (a) with any of the Cas9 domains or fusion proteins provided herein, and with at least one guide RNA, wherein the guide RNA is about 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence in the nucleic acid molecule; or (b) with a Cas9 domain, a fusion protein comprising a Cas9 domain, or a complex comprising a Cas9 domain, wherein the Cas9 domain is associated with at least one gRNA as provided herein. In some embodiments, the contacting is *in vitro*. In some embodiments, the contacting is *in vivo* in a subject. In some embodiments, the cell is a prokaryotic cell. In some embodiments, the prokaryotic cell is a bacterium. In some embodiments, the bacterium is *E. coli*. In some embodiments, the cell is a eukaryotic cell. In some embodiments, the eukaryotic cell is a mammalian cell. In some embodiments, the mammalian cell is a human cell. In some embodiments, the cell is a plant or fungal cell.

[00401]In another aspect, provided herein are methods for administering to a subject (a) any of the Cas9 domains or fusion proteins provided herein, and at least one guide RNA, wherein the guide RNA is about 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence in the nucleic acid molecule; or (b) a Cas9 domain, a fusion protein comprising a Cas9 domain, or a complex comprising a Cas9 domain, wherein the Cas9 domain is associated with at least one gRNA as provided herein. In some embodiments, an effective amount of the Cas9 domain, fusion protein, or complex is administered to the subject. In some embodiments, the effective amount is an amount effective for treating a disease or disorder, wherein the disease comprises one or more point mutations in a nucleic acid sequence associated with the disease or disorder.

[00402]In some embodiments, the 3′ end of the target sequence is not immediately adjacent to the canonical PAM sequence (5′-NGG-3′). In some embodiments, the 3′ end of the target sequence is directly adjacent to a sequence selected from the group consisting of NGT, NGA, NGC, and NNG, wherein N is an A, G, T, or C. In some embodiments, the 3′ end of the target sequence is directly adjacent to a sequence selected from the group consisting of CGG, AGT, TGG, AGT, CGT, GGG, CGT, TGT, GGT, AGC, CGC, TGC, GGC, AGA, CGA, TGA, GGA, GAA, GAT, and CAA.

[00403]In some embodiments, the target sequence comprises a sequence associated with a disease or disorder. In some embodiments, the target sequence comprises a point mutation associated with a disease or disorder. In some embodiments, the activity of the Cas9 domain, the Cas9 fusion protein, or the complex results in a correction of the point mutation. In some embodiments, the target sequence comprises a T→C point mutation associated with a disease or disorder, wherein the deamination of the mutant C base results in a sequence that is not associated with a disease or disorder. In some embodiments, the target sequence comprises a A→G, wherein deamination of the C that is base-paired to the mutant G base results in a sequence that is not associated with a disease or disorder. In some embodiments, the target sequence encodes a protein and wherein the point mutation is in a codon and results in a change in the amino acid encoded by the mutant codon as compared to the wild-type codon. In some embodiments, the deamination of the mutant C results in a change of the amino acid encoded by the mutant codon. In some embodiments, the deamination of the mutant C results in the codon encoding the wild-type amino acid. In some embodiments, the target DNA sequence comprises a G→A point mutation associated with a disease or disorder, and wherein the deamination of the mutant A base results in a sequence that is not associated with a disease or disorder. In some embodiments, the target DNA sequence comprises a C→T point mutation associated with a disease or disorder, wherein deamination of the A that is base-paired with the mutant T results in a sequence that is not associated with a disease or disorder. In some embodiments, the target DNA sequence encodes a protein and wherein the point mutation is in a codon and results in a

change in the amino acid encoded by the mutant codon as compared to the wild-type codon. In some embodiments, the deamination of the mutant A results in a change of the amino acid encoded by the mutant codon. In some embodiments, the deamination of the mutant A results in the codon encoding the wild-type amino acid. In some embodiments, the contacting is *in vivo* in a subject. In some embodiments, the subject has or has been diagnosed with a disease or disorder. In some embodiments, the disease or disorder is cystic fibrosis, phenylketonuria, epidermolytic hyperkeratosis (EHK), Charcot-Marie-Toot disease type 4J, neuroblastoma (NB), von Willebrand disease (vWD), myotonia congenital, hereditary renal amyloidosis, dilated cardiomyopathy (DCM), hereditary lymphedema, familial Alzheimer's disease, HIV, Prion disease, chronic infantile neurologic cutaneous articular syndrome (CINCA), desmin-related myopathy (DRM), a neoplastic disease associated with a mutant PI3KCA protein, a mutant CTNNB1 protein, a mutant HRAS protein, or a mutant p53 protein. In some embodiments, the target sequence comprises a sequence located in a genomic locus. In some embodiments, the genomic locus is a HEK site. In some embodiments, the HEK site is HEK site 3 or HEK site 4. In some embodiments, the HEK site comprises a CGG, GGG, TGT, GGT, AGC, CGC, TGC, AGA, or TGA PAM sequence. In some embodiments, the genomic locus is *EMX1*. In some embodiments, the *EMX1* locus comprises a GGG or CAA PAM sequence. In some embodiments, the genomic locus is *VEGFA*. In some embodiments, the *VEGFA* locus comprises a AGT, GGC, GGA, or GAT PAM sequence. In some embodiments, the genomic locus is *FANCF*. In some embodiments, the *FANCF* locus comprises a CGT, GAA, GAT, TGG, AGT, TGT, GGT, CGC, TGC, GGC, AGA, or TGA PAM sequence.

[00404]Some embodiments provide methods for using the Cas9 DNA editing fusion proteins provided herein. In some embodiments, the fusion protein is used to introduce a point mutation into a nucleic acid by deaminating a target nucleobase, *e.g.*, a C or A residue. In some embodiments, the deamination of the target nucleobase results in the correction of a genetic defect, *e.g.*, in the correction of a point mutation that leads to a loss of function in a gene product. In some embodiments, the

genetic defect is associated with a disease or disorder, *e.g.*, a lysosomal storage disorder or a metabolic disease, such as, for example, type I diabetes. In some embodiments, the methods provided herein are used to introduce a deactivating point mutation into a gene or allele that encodes a gene product that is associated with a disease or disorder. For example, in some embodiments, methods are provided herein that employ a fusion protein comprising a Cas9 domain (*e.g.*, a base editor) to introduce a deactivating point mutation into an oncogene (*e.g.*, in the treatment of a proliferative disease). A deactivating mutation may, in some embodiments, generate a premature stop codon in a coding sequence, which results in the expression of a truncated gene product, *e.g.*, a truncated protein lacking the function of the full-length protein.

[00405]In some embodiments, the purpose of the methods provide herein is to restore the function of a dysfunctional gene via genome editing. The Cas9-deaminase fusion proteins provided herein can be validated for gene editing-based human therapeutics *in vitro*, *e.g.*, by correcting a disease-associated mutation in human cell culture. It will be understood by the skilled artisan that the fusion proteins provided herein, *e.g.*, the fusion proteins comprising a Cas9 domain and a cytidine deaminase domain can be used to correct any single T→C or A→G point mutation. In the first case, deamination of the mutant C back to U corrects the mutation, and in the latter case, deamination of the C that is base-paired with the mutant G, followed by a round of replication, corrects the mutation. The fusion proteins comprising a Cas9 domain and one or more adenosine deaminase domains can be used to correct any single G→A or C→T point mutation. In the first case, deamination of the mutant A to I corrects the mutation, and in the latter case, deamination of the A that is base-paired with the mutant T, followed by a round of replication, corrects the mutation.

[00406]An exemplary disease-relevant mutation that can be corrected by the provided fusion proteins *in vitro* or *in vivo* is the H1047R (A3140G) polymorphism in the PI3KCA protein. The phosphoinositide-3-kinase, catalytic alpha subunit (PI3KCA) protein acts to phosphorylate the 3-OH group of the inositol ring of phosphatidylinositol. The PI3KCA gene has been found to be mutated

in many different carcinomas, and thus it is considered to be a potent oncogene.[50] In fact, the A3140G mutation is present in several NCI-60 cancer cell lines, such as, for example, the HCT116, SKOV3, and T47D cell lines, which are readily available from the American Type Culture Collection (ATCC).[51]

[00407]In some embodiments, a cell carrying a mutation to be corrected, *e.g.*, a cell carrying a point mutation, *e.g.*, an A3140G point mutation in exon 20 of the PI3KCA gene, resulting in a H1047R substitution in the PI3KCA protein, is contacted with an expression construct encoding a Cas9 deaminase fusion protein and an appropriately designed sgRNA targeting the fusion protein to the respective mutation site in the encoding PI3KCA gene. Control experiments can be performed where the sgRNAs are designed to target the fusion enzymes to non-C residues that are within the PI3KCA gene. Genomic DNA of the treated cells can be extracted, and the relevant sequence of the PI3KCA genes PCR amplified and sequenced to assess the activities of the fusion proteins in human cell culture.

[00408]It will be understood that the example of correcting point mutations in PI3KCA is provided for illustration purposes and is not meant to limit the instant disclosure. The skilled artisan will understand that the instantly disclosed DNA-editing fusion proteins can be used to correct other point mutations and mutations associated with other cancers and with diseases other than cancer including other proliferative diseases.

[00409]The successful correction of point mutations in disease-associated genes and alleles opens up new strategies for gene correction with applications in therapeutics and basic research. Site-specific single-base modification systems like the disclosed fusions of Cas9 domains and deaminase domains also have applications in "reverse" gene therapy, where certain gene functions are purposely suppressed or abolished. In these cases, site-specifically mutating Trp (TGG), Gln (CAA and CAG), or Arg (CGA) residues to premature stop codons (TAA, TAG, TGA) can be used to abolish protein function *in vitro*, *ex vivo*, or *in vivo*.

[00410]The instant disclosure provides methods for the treatment of a subject diagnosed with a disease associated with or caused by a point mutation that can be corrected by a fusion protein comprising a Cas9 domain and nucleic acid editing domain (*e.g.*, a deaminase domain) provided herein. For example, in some embodiments, a method is provided that comprises administering to a subject having such a disease, *e.g.*, a cancer associated with a PI3KCA point mutation as described above, an effective amount of a Cas9 deaminase fusion protein that corrects the point mutation or introduces a deactivating mutation into the disease-associated gene. In some embodiments, the disease is a proliferative disease. In some embodiments, the disease is a genetic disease. In some embodiments, the disease is a neoplastic disease. In some embodiments, the disease is a metabolic disease. In some embodiments, the disease is a lysosomal storage disease. Other diseases that can be treated by correcting a point mutation or introducing a deactivating mutation into a disease-associated gene will be known to those of skill in the art, and the disclosure is not limited in this respect.

[00411]The instant disclosure provides methods for the treatment of additional diseases or disorders, *e.g.*, diseases or disorders that are associated or caused by a point mutation that can be corrected by deaminase-mediated gene editing. Some such diseases are described herein, and additional suitable diseases that can be treated with the strategies and fusion proteins provided herein will be apparent to those of skill in the art based on the instant disclosure. Exemplary suitable diseases and disorders are listed below. It will be understood that the numbering of the specific positions or residues in the respective sequences depends on the particular protein and numbering scheme used. Numbering might be different, *e.g.*, in precursors of a mature protein and the mature protein itself, and differences in sequences from species to species may affect numbering. One of skill in the art will be able to identify the respective residue in any homologous protein and in the respective encoding nucleic acid by methods well known in the art, *e.g.*, by sequence alignment and determination of homologous residues. Exemplary suitable diseases and disorders include, without limitation, cystic fibrosis (see, *e.g.*, Schwank *et al.*, Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of

cystic fibrosis patients. *Cell stem cell.* **2013**; 13: 653-658; and Wu et. al., Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell stem cell.* **2013;** 13: 659-662, neither of which uses a deaminase fusion protein to correct the genetic defect); phenylketonuria – *e.g.*, phenylalanine to serine mutation at position 835 (mouse) or 240 (human) or a homologous residue in phenylalanine hydroxylase gene (T>C mutation) – see, *e.g.*, McDonald *et al.*, *Genomics.* **1997;** 39:402-405; Bernard-Soulier syndrome (BSS) – *e.g.*, phenylalanine to serine mutation at position 55 or a homologous residue, or cysteine to arginine at residue 24 or a homologous residue in the platelet membrane glycoprotein IX (T>C mutation) – see, *e.g.*, Noris *et al.*, *British Journal of Haematology.* **1997;** 97: 312-320, and Ali *et al.*, *Hematol.* **2014;** 93: 381-384; epidermolytic hyperkeratosis (EHK) – *e.g.*, leucine to proline mutation at position 160 or 161 (if counting the initiator methionine) or a homologous residue in keratin 1 (T>C mutation) – see, *e.g.*, Chipev *et al.*, *Cell.* **1992;** 70: 821-828, see also accession number P04264 in the UNIPROT database at www[dot]uniprot[dot]org; chronic obstructive pulmonary disease (COPD) – *e.g.*, leucine to proline mutation at position 54 or 55 (if counting the initiator methionine) or a homologous residue in the processed form of α_1-antitrypsin or residue 78 in the unprocessed form or a homologous residue (T>C mutation) – see, *e.g.*, Poller *et al.*, *Genomics.* **1993;** 17: 740-743, see also accession number P01011 in the UNIPROT database; Charcot-Marie-Toot disease type 4J – *e.g.*, isoleucine to threonine mutation at position 41 or a homologous residue in FIG4 (T>C mutation) – see, *e.g.*, Lenk *et al.*, PLoS Genetics. 2011; 7: e1002104; neuroblastoma (NB) – *e.g.*, leucine to proline mutation at position 197 or a homologous residue in Caspase-9 (T>C mutation) – see, *e.g.*, Kundu *et al.*, *3 Biotech.* **2013, 3:225-234**; von Willebrand disease (vWD) – *e.g.*, cysteine to arginine mutation at position 509 or a homologous residue in the processed form of von Willebrand factor, or at position 1272 or a homologous residue in the unprocessed form of von Willebrand factor (T>C mutation) – see, *e.g.*, Lavergne *et al.*, *Br. J. Haematol.* **1992**, see also accession number P04275 in the UNIPROT database**;** 82: 66-72; myotonia congenital – *e.g.*, cysteine to arginine mutation at position 277 or a homologous residue in the muscle

chloride channel gene CLCN1 (T>C mutation) – see, *e.g.*, Weinberger *et al.*, *The J. of Physiology.* **2012;** 590: 3449-3464; hereditary renal amyloidosis – *e.g.*, stop codon to arginine mutation at position 78 or a homologous residue in the processed form of apolipoprotein AII or at position 101 or a homologous residue in the unprocessed form (T>C mutation) – see, *e.g.*, Yazaki *et al.*, *Kidney Int.* **2003;** 64: 11-16; dilated cardiomyopathy (DCM) – *e.g.*, tryptophan to Arginine mutation at position 148 or a homologous residue in the FOXD4 gene (T>C mutation), see, *e.g.*, Minoretti et. al., *Int. J. of Mol. Med.* **2007;** 19: 369-372; hereditary lymphedema – *e.g.*, histidine to arginine mutation at position 1035 or a homologous residue in VEGFR3 tyrosine kinase (A>G mutation), see, *e.g.*, Irrthum *et al.*, *Am. J. Hum. Genet.* **2000;** 67: 295-301; familial Alzheimer's disease – *e.g.*, isoleucine to valine mutation at position 143 or a homologous residue in presenilin1 (A>G mutation), see, *e.g.*, Gallo et. al., *J. Alzheimer's disease.* **2011;** 25: 425-431; Prion disease – *e.g.*, methionine to valine mutation at position 129 or a homologous residue in prion protein (A>G mutation) – see, *e.g.*, Lewis et. al., *J. of General Virology.* **2006;** 87: 2443-2449; chronic infantile neurologic cutaneous articular syndrome (CINCA) – *e.g.*, Tyrosine to Cysteine mutation at position 570 or a homologous residue in cryopyrin (A>G mutation) – see, *e.g.*, Fujisawa et. al. *Blood.* **2007;** 109: 2903-2911; and desmin-related myopathy (DRM) – *e.g.*, arginine to glycine mutation at position 120 or a homologous residue in αB crystallin (A>G mutation) – see, *e.g.*, Kumar *et al.*, *J. Biol. Chem.* **1999;** 274: 24137-24141. The entire contents of all references and database entries is incorporated herein by reference.

Pharmaceutical compositions

[00412]Other aspects of the present disclosure relate to pharmaceutical compositions comprising any of the various components described herein (e.g., including, but not limited to, the napDNAbps, fusion proteins, guide RNAs, and complexes comprising fusion proteins and guide RNAs).

[00413]The term "pharmaceutical composition", as used herein, refers to a composition formulated for pharmaceutical use. In some embodiments, the pharmaceutical composition further comprises a pharmaceutically acceptable carrier. In some embodiments, the pharmaceutical composition

comprises additional agents (e.g. for specific delivery, increasing half-life, or other therapeutic compounds).

[00414]As used here, the term "pharmaceutically-acceptable carrier" means a pharmaceutically-acceptable material, composition or vehicle, such as a liquid or solid filler, diluent, excipient, manufacturing aid (e.g., lubricant, talc magnesium, calcium or zinc stearate, or steric acid), or solvent encapsulating material, involved in carrying or transporting the compound from one site (e.g., the delivery site) of the body, to another site (e.g., organ, tissue or portion of the body). A pharmaceutically acceptable carrier is "acceptable" in the sense of being compatible with the other ingredients of the formulation and not injurious to the tissue of the subject (e.g., physiologically compatible, sterile, physiologic pH, etc.). Some examples of materials which can serve as pharmaceutically-acceptable carriers include: (1) sugars, such as lactose, glucose and sucrose; (2) starches, such as corn starch and potato starch; (3) cellulose, and its derivatives, such as sodium carboxymethyl cellulose, methylcellulose, ethyl cellulose, microcrystalline cellulose and cellulose acetate; (4) powdered tragacanth; (5) malt; (6) gelatin; (7) lubricating agents, such as magnesium stearate, sodium lauryl sulfate and talc; (8) excipients, such as cocoa butter and suppository waxes; (9) oils, such as peanut oil, cottonseed oil, safflower oil, sesame oil, olive oil, corn oil and soybean oil; (10) glycols, such as propylene glycol; (11) polyols, such as glycerin, sorbitol, mannitol and polyethylene glycol (PEG); (12) esters, such as ethyl oleate and ethyl laurate; (13) agar; (14) buffering agents, such as magnesium hydroxide and aluminum hydroxide; (15) alginic acid; (16) pyrogen-free water; (17) isotonic saline; (18) Ringer's solution; (19) ethyl alcohol; (20) pH buffered solutions; (21) polyesters, polycarbonates and/or polyanhydrides; (22) bulking agents, such as polypeptides and amino acids (23) serum component, such as serum albumin, HDL and LDL; (22) C2-C12 alcohols, such as ethanol; and (23) other non-toxic compatible substances employed in pharmaceutical formulations. Wetting agents, coloring agents, release agents, coating agents, sweetening agents, flavoring agents, perfuming agents, preservative and antioxidants can also be present in the

formulation. The terms such as "excipient", "carrier", "pharmaceutically acceptable carrier" or the like are used interchangeably herein.

[00415]In some embodiments, the pharmaceutical composition is formulated for delivery to a subject, e.g., for gene editing. Suitable routes of administrating the pharmaceutical composition described herein include, without limitation: topical, subcutaneous, transdermal, intradermal, intralesional, intraarticular, intraperitoneal, intravesical, transmucosal, gingival, intradental, intracochlear, transtympanic, intraorgan, epidural, intrathecal, intramuscular, intravenous, intravascular, intraosseus, periocular, intratumoral, intracerebral, and intracerebroventricular administration.

[00416]In some embodiments, the pharmaceutical composition described herein is administered locally to a diseased site (e.g., tumor site). In some embodiments, the pharmaceutical composition described herein is administered to a subject by injection, by means of a catheter, by means of a suppository, or by means of an implant, the implant being of a porous, non-porous, or gelatinous material, including a membrane, such as a sialastic membrane, or a fiber.

[00417]In other embodiments, the pharmaceutical composition described herein is delivered in a controlled release system. In one embodiment, a pump may be used (see, e.g., Langer, 1990, Science 249:1527-1533; Sefton, 1989, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald et al., 1980, Surgery 88:507; Saudek et al., 1989, N. Engl. J. Med. 321:574). In another embodiment, polymeric materials can be used. (See, e.g., Medical Applications of Controlled Release (Langer and Wise eds., CRC Press, Boca Raton, Fla., 1974); Controlled Drug Bioavailability, Drug Product Design and Performance (Smolen and Ball eds., Wiley, New York, 1984); Ranger and Peppas, 1983, Macromol. Sci. Rev. Macromol. Chem. 23:61. See also Levy et al., 1985, Science 228:190; During et al., 1989, Ann. Neurol. 25:351; Howard et al., 1989, J. Neurosurg. 71:105.) Other controlled release systems are discussed, for example, in Langer, supra.

[00418]In some embodiments, the pharmaceutical composition is formulated in accordance with routine procedures as a composition adapted for intravenous or subcutaneous administration to a

subject, e.g., a human. In some embodiments, pharmaceutical composition for administration by injection are solutions in sterile isotonic aqueous buffer. Where necessary, the pharmaceutical can also include a solubilizing agent and a local anesthetic such as lignocaine to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the pharmaceutical is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the pharmaceutical composition is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients can be mixed prior to administration.

[00419]A pharmaceutical composition for systemic administration may be a liquid, e.g., sterile saline, lactated Ringer's or Hank's solution. In addition, the pharmaceutical composition can be in solid forms and re-dissolved or suspended immediately prior to use. Lyophilized forms are also contemplated.

[00420]The pharmaceutical composition can be contained within a lipid particle or vesicle, such as a liposome or microcrystal, which is also suitable for parenteral administration. The particles can be of any suitable structure, such as unilamellar or plurilamellar, so long as compositions are contained therein. Compounds can be entrapped in "stabilized plasmid-lipid particles" (SPLP) containing the fusogenic lipid dioleoylphosphatidylethanolamine (DOPE), low levels (5-10 mol%) of cationic lipid, and stabilized by a polyethyleneglycol (PEG) coating (Zhang Y. P. *et al.*, *Gene Ther.* 1999, 6:1438-47). Positively charged lipids such as N-[1-(2,3-dioleoyloxi)propyl]-N,N,N-trimethyl-amoniummethylsulfate, or "DOTAP," are particularly preferred for such particles and vesicles. The preparation of such lipid particles is well known. See, *e.g.*, U.S. Patent Nos. 4,880,635; 4,906,477; 4,911,928; 4,917,951; 4,920,016; and 4,921,757; each of which is incorporated herein by reference.

[00421]The pharmaceutical composition described herein may be administered or packaged as a unit

dose, for example. The term "unit dose" when used in reference to a pharmaceutical composition of the present disclosure refers to physically discrete units suitable as unitary dosage for the subject, each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required diluent; *i.e.*, carrier, or vehicle.

[00422]Further, the pharmaceutical composition can be provided as a pharmaceutical kit comprising (a) a container containing a compound of the invention in lyophilized form and (b) a second container containing a pharmaceutically acceptable diluent (e.g., sterile water) for injection. The pharmaceutically acceptable diluent can be used for reconstitution or dilution of the lyophilized compound of the invention. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

[00423]In another aspect, an article of manufacture containing materials useful for the treatment of the diseases described above is included. In some embodiments, the article of manufacture comprises a container and a label. Suitable containers include, for example, bottles, vials, syringes, and test tubes. The containers may be formed from a variety of materials such as glass or plastic. In some embodiments, the container holds a composition that is effective for treating a disease described herein and may have a sterile access port. For example, the container may be an intravenous solution bag or a vial having a stopper pierceable by a hypodermic injection needle. The active agent in the composition is a compound of the invention. In some embodiments, the label on or associated with the container indicates that the composition is used for treating the disease of choice. The article of manufacture may further comprise a second container comprising a pharmaceutically-acceptable buffer, such as phosphate-buffered saline, Ringer's solution, or dextrose solution. It may further include other materials desirable from a commercial and user standpoint, including other buffers, diluents, filters, needles, syringes, and package inserts with instructions for use.

Delivery methods

[00424]In some aspects, the invention provides methods comprising delivering one or more polynucleotides, such as or one or more vectors as described herein encoding one or more components described herein, one or more transcripts thereof, and/or one or proteins transcribed therefrom, to a host cell. In some aspects, the invention further provides cells produced by such methods, and organisms (such as animals, plants, or fungi) comprising or produced from such cells. In some embodiments, a base editor as described herein in combination with (and optionally complexed with) a guide sequence is delivered to a cell. Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding components of a base editor to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (e.g. a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, Science 256:808-813 (1992); Nabel & Felgner, TIBTECH 11:211-217 (1993); Mitani & Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6(10):1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer & Perricaudet, British Medical Bulletin 51(1):31-44 (1995); Haddada et al., in Current Topics in Microbiology and Immunology Doerfler and Bihm (eds) (1995); and Yu et al., Gene Therapy 1:13-26 (1994).

[00425]Methods of non-viral delivery of nucleic acids include lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable

for efficient receptor-recognition lipofection of polynucleotides include those of Feigner, WO 91/17424; WO 91/16024. Delivery can be to cells (e.g. in vitro or ex vivo administration) or target tissues (e.g. in vivo administration). The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

[00426]The use of RNA or DNA viral based systems for the delivery of nucleic acids take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo) or they can be used to treat cells in vitro, and the modified cells may optionally be administered to patients (ex vivo). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[00427]The tropism of a viruses can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression.

Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., J. Virol. 66:2731-2739 (1992); Johann et al., J. Virol. 66:1635-1640 (1992); Sommnerfelt et al., Virol. 176:58-59 (1990); Wilson et al., J. Virol. 63:2374-2378 (1989); Miller et al., J. Virol. 65:2220-2224 (1991); PCT/US94/05700). In applications where transient expression is preferred, adenoviral based systems may be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors may also be used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., Virology 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., J. Virol. 63:03822-3828 (1989).

[00428]Packaging cells are typically used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producing a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains

a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line may also be infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV. Additional methods for the delivery of nucleic acids to cells are known to those skilled in the art. See, for example, US20030087817, incorporated herein by reference.

Kits, vectors, cells

[00429]Some aspects of this disclosure provide kits comprising a nucleic acid construct, comprising (a) a nucleotide sequence encoding a Cas9 domain or a fusion protein comprising a Cas9 domain as provided herein; and (b) a heterologous promoter that drives expression of the sequence of (a). In some embodiments, the kit further comprises an expression construct encoding a guide RNA backbone, wherein the construct comprises a cloning site positioned to allow the cloning of a nucleic acid sequence identical or complementary to a target sequence into the guide RNA backbone.

[00430]Some aspects of this disclosure provide polynucleotides encoding a Cas9 domain or a fusion protein comprising a Cas9 domain as provided herein. Some aspects of this disclosure provide vectors comprising such polynucleotides. In some embodiments, the vector comprises a heterologous promoter driving expression of polynucleotide.

[00431]In one aspect, provided herein are methods comprising contacting a cell with a kit provided herein. In another aspect, provided herein are methods comprising contacting a cell with a vector provided herein. In some embodiments, the vector is transfected into the cell. In some embodiments, the vector is transfected into the cell using a suitable transfection reaction. Transfection reactions may be carried out, for example, using electroporation, heat shock, or a composition comprising a cationic lipid. Cationic lipids suitable for the transfection of nucleic acid molecules are provided in, for example, Patent Publication WO2015/035136, published March 12, 2015, entitled “Delivery System

for Functional Nucleases"; the entire contents of which is incorporated by reference herein.

[00432]Some aspects of this disclosure provide cells comprising a Cas9 domain, a fusion protein, a nucleic acid molecule, and/or a vector as provided herein.

[00433] The description of exemplary embodiments of the reporter systems (*e.g.*, GFP) herein is provided for illustration purposes only and not meant to be limiting. Additional reporter systems, *e.g.*, variations of the exemplary systems described in detail above, are also embraced by this disclosure.

REFERENCES

[00434] 1. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).

[00435] 2. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, 1248 (2016).

[00436] 3. Gaudelli, N. M. et al. Programmable base editing of A.T to G.C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).

[00437] 4. Rees, H. A. et al. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* **8**, 15790 (2017).

[00438] 5. Kim, J.-S. Precision genome engineering through adenine and cytosine base editing. *Nat. Plants* **4**, 148–151 (2018).

EXAMPLES

EXAMPLE 1

Identification of PAM sequences that SpCas9 and xCas9 have low activity

[00439] A key limitation to the use of CRISPR-Cas9 domains for genome editing and other applications is the requirement that a protospacer adjacent motif (PAM) be present at the target site.

For the most commonly used Cas9 from *Streptococcus pyogenes* (SpCas9), this PAM requirement is NGG. No natural or engineered Cas9 variants shown to function efficiently in mammalian cells offer a PAM less restrictive than NGG. Phage-assisted continuous evolution (PACE) was used to evolve the wild type SpCas9 and an expanded PAM SpCas9 variant (xCas9) that can recognize a broad range of PAM sequences. The PAM compatibility of xCas9 is the broadest reported to date among Cas9s active in mammalian cells, and supports applications in human cells including targeted transcriptional activation, nuclease-mediated gene disruption, and both cytidine and adenine base editing.

[00440] Here, phage-assisted continuous evolution (PACE) is used for identification on PAMs that spCas9 and xCas9 have low activity. During PACE, host *E. coli* cells continuously dilute an evolving population of bacteriophages (selection phage, SP). Since dilution occurs faster than cell division but slower than phage replication, only the SP, and not the host cells, can accumulate mutations. Each SP carries a gene to be evolved instead of a phage gene (gene III) that is required for the production of infectious progeny phage. SP containing desired gene variants trigger host-cell gene III expression from the accessory plasmid (AP) and the production of infectious SP that propagate the desired variants. Phage encoding inactive variants do not generate infectious progeny and are rapidly diluted out of the culture vessel (**FIG. 1A**). As phage replication can occur in as little as 10 minutes, PACE enables hundreds of generations of directed evolution to occur per week without researcher intervention.

[00441] To link Cas9 DNA recognition to phage propagation during PACE, a bacterial one-hybrid selection in which the SP encodes a catalytically dead SpCas9 (dCas9) fused to the ω subunit of bacterial RNA polymerase was developed (**FIG. 1A**). When this fusion binds an AP-encoded sgRNA and a PAM and protospacer upstream of gene III in the AP, RNA polymerase recruitment causes gene III expression and phage propagation (**FIG. 1B**). A library of all 64 possible NNN PAM sequences at the target protospacer in the AP, so that SP encoding Cas9 variants with broader PAM compatibility would replicate in a larger fraction of host cells and thus experience a fitness advantage, was

generated. After overnight propagation. As expected, xCas9 are less stringent on PAM requirement. Both SpCas9 or xCas9 exhibited low activity on NAA, NAC, and NAT PAMs (**FIG. 1C**). The following experiments were designed to identify Cas9 variants that are able to bind to NAA, NAC, and NAT PAMs.

EXAMPLE 2

Phage Assisted Non-continuous Evolution (PANCE) of Cas9 variants for expanded PAM compatibility

[00442] Phage-assisted non-continuous evolution (PANCE) system was used to further evolve SpCas9 and xCas9 for identification of Cas9 variants that can recognize non-NGG PAMs. In the PANCE system, the SP is iteratively passaged through serial dilution in host cells in order to evolve SpCas9 and/or xCas9 proteins that bind to all possible The PANCE system preferentially replicates Cas9 variants that bind a greater variety of PAM sequences, similar to PACE, but with lower stringency since there is no outflow of phage. Although lower in stringency, the PANCE system allows for higher throughput, enabling evolution towards multiple targets (e.g., NAA, NAC, NAT PAMS) simultaneously.

[00443] In this experiment, SPs were iteratively passaged through serial dilution in host cells to evolve either SpCas9 or xCas9 proteins capable of binding to all 16 NAN PAM target sequences. In PANCE, *E. coli* host cells transformed with an AP and mutagenesis plasmid (MP) or dilution plasmid (DP) are plated in individual wells of a multi-well plate and grown to log phase. Selection phages, are then introduced and mutagenesis is induced with arabinose or aTc. The SPs are then grown for at least 6 additional hours, before being collected and used to infect the next multi-well plate of *E. coli* host cells that have grown to log phase (**FIG. 2A**). Each one of these infection-incubation-collection cycles is referred to as a "passage".

[00444] Increased recognition of non-NGG PAMs were observed in both SpCas9 and xCas9 as they were evolved through more passages in PANCE. FIG. 2B shows evolving SpCas9 and xCas9's ability to recognize all 64 PAMs for passage 2, passage 12 and passage 16. After performing 19 rounds of

selection in PANCE and sequencing the surviving phage pools (**FIG. 36**), mutations largely differing according to the third base of the NAN PAM targeted for evolution were observed. For example, variants selected on NAA enriched for Gly, Ile, or Lys at position 1333, while those selected for NAT enriched for Gln or Leu at position 1335. Finally, variants evolved to bind NAC enriched simultaneously for Gln at position 1335 and Asn at position 1337.

[00445] The clones of mutated SpCas9 and xCas9 variants that were able to recognize NAA PAMs were isolated and sequenced for identification of mutations in Cas9. FIG. 3A shows mutations in SpCas9 at passage 12 that can recognize CAA, GAT, ATG, or AGC PAMs. FIG. 4A shows mutations in SpCas9 at passage 19 that can recognize ATG, CAA, or GAA PAMs. Further, the wild type SpCas9 clones, e.g., CAA-3, GAT-2, ATG-2, ATG-3, or AGC-3 in passage 12 were tested using the luciferase assay described above for their ability to recognize all 64 PAMs, as shown in FIG. 3B. Similarly, the wild type SpCas9 clones, e.g., CAA-1, CAA-2, GAA-1, GAA-2, GAC-5, GAT-1, GAT-3, AGC-1, AGC-3, AGC-6. ATG-3, or ATG-6 in passage 19 were tested using the luciferase assay described above for their ability to recognize all 64 PAMs, as shown in FIG. 4B.

[00446] Similarly, FIG. 5A shows mutations in xCas9 at passage 12 that can recognize TAT, GTA, or CAC PAMs, and FIG. 6A shows mutations in xCas9 at passage 19 that can recognize AAA, GCC, or TAA PAMs. Further, xCas9 mutant clones, e.g., TAT-1, TAT-3, GTA-1, GTA-3, or CAC-2 in passage 12 were tested using the luciferase assay described above for their ability to recognize all 64 PAMs, as shown in FIG. 5B. Similarly, xCas9 mutant clones, e.g., AAA-1, TAA-2, TAA-5, TAT-5, CAC-5, CAC-6, GTA-2, GTA-7, GCC-2, GCC-5, or GCC-8 in passage 18 were tested using the luciferase assay described above for their ability to recognize all 64 PAMs, as shown in FIG. 6B.

[00447] To test if mutations evolved during PANCE in bacteria are compatible with xCas9 function in mammalian cells, SpCas9 and xCas9 variants were characterized for their activity and PAM compatibility in human cells in two contexts: adenine base editing and genomic DNA cutting. Additionally, to further characterize genomic DNA cleavage in human cells by xCas9 variants, we

targeted endogenous genomic sites in HEK293T cells and measured indel formation by high-throughput sequencing (HTS).

[00448] To evaluate C•G-to-T•A base editing activity of xCas9 variants, SpCas9 was substituted with xCas9 3.7 and 3.6 in the third-generation (BE3) base editor architecture. Both xCas9–BE3s were transfected into mammalian cells to compare editing efficiency. The xCas9-BE3 protein demonstrated base editing activity only on CGT and CGG PAMs, whereas the ATG2-BE3 protein demonstrated base editing activity on CAG and ATG PAMs, the CAA3-BE3 protein demonstrated base editing activity on CGG PAMs, and the TAT1-BE3 protein demonstrated base editing activity on CAT PAMs (**FIG. 7**).

[00449] The xCas9 protein produced indels in CAG, ATG, CAT, CGT, and CGG PAMs, whereas the ATG2 protein produced indels in CAG and CGG PAMs, the CAA3 protein produced indels in CAT and CGG PAMs, and the TAT1 protein produced indels in CAT PAMs (**FIG. 7**).

[00450] Thus, the PANCE evolved spCas9 variants have some activity *in vitro* on non-NGG PAMs.

[00451] Additionally, the xCas9-passage 12-TAT1 (N6) variant was subjected to further PANCE evolution. A comparison of xCas9-passage 12-TAT1 to SpCas9 in various amino acid residues was shown in **FIG. 9A.** The clones resulting from further PANCE evolution of the xCas9-passage 12-TAT1 (N6) variant are shown in **FIGs. 10-11**. FIG. 12 shows evolving's xCas9-passage 12-TAT1 variant's ability to recognize all 64 PAMs for passage 2, passage 12 and passage 16.

EXAMPLE 3

Selection improvement allows the evolution of NAA PAM binding activity

[00452] Despite enriching for multiple consensus mutations in the PAM-interacting domain (PID), (D1135N/E1219V/Q1221H/H1264Y/A1320V/R1333K), the NAA-targeted PANCE-evolved mutants exhibited low activity when subcloned into C to T base editors (CBEs) and tested for base conversion on sites containing NAA PAMs in mammalian cells (**FIGs. 7, 37C**). One possible explanation is that

evolving increased binding activity might require increased selection stringency, and three strategies were implemented to accomplish this.

[00453] First, two variants evolved to bind a CAA PAM in the initial PANCE assay were selected and subjected to PACE using a dual-AP system. Here, each AP provides one half of slit-intein pIII under control of an orthogonal Cas9 1-hybrid circuit, requiring ω-dCas9 to successfully bind two distinct protospacer-PAM motifs to produce full-length pIII (**FIGs. 13A, 13B, 37A**). These experiments led to the acquisition of a few additional consensus mutations (**FIGs. 14A, 37B**). This in turn led to improvements in CBE on sites (**FIGs. 14B, 37C**) and increased percentages of indels in mammalian cells (**FIG. 14C**).

[00454] Next, the total amount of Cas9 present in the selection was limited by using a split-intein to divide ω-dCas9 into two halves and encoding only the C-terminal half (which contains the PID) on the SP. Production of large amounts of ω-dCas9 by the SP might lead to saturation of binding to protospacer-PAM sites AP despite the presence of a non-optimal PAM (**FIG. 15A**). Indeed, using higher concentrations of SpCas9 in *in vitro* PAM depletion assays can lead to depletion of non-canonical PAM sequences (REF). Here, residues 574-1368 of Cas9 fused to NpuC ($dCas9_C$) reside on the phage, while ω-dCas9 (1-573) fused to NpuN (ω-$dCas9_N$) is provided on a complimentary plasmid (CP) (**FIGs. 15B, 37A**). This strategy allows the total amount of full-length SpCas9 produced in the host cells in PACE to be user-defined on the CP.

[00455] The consensus mutations obtained from the dual-AP selection were subcloned into a split-intein ω-dCas9 format. However, several mutations (T10A/I322V/S409I,E427G) had accumulated in the 1-573 region over the course of the previous selection. These mutations were incorporated into ω-$dCas9_N$ and investigated their effect on Cas9 DNA binding in overnight phage propagation assays using an evolved $dCas9_C$ phage clone (P4.72.5). High phage propagation was observed on host cells containing a CP encoding ω-$dCas9_N$(T10A/I322V/S409I/E427G), suggesting

that the mutations might have a beneficial effect on Cas9 binding. Therefore, these four mutations were incorporated into ω-dCas9$_N$ for all future evolutions (hereon referred to as ω-dCas9$_{N\text{-}mut}$).

[00456] Thus, the evolved dCas9$_C$ was subjected to two subsequent evolutions using host cells encoding a medium-copy AP containing an AAA PAM and low-copy CPs providing ω-dCas9$_{N\text{-}mut}$ from increasingly weak constitutive promoters. These rounds lead to the accumulation of additional mutations in the PID, including D1180G, which was present in several sequenced clones (**FIGs. 16A, 37B**). The Cas9s evolved through this split-intein method exhibited a large increase in mammalian cell base editing activity, with more than double the activity of our previous variants on most NAA sites tested (**FIGs. 17, 37C**). Additionally, the Cas9s evolved through this split-intein method exhibited a large increase in percentage of indels in most NAA PAMs tested (**FIG. 18**).

[00457] Finally, to further increase selection stringency, gVI, whose protein product pVI is essential for phage propagation, was removed from the phage genome for use as an orthogonal selection marker for phage propagation on a second AP (**FIGs. 27A**). Both previously described selection principles were employed, requiring a split-intein ω-dCas9 to bind two distinct protospacers on APs providing both gIII and gVI (**FIG. 37A**). Thus, three dCas9$_C$ clones from pervious evolutions (P13.3.3, P10.5.192.7, P10.6.192.1) were subjected to this highest-stringency selection in PACE, resulting in additional mutations in the PID-notably R1114G and L1318S, which both converged to a high degree (**FIG. 37B**).

[00458] Unfortunately, these variants proved to be inactive in mammalian cell CBE experiments (**FIG. 37C**). The large numbers of mutations present in these highly evolved variants, especially those outside the PID, might prove deleterious to expression and/or nuclease activity. To address this, DNA shuffling was performed of the C-terminal portion (residues 574-1368) of the pool of variants from this final evolution with that of wild-type Cas9 and re-subjected the resulting library to the most stringent binding selection. This led to the isolation of several clones that exhibited improved CBE activity at both NAA and NGA sites in mammalian cells, most notably clone P16s.4-5

(R1114G/D1135N/V1139A/D1180G/E1219V/Q1221H/A1320V/R1333K) (**FIG. 37B**), which exhibited the highest levels of activity across all sites tested amongst the variants (**FIG. 37C**).

EXAMPLE 4

Evolution of Cas9 variants that recognize NAC or NAT PAM sequences

[00459] The strategy evolved in Example III was employed in evolving toward NAT and NAC PAMs in SpCas9 and xCas9 proteins to minimize the accumulation of potentially deleterious bystander mutations. To ensure the variants retained nuclease activity, the dCas9 from the SP pool was evolved to bind either a TAT or CAC PAM in PANCE to a nuclease-active form and passed the resulting library through a modified version of a previously reported bacterial DNA cleavage selection (data not shown). Here, Cas9 variants are challenged for their ability to bind to and cleave a protospacer-PAM sequence on a high-copy plasmid that also encodes a conditionally toxic gene (sacB). The surviving cells should then encode Cas9 variants with mutations that confer binding to a specific PAM and are compatible with nuclease activity.

[00460] From these experiments, two clones were isolated that exhibited DNA cleavage activity on a selection plasmid containing a TAT PAM with PID consensus mutations of D1135N/E1219V/Q1221H/P1321S/R1335L, and one clone that cleaved a selection plasmid with a CAC PAM with PID mutations N1135D/E1219V/D1332N/R1335Q/T1337N (**FIGs. 37D, 37E**). These nuclease-active TAT and CAC variants were then converted into split-intein ω –dCas9 format and evolved in PACE using host cells encoding APs with either NAT (AAT or TAT) or NAC (AAC, TAC, or CAC) PAMs, respectively. These experiments resulted in the enrichment of several additional PID mutations, including R1114G, which arose independently in all three PAM trajectories (NAA, NAT, and NAC) (**FIGs. 37D, 37E**), suggesting that this mutation may be generally beneficial for modifying PAM recognition by the PID.

[00461] Next, gVI was removed from the genome of these evolved SP pools, which were subjected to additional selection in PACE using a dual-AP system containing two distinct protospacers

and either an AAT or TAC PAM driving gIII/gVI expression. A Y1131C mutation was enriched in the SP pool evolved on AAT (**FIG. 37E**); however, variants carrying this mutation were inactive in mammalian cell BE experiments (Supplementary Figure XX). Because no additional functional mutations in the PID were observed, the most active NAT PAM-targeting variant was selected from the split-intein ω–dCas9 evolution (clone P12.3.b9-8) to move forward with. This variant contained the PID mutations R1114G/D1135N/D1180G/G1218S/E1219V/Q1221H/P1249S/E1253K/ P1321S/D1332G/R1335L (**FIG. 37E**).

[00462] Several additional mutations were also enriched in the SP pool selected for binding to a TAC PAM in the split-intein ω-dCas9/dual protospacer PACE. The C-terminal portion (residues 574-1368) of this pool was shuffled with that of wild-type Cas9 and re-challenged the resulting library with our most stringent binding selection. From the surviving SP pool, clone P17s.1.7-4 with the PID mutations R1114G/D1135N/E1219V/D1332N/R1335Q/T1337N/S1338T/H1249R was isolated from the surviving pool (**FIG. 37C**).

EXAMPLE 5

Mutations outside of the PID

[00463] Structural studies of the SpCas9 suggest that residues in the PID mediate PAM specificity (REF). Indeed, most of the previous efforts to engineer or evolve SpCas9 to accept alternative PAMs have focused on modulating this region of the protein. However, because PANCE and PACE experiments involved mutagenesis of either the entire SpCas9 sequence or residues 574-1368 (in the case of split-intein ω –dCas9), there was an enrichment of a number of mutations outside of the PID. Because many of these mutations fell within the RuvC or HNH nuclease domains, some may negatively impact Cas9 nuclease activity. However, other mutations in the helical domain consistently enriched across several independent evolving populations, suggesting that they may confer a beneficial effect on Cas9 DNA binding/unwinding.

[00464] Therefore, to minimize the deleterious effects from bystander mutation accumulation in the nuclease domains but also to preserve beneficial mutations in the helical domain, the evolved PIDs from Example 4 were transferred onto a fixed N-terminal sequence that included the mutations T10A/I322V/S409I/E427G shown to improve phage propagation in the split-intein ω–dCas9 selection, as well as R654L/R753G, which consistently enriched across multiple independently evolving SP pools. The addition of these mutations to CBEs containing the PIDs of NAA variant P16.4-5 and NAC variant P17.1.7-4 improved CBE activity in mammalian cells across several sites when compared to just the PID mutations alone (data not shown). A smaller effect was observed for NAT variant P12.3.b9-8, but because there did not appear to be a decrease in overall CBE activity (data not shown), there N-terminal mutations were incorporated into all three final variants, from hereon referred to as NRRH, NRCH, and NRTH, which are derived from clones P16.4-5, P17.1.7-4, and P12.3.b9-8, respectively.

EXAMPLE 6

Characterization of PAM specificity through bacterial depletion

[00465] To better characterize the PAM specificities of the evolved variants, bacterial PAM depletion was performed using a library consisting of 4Ns following the protospacer (**FIGs. 19A-19C**). For comparison, depletion experiments were also performed with wild-type Cas9 that acts on an NGG PAM sequence (SpCas9-NG) in parallel. Cells were plated after 1 or 3 h or overnight expression of the SpCas9 variant from an inducible promoter to better resolve any kinetic differences in PAM sequence preference. As expected, depletion scores of any given PAM increased with longer induction times (data not shown), with the shortest induction times resulting in the most noticeable sequence preferences (data not shown).

[00466] For example, at 1 hour (h) induction, NRRH exhibited a strong preference for C at the 4th PAM position, a mixed preference for G/A at positions 2 and 3 and a moderate preference for G at position 1 (**FIGs. 20, 38A**). However, longer induction times resulted in more relaxed specificity at

all positions. Similarly, NRCH showed a strong preference for G at position 2 and a moderate preference for pyrimidines at position 4 (**FIG. 38A**) at 1 h induction, but only a mixed enrichment for G/A at position 2 was observable at longer induction times (**FIG. 38A**). Finally, at 1 h induction, NRTH enriched strongly for G and T at positions 2 and 3, respectively (**FIG. 38A**), but by 3 h we observed a shift in the nucleotide preference at position 2 to a mix of G and A, suggesting that this variant recognizes and cleaves NAT PAMs more slowly when compared to NGT PAMs. Additionally, this suggests that NRTH may preferentially recognize NRT over NGG PAMs.

[00467] Interestingly, SpCas9-NG displayed a moderate preference for G at the 3rd and 4th PAM position at short induction times. This is consistent with SpCas9-NG's T1337R mutation, which is also found in SpCas9 VRER and VRQR [REF] and is the cause for the increased specificity for G at the 4th PAM position of these variants. Similar to the evolved Cas9 variants, SpCas9-NG's PAM sequence requirements also became more relaxed with longer induction times (data not shown).

[00468] Further, the P11 clone, which also possesses the P4.2.72.4 spCas9 mutations, was evolved using split-intein Cas9 mutants on AAA PAM bacterial depletion to generate clones with new mutations (**FIG. 21**). The ability of the newly P11-SacB-1 and P11-SacB-2 clones to perform base-editing and generate indels was evaluated *in vitro* in HEK293T cells (**FIGs. 22-23**). Both the P11-SacB-1 and P11-SacB-2 clones had higher base editing activity and a greater percentage of indels generated compared to xCas9 proteins (**FIGs. 22-23**).

[00469] Similarly, the P12 clone was evolved using split-intein Cas9 mutants on AAT or TAT PAM bacterial depletion to generate clones with new mutations (**FIGs. 24A-24B**). The ability of these newly-generated P12.3.b9-8 and P12.3.b10 clones to perform base-editing and generate indels was evaluated *in vitro* in HEK293T cells (**FIGs. 25A, 25B, 26A, 26B**).

EXAMPLE 7

Survival-based selection for isolating nuclease-active Cas9 variants

[00470] A survival-based selection method for isolating nuclease-active SpCas9 clones was generated (**FIG. 28**). The SacB gene produces a toxic protein, and clones that survive this selection will have active nuclease that can cut the SacB gene. The original TAT clone was generated from PANCE on a TAT PAM, but lacked nuclease activity. This TAT cloned was subcloned from a pool of N4.TAT selection phage (SP) into a Cas9 plasmid and selection was performed for variants that cut a SacB selection plasmid with a TAT PAM. Two additional TAT clones, SacB-TAT-1 and SacB-TAT-2, were isolated (**FIGs. 29A, 29B**).

[00471] These SacB-TAT-1 and SacB-TAT-2 clones were evaluated for their ability to perform base editing and generating indels *in vitro* in HEK293T cells (**FIGs. 30A, 30B, 31**). The SacB-TAT-1 and SacB-TAT-2 clones both possessed higher base editing activity on GAT, CAT, and GAAP AMs compared with xCas9 (**FIG. 30A**), as well as higher indel generation on GAT and TAT PAMs compared with xCas9 and spCas9 (**FIGs. 30B, 31**).

EXAMPLE 8

Evolved Cas9 to generate indels at endogenous human genomic loci

[00472] The activity of the evolved SpCas9 and xCas9 variant proteins was assessed in HEK293T cells through indel formation at endogenous target sites spanning all 64 NANN PAMs. For comparison, the activity of the SpCas9 wild-type (SpCas9-NG) protein was tested at these sites in parallel. Generally, each of the variants displayed the highest indel formation activity on target sites containing a PAM it was evolved to recognize, with NRRH and NRTH showing an average of 23.0±7.8% and 22.9±7.2% indel formation on target sites containing NAAN and NATN PAMs, respectively. Sites containing NACN PAMs were edited at slightly lower efficiencies, with NRCH averaging 18.0±5.9% indel formation. Additionally, NRRH displayed 20% average indel formation on sites containing a NAG PAM, even though it had not been evolved to bind this PAM sequence (**FIG. 38B**).

[00473] Interestingly, indel formation was observed with SpCas9-NG at a number of NANN sites. Although its average indel formation across these sites was lower than the evolved variants, SpCas9-NG displayed activity at sites with NANG PAMs (12.2±3.0%, 11.9±5.2%, 21.2±6.2%, and 18.3±4.4% average indel formation for NAAG, NACG, NATG, and NAGG, respectively) (**FIG. 38B**). In contrast, the evolved variants showed the lowest average activity at sites with PAM sequences with a G at position 4, and the highest at sites with a non-G (H) at this position (27.3±8.6%, 23.7±6.8, 26.9±8.1%, and 26.8±7.6% average indel formation for NRRH, NRCH, NRTH, and NRRH on NAAH, NACH, NATH, and NAGH PAMs, respectively) (**FIGs. 38B, 38C**). These results are consistent with the sequence preferences predicted by the bacterial PAM depletion experiments, and suggest that the variants and SpCas9-NG exhibit orthogonal PAM specificities.

[00474] The indel formation activity of evolved variants and SpCas9-NG were tested on a number of endogenous target sites containing NGN PAMs, with SpCas9-NG, NRCH, and NRTH performing best on NGA, NGC, and NGT PAMs, respectively, with 41.1±10.7%, 42.4±4.4%, and 67.7±6.8% average indel formation (data not shown). Similar to above, a preference for H at position 4 of the PAM by our variants was observed in these experiments.

[00475] Thus, increasing the DNA targeting capabilities of SpCas9 and xCas9 variants towards NRN PAMs could also greatly increase the proportion of genomic off-target sequences accessible by these Cas9 variants.

EXAMPLE 9

Evolved Cas9s are compatible with base editing technology

[00476] Next, the ability of evolved Cas9 variant proteins to support base editing was determined. C to T base editors (CBEs) were generated by incorporating the evolved Cas9 variants into BE4max (REF) in place of wt-Cas9. The activity of these CBEs was analyzed at the same 64 endogenous examined above for indel formation. As before, each of the three variants showed the highest average activity on sites containing the PAM it was evolved to recognize. BE4max-NRRH

and BE4max-NRTH performed best on NAAN and NATN PAMs, with an average of 11.7±3.7% and 17.3±4.0% C•G to T•A conversion, respectively. CBE activity on NACN PAMs was slightly less efficient, with BE4max-NRCH enabling the highest editing activity at these sites at an average of 10.8±3.0% base conversion. Both BE4max-NRRH and BE4max-NG edit NAGN sites similarly, at 11.4±3.6 and 11.6±4.8% average base conversion (**FIG. 39A**).

[00477] Improved base editing activity was again observed on sites with NANH PAMS, where C•G to T•A conversion at NAAH, NACH, NATH, and NAGH sites increasing to 14.4±4.1%, 13.0±2.6%, 21.0±4.2%, and 14.5±4.0 for BE4max-NRRH, -NRCH, -NRTH, and -NRRH, respectively (**FIGs. 39A, 39B**). BE4max-NG performs well at sites containing NANG PAMs, with 13.6±4.4% average editing (**FIG. 39A**). These editors also function on sites with NGN PAMs (data not shown). As expected, the CBE activity across all 64 sites is much more variable than that of indel formation, since there are increased requirements for efficient base editing such as sequence context and position of the C within the window. Finally, the Cas9 variants are also compatible with A to T base editors, exhibiting similar performance on a subset of sites containing NAN and NGN PAMs when substituted in place of wt-Cas9 in ABEmax (**FIG. 39C**).

EXAMPLE 10

Characterization of evolved Cas9s and SpCas9-NG using a mammalian library for base editing activity

[00478] Finally, to thoroughly profile the PAM preferences of these variants, the base editing efficiencies of the three evolved variants, SpCas9-NG, and wt-Cas9 were evaluated on a library of 11,776 unique sequences in mammalian cells. This library was designed using 46 distinct protospacers derived from sequences found in the human genome, each with different sequence contexts surrounding a fixed C in the 4th position. Each protospacer is adjacent to a PAM sequence of 4Ns, and is additionally flanked with designated primer binding sites for amplification for high-throughput sequencing (HTS) analysis (**FIG. 40B**).

[00479] Characterization of the evolved variants in this library format recapitulated the same preferences observed with both bacterial PAM depletion and base editing on endogenous mammalian genomic sites. For instance, our evolved variants exhibited the highest editing activity on the third base towards which it was evolved (**FIG. 40E**) or when a non-G was at the 4th position of the PAM, performing best when a pyrimidine was at this position (**FIG. 40F**). Additionally, our evolved variants, in particular NRRH, performed best when a G or C was present at position 1 of the PAM, whereas wt-Cas9 exhibited only slight preference for G at this position (data not shown).

[00480] The U6 promoter, commonly used to express sgRNAs in mammalian cells, initiates transcription with a 5' G. If a G is not natively present at the 5' end of the protospacer, guide sequences are typically either extended to the next native G or transcribed with a mismatched G at position 21 of the guide sequence. However, high-fidelity (HF) Cas9s, which are less tolerant of mismatches between the protospacer and sgRNA, exhibit decreased efficiency when using a 21 nucleotide (nt) with a mismatched 5' G [REF]. Because PACE has previously led to Cas9s with HF properties, including sgRNA mismatch intolerance [REF], we sought to determine if our new variants shared the same characteristics.

[00481] The average base editing activity of the evolved variants was evaluated across all sites containing either a 20 nt protospacer with a matched 5' G, a 21 nt protospacer with a matched 5' G, or a 21 nt protospacer with a mismatched 5'G. Both the evolved variants and wt-Cas9 showed the highest base editing activity with a 20 nt protospacer and a matched 5' G. When examining all NNNN PAMs, both the variants and wt-Cas9 showed a significant decrease in base editing efficiency when the protospacer was increased to 21 nt, regardless if the 5' G was matched with the target sequence (**FIG. 40C**). The magnitude of this decrease was greater for the evolved variants when compared to wt-Cas9. Interestingly, the deleterious effect of using a 21 nt protospacer on editing efficiency is ameliorated when targeting sites with a NGNN PAM (data not shown), and almost completely absent when targeting sites with a NGGN PAM (**FIG. 40D**). This is especially true for wt-Cas9, which shows

no significantly decreased base editing activity on sites with a 21 nt protospacer when the PAM is NGG.

EXAMPLE 11

Evolved Cas9s correct disease-associated SNPs by accessing non-G PAMs

[00482] To demonstrate the utility of the evolved variants in a disease-relevant context, the Glu to Val point mutation at position 6 of the sickle-hemoglobin (HbS) variant of β-globin, which is causative of red blood cell sickling in sickle-cell anemia, was targeted [REF]. The HbS mutation arises from a GAG to GTG codon change, which cannot be fully reverted through current base editing technologies. However, this SNP can be partially corrected with ABE to a GCG (Ala) through A·T to G·C conversion on the opposite strand. This genotype, known as the Makassar mutation, has been shown to result in phenotypically normal hemoglobin.

[00483] Unfortunately, the only NGG or NGN PAMs available at this site place the target A at either position 2 or 9, respectively, which fall outside the optimal editing window for ABE. However, two alternative target protospacer sequences that fall adjacent to a CAT or CAC PAM place the target A at either position 4 or 7, respectively, with an off-target A present at either position 6 or 9 leading to a silent CCT to CCC (Pro to Pro) mutation. Thus, the ability of the evolved variants, along with SpCas9-NG, to convert the sickle-cell SNP to the Makassar mutation using these alternative sites with non-G PAMs was evaluated.

[00484] In experiments using HEK293T cells engineered with a GAG to GTG mutation at codon 6 of β-globin, while the evolved variants supported considerable A to G conversion at both sites, SpCas9-NG edited efficiently only using the protospacer sequence containing a CAT PAM. This is perhaps due to the presence of a G at the 4th position of this PAM sequence (**FIGs. 41B, 41C**), which appears to improve SpCas9-NG's recognition of NAN PAMs (see above). Unfortunately, editing using the CAT PAM protospacer occurred primarily at the off-target base (position 6), with the target A (position 4) showing less than 10% conversion across all editors (**FIG. 41C**). Base

conversion using the CAC PAM protospacer, however, was much more efficient. As expected, ABEmax-NRCH showed the highest editing activity, with 40.6±6.5% base conversion at the target A (position 7) and 13.0±5.6% at the off-target A (position 9).

[00485] ABEmax-NRRH and -NRTH were also able to achieve 28.9±7.4% and 14.1±4.8% conversion, respectively. The high activity of all three evolved variants at this site likely stems from the presence of a C at the 4th position of the CAC PAM sequence. In comparison, ABEmax-NG showed negligible (1.0±0.8%) base conversion activity at this site (**FIG. 41B**). Collectively, these results suggest that both the evolved variants and SpCas9-NG have the potential to edit disease relevant SNPs using non-G PAMs, and furthermore highlight the utility of targeting a SNP using multiple protospacer/PAM sequences.

[00486] Together with SpCas9-NG, the evolved variants NRRH, NRCH, and NRTH should expand the targeting scope of SpCas9 to sites with NR PAMs, increasing the number of pathogenic SNPs correctable by either CBE or ABE. Based on analysis of the ClinVar database, 95.0% of pathogenic SNPs correctable through a C·G to T·A conversion and 94.7% of pathogenic SNPs correctable through an A·T to G·C conversion can be targeting using an NR PAM. Additionally, expansion to NR PAMs increases the number of possible protospacers available for targeting a given SNP for correction with base editors: on average, there are XX protospacers per disease SNP targetable with CBE and XX protospacers for those targetable with ABE with NR PAMs, compared to XX targetable with CBE and XX targetable with ABE, respectively, when using NG PAMs.

EXAMPLE 12

[00487] Characterizing Mutants that Work on NRRH, NRCH, and NRTH PAMs

[00488] SpCas9 mutant proteins were identified that work best on NRRH, NRCH, and NRTH PAMs. The SpCas9 mutant protein that works best on NARH ("es" variant), has an amino acid sequence as presented in SEQ ID NO: 22 (underligned residues are mutated from SpCas9)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY

HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
IIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGFNSPTAAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIGFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLHKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGVPAAFKYFDTTIDKKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD (SEQ ID NO: 22).

[00489] The SpCas9 mutant protein that works best on NRCH (“fn” variant), has an amino acid sequence as presented in SEQ ID NO: 23 (underligned residues are mutated from SpCas9)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
IIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR

QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTINRKQYNTTKEVLDATLIRQSITGLYETRI
DLSQLGGD (SEQ ID NO: 23)

[00490] The SpCas9 mutant protein that works best on NRTH (“ax” variant), has an amino acid sequence as presented in SEQ ID NO: 24 (underligned residues are mutated from SpCas9)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY
HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYD
DDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLKALVR
QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
IIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW
NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ
KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEEN
EDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTIL
DFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE
VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK
MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLS
MPQVNIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKK
LKSVKELLGITIMERSSFEKNPIGFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASASVLHKGN
ELALPSKYVNFLYLASHYEKLKGSSEDNKQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS
AYNKHRDKPIREQAENIIHLFTLTNLGASAAFKYFDTTIGRKLYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD (SEQ ID NO: 24)

[00491] The base-editing activity of the ax, es, fn, and SpCas9 (“NG”) proteins was characterized *in vitro* in HEK293T cells on NAA, NAC, NAT, and NAG PAMs (**FIGs. 33A-33D; 34A-34B**). The es protein had increased activity on CAAA, CAAC, AAAT, and GAAC PAMs, the

fn protein had increased activity on AACC, AACT, TACT, TACC, CACT, and CACC PAMs, the ax protein had increased activity on AATA, TATT, TATA, TATC, CATA, CATT, CATC, GATA, GATT, and GATC PAMS compared with other SpCas9 proteins (**FIGs. 33A-33C; 34A-34B**).

[00492] The A to G base editing activity of es and fn SpCas9 proteins were also characterized *in vitro* in HEK293T cells on NAA, NGA, NAC, and NGC PAMs (**FIGs. 35A-35C**). The es, fn, or wild-type SpCas9 proteins were incorporated into the ABEMAX A to G gene editing fusion protein. The es protein had increased base-editing activity on AAAT, CAAC, GAAC, AACC, TACT, TACC, CACT, CACC, AGCC, AAGA, and AAGC PAMs compared with NG SpCas9 protein (**FIGs. 35A, 35B**). The fn protein had increased base-editing activity on GGGT and TGGC compared with NG SpCas9 protein (**FIG. 35C**).

EXAMPLE 13

[00493] Continuous evolution of SpCas9 variants compatible with non-G PAMs

[00494] Streptococcus pyogenes Cas9 (SpCas9) is a widely used genome editing tool, but can only access a small fraction of DNA sites due to its requirement for an NGG protospaceradjacent motf (PAM). This limits SpCas9's utility for precision genome editing applications such as base editing (Rees and Liu, 2018), homology-directed repair (Paquet et al., 2016), and predictable template-free end-joining repair (Shen et al., 2018). While SpCas9 variants with alternative PAM requirements have been reported, their targeting scope remains primarily restricted to PAMs containing G. Here, we report the laboratory evolution of three new SpCas9 variants collectively capable of recognizing NRNH PAMs (where R = A or G and H = A, C, or T) using an improved phage-assisted continuous evolution (PACE) selection for DNA binding. We show that these variants recognize NAAH, NACH, NATH, and NAGH PAMs to effect indel formation, cytosine base editing, and adenine base editing using a panel of 64 endogenous human genome target sites containing all NANN PAMs. Additionally, we profile the editing efficiencies of our evolved

SpCas9s and the previously-reported SpCas9-NG as base editors on a 11,776-member genomically integrated protospacer/sgRNA pair library spanning all NNNN PAMs in HEK293T cells to provide an exhaustive characterization of their PAM preferences in a human cell setting. Finally, we demonstrate the ability of our variants to enable A•T-to- G•C base editing of the founder sickle-cell anemia mutation of β-globin using a previously inaccessible CAC PAM. Together with previously reported SpCas9 mutants, these newly evolved variants expand the targeting scope of SpCas9 to include a majority of NR PAM sequences, greatly increasing the fraction of genomes accessible to Cas9-mediated genome editing.

[00495] The CRISPR-Cas9 system, originally evolved as a mechanism for adaptive immunity in bacteria, has in recent years transformed the life sciences by enabling a wide range of techniques for targeted genome manipulation including gene disruption, homologydirected repair, gene regulation, and base editing (Komor et al., 2017). The applicability of these techniques is limited by the requirement of Cas9 for a protospacer-adjacent motif (PAM) in order to bind a DNA sequence. For example, wild-type Streptococcus pyogenes Cas9 (SpCas9), the most widely-used and well-characterized Cas9 homolog (Komor et al., 2017), recognizes an NGG PAM immediately 3' of the target DNA sequence, and with rare exception will not efficiently engage DNA sequences lacking an NGG PAM (Jinek et al., 2012). To address this limitation and expand the range of targetable genomic loci, researchers have used naturally occurring Cas9 orthologs with different PAM specificities (Cebrian-Serrano and Davies, 2017). The majority of these natural Cas9 variants, however, are less well-characterized, less active in a variety of conditions, and/or more stringent in their PAM requirements than SpCas9.

[00496] Motivated by the limited set of natural Cas9 homologs that have successfully been used for genome editing, researchers have engineered or evolved both Staphylococcus aureus Cas9 (SaCas9) (Kleinstiver et al., 2015a) and SpCas9 (Hu et al., 2018; Kleinstiver et al., 2015b; Nishimasu et al., 2018) to increase their PAM targeting scope. These efforts have led to an expansion of SpCas9's

potential PAM compatibility from NGG to most NG sites (Hu et al., 2018; Nishimasu et al., 2018). However, despite this substantial increase in SpCas9's DNA targeting capability, non G-rich locations in the genome remain difficult to access, despite their abundance. The restriction on Cas9 targeting is especially problematic when using precision genome editing techniques which require strict placement of the Cas9 in relation to the desired genomic edit, such as homology-directed repair (HDR) (Paquet et al., 2016), predictable template-free end-joining (Shen et al., 2018), and base editing (Rees and Liu, 2018).

[00497] Base editing is a widely used genome editing technology in which a target base is directly converted to another base through deamination of cytosine to uracil (cytosine base editor, CBE) (Komor et al., 2016), or adenine to inosine (adenine base editor, ABE) (Gaudelli et al., 2017) by a Cas9-directed deaminase, ultimately resulting in a C•G-to- T•A, or A•T-to-G•C conversion, respectively. This technology is particularly sensitive to Cas9 positioning: activity for SpCas9-derived editors, for example, is optimal when the PAM is located approximately 13-17 nt away from the target base (Rees and Liu, 2018). In addition, for any given base edit, it may be desirable to screen multiple target sequence windows to maximize on-target activity while minimizing editing of other bases (Jin et al., 2019; Lee et al., 2018a; Xin et al., 2019; Zuo et al., 2019). Taken together, these requirements highlight the major ongoing need to access additional PAM sequences. Here we report the directed evolution of three new SpCas9 variants capable of recognizing NRRH, NRTH, and NRCH PAMs, respectively, where R = A or G, and H = A, C, or T. These variants were evolved through improved phage-assisted continuous evolution (PACE) selections for SpCas9 binding to specific sequences with non-NGG PAMs. We extensively characterized these three new variants, as well as SpCas9-NG (Nishimasu et al., 2018), a previously-reported engineered SpCas9 that recognizes NG PAMs, on 64 endogenous human genomic target sites, as well as a library of 11,776 integrated target sites. The new variants reported here, together with previously reported NG-

compatible Cas9 variants, expand the potentially accessible PAM sequence space of SpCas9 to cover the vast majority of NR sequences.

Results

Initial evolution of SpCas9 toward non-G PAM sequences

[00498] Phage-assisted continuous evolution (PACE), a method for the rapid directed evolution of biomolecules, has been used to evolve a wide range of proteins including RNA polymerases (Carlson et al., 2014; Dickinson et al., 2013; Esvelt et al., 2011; Pu et al., 2017), proteases (Dickinson et al., 2014; Packer et al., 2017), antibody-like proteins (Badran et al., 2016; Wang et al., 2018), insecticidal proteins (Badran et al., 2016), metabolic enzymes (Roth et al., 2019), aminoacyl-tRNA synthetases (Bryson et al., 2017), and DNA-binding proteins (Hu et al., 2018; Hubbard et al., 2015). In PACE, a population of bacteriophage (selection phage, SP) is continuously diluted by E. coli host cells (Esvelt et al., 2011). These SP lack gene III (gIII), which encodes the coat protein pIII that is essential for phage infectivity, and instead express the protein to be evolved.

[00499] SP carrying protein variants with desired activity are able to trigger the production of pIII from an accessory plasmid (AP) in the host cells, thus generating infectious progeny and allowing the SP population to persist despite continuous dilution. Conversely, SP encoding inactive variants cannot trigger pIII production, and produce non-infectious progeny that are rapidly diluted out of the system. The SP genome is continuously mutagenized by a mutagenesis plasmid (MP), thus generating diversity in the evolving protein of interest.

[00500] PACE was used to evolve SpCas9 variants with broadened PAM compatibility by linking PAM recognition to SP propagation through a bacterial one-hybrid protein:DNA binding selection (Hu et al., 2018). In this selection system, binding of a nuclease-inactive dSpCas9 variant fused to the *E. coli* RNA polymerase omega subunit (ώ–dSpCas9) to a target protospacer-PAM sequence recruits E. coli RNA polymerase to drive gIII transcription from an adjacent s70 promoter (FIG. 36 (A)). Only SP carrying w– dSpCas9 variants capable of binding to the target PAM sequence will produce

infectious progeny phage and replicate during PACE (Hu et al., 2018). Evolving SpCas9 against a mixture of all possible NNN or HHH (H = non-G) PAMs using this selection led to xCas9, which can bind some NG PAMs, but very few non-G PAMs (Hu et al., 2018). We hypothesized that during the evolution of xCas9, the use of a complex mixture of many PAMs reduced the selection pressure for binding activity on any specific PAM. Therefore, we reasoned that selecting for binding to specific PAM sequences in parallel PACE experiments might result in SpCas9 variants with better recognition of non-canonical PAMs.

[00501] To determine which non-G PAMs might be accessible upon extensive SpCas9 evolution, we performed phage propagation assays, which serve as a proxy for a protein's activity on a defined target, of SP encoding either SpCas9 or xCas9 on host cells containing APs spanning all 64 NNN PAM sequences (FIG. 36(B)). While SpCas9 and xCas9 demonstrated phage propagation activity on many G-containing PAMs, SP encoding xCas9 and, to a more limited extent, SpCas9, also showed modest propagation on host cells containing NAN PAM APs (FIG. 36(B)). Thus, we decided to focus our evolution efforts on the NAN subset of PAM sequence space.

[00502] We began by using phage-assisted non-continuous evolution (PANCE) (Roth et al., 2019; Suzuki et al., 2017), in which SP are iteratively passaged through serial dilution in plate wells containing host cells, to evolve either SpCas9 or xCas9 for binding to each of the 16 possible NAN PAM target sequences in parallel (FIG. 36(C)). While slower than PACE, PANCE is less stringent, enabling weakly active variants to replicate (Roth et al., 2019) and can be performed in higher throughput, allowing us to evolve simultaneously towards many different targets. After performing 19 rounds of serial dilution in PANCE (total net phage replication of ~1038-fold) on each of the 16 NAN PAM variants in parallel, we observed mutations largely differing according to the 3rd base of the NAN PAM targeted for evolution (FIG. 36(D)). For example, variants selected on NAA enriched Gly, Ile, or Lys at position 1333, while those selected on NAT enriched Gln or Leu at position 1335.

[00503] Finally, variants evolved to bind NAC simultaneously acquired Gln at position 1335 and Asn at position 1337. Given this early divergence, we decided to divide the evolution of these SpCas9 variants into three separate non-G PAM trajectories: HAA, HAT, and HAC. Because our goal was to evolve SpCas9 to recognize non-G PAMs, we chose to exclude NAG from our targets; additionally, NG-targeting SpCas9 variants have been reported (Hu et al., 2018; Nishimasu et al., 2018), which in theory should allow targeting of sites with NAG PAMs by simply shifting the protospacer sequence by a single nt in the 3' direction.

New Cas9 PACE selections enable evolution of NAA PAM binding activity

[00504] The NAA PAM trajectory was initially focused on. Despite enriching for multiple consensus mutations in the PAM-interacting domain (PID; residues 1099-1368) (D1135N, E1219V, Q1221H, H1264Y, A1320V, R1333K), our NAA-targeted PANCE evolved variants exhibited low base editing activity when subcloned into C to T base editors (CBEs) and tested on sites containing NAA PAMs in mammalian cells (clone GAA.N1-4; FIG. 37C). We hypothesized that evolving increased binding activity might benefit editing efficiencies, and implemented three strategies to increase selection stringency.

[00505] First, we required that the evolving SpCas9 also bind a second, distinct protospacer by using a dual-AP system. In this system, each AP provides one half of split-intein pIII (Wang et al., 2018) under control of the Cas9 1-hybrid circuit. Binding of the SpCas9 variant to both sites produces both pIII-intein halves, which must be coexpressed to splice and generate functional full-length pIII (FIG. 37A). We chose two variants evolved in PANCE (GAA.N1-2 and GAA.N1-4; FIG. 37D and 37B) and subjected them to PACE using this dual-AP system. These experiments, which also targeted a CAA PAM, lead to the acquisition of five additional consensus mutations (A10T, I322V, S409I, E427G and G715C; FIG. 44B), which together in clone CAA.P1-1 improved CBE activity on sites with NAA PAMs in mammalian cells 4.2-fold on average when compared to PANCE evolved variant GAA.N1-4 (FIG. 37C).

[00506] Second, we reasoned that production of large amounts of w–dSpCas9 by the SP might saturate binding to protospacer-PAM sites even if the affinity of the SpCas9 variant for that PAM was modest. Indeed, previous reports have shown that using higher concentrations of SpCas9 can lead to recognition of non-canonical PAM sequences (Karvelis et al., 2015), despite modest binding of these sequences by SpCas9. Unfortunately, as both the promoter and ribosome-binding site for w–dSpCas9 are encoded on the SP, the total amount of w–dSpCas9 produced is subject to selection in PACE and thus falls outside of experimenter control.

[00507] Therefore, we sought to limit the total amount of SpCas9 present in the selection by using a split-intein to divide w– dSpCas9. Here, only the C-terminal segment of dSpCas9 (residues 574-1368) fused to NpuC (dSpCas9C) is encoded on the evolving SP, and the w–N-terminal portion (residues 1-573) fused to NpuN (w–dSpCas9N) is provided on an immutable complementary plasmid (CP) in the host cells (FIG. 37A and 43B). This strategy (hereafter, "split-SpCas9") allows the total amount of full-length SpCas9 produced in the host cells in PACE to be limited by the expression level of w–dSpCas9N from the CP.

[00508] We subcloned the mutations obtained from clone CAA.P1-1 (FIG. 37B), evolved using the dual-AP selection, into the split-SpCas9 format. Four mutations (T10A, I322V, S409I, and E427G) had accumulated in residues 1-573 of this clone. To investigate their effect, we compared the activity of w–dSpCas9N with that of w–dSpCas9N(T10A I322V S409I E427G) (hereafter referred to as w–dSpCas9N-mut) in overnight phage propagation assays using phage encoding dSpCas9C derived from CAA.P1-1. We observed greater phage propagation on host cells with a CP encoding w–dSpCas9N-mut(FIG. 43D), suggesting that these four mutations might have a beneficial effect on SpCas9 binding. Therefore, we used w–dSpCas9N-mut in the CP supporting all subsequent evolution efforts.

[00509] We subjected our evolved CAA.P1-1 dSpCas9C to two subsequent PACE campaigns (8 and 3 days, respectively, at average flow rates of 1.3 V/h) using host cells harboring an AP containing an AAA or CAA PAM target site and CPs providing successively decreasing amounts of w–dSpCas9N-

mut (see Methods for details). These rounds lead to the accumulation of additional mutations in the PID, including D1180G, which was present in several sequenced clones (CAA.P2-2, AAA.P3-1, CAA.P3-1,2; FIG. 37B).

[00510] Among 10 surviving clones randomly chosen for sequencing, we observed 7-17 nonsilent mutations per clone (FIG. 37B). From these, the SpCas9 variant CAA.P2-2 exhibited a large increase in mammalian cell base editing activity, with more than double the activity of our previous variants on most NAA sites tested (FIG. 37C). Third, to further increase selection stringency, we removed gene VI (gVI), which is essential for phage propagation, (Brödel et al., 2016) from the SP for use as a second selection marker (in addition to gIII) in PACE. This strategy allowed us to combine both selection modifications described above by requiring a split-dSpCas9 to bind each of two distinct protospacers in order to express both gIII and gVI (FIG. 37A).

[00511] Thus, three dSpCas9C clones from our previous evolutions (CAA.P2-1, CAA.P2-2, and CAA.P3-1) were subjected to this highest-stringency selection in PACE, resulting in additional mutations in the PID. Most notably, R1114G and L1318S were both highly enriched among sequenced surviving variants, which on average contained 20 non-silent mutations relative to SpCas9 (TAA.P4; FIG. 37B). When tested in mammalian cell CBE experiments, these variants showed little editing activity (FIG. 37C). We theorized that the large number of mutations present in these highly evolved variants, especially those outside of the PID, might prove deleterious to expression and/or inactivate either nuclease domain. To address this possibility, we performed DNA shuffling of the C-terminal portion (residues 574-1368) of the pool of variants from this final evolution with wild-type SpCas9(574-1368), allowing deleterious mutations to exit while shuffling mutations between the pool members, and re-subjected the resulting library to our most stringent binding selection. This "backcrossing" process led to the isolation of clone TAA.P4s-4 (R1114G, D1135N, V1139A, D1180G, E1219V, Q1221H, A1320V, R1333K) (FIG. 37B and 37D), which demonstrated a 1.2- fold

increase relative to the previous best PACE mutant across all HAA sites tested amongst our variants (FIG. 37C).

Evolution of SpCas9 variants that recognize NAT or NAC PAM sequences

[00512] Based on the outcomes of the NAA PAM evolution campaigns, we approached the evolution of SpCas9 variants capable of recognizing NAT and NAC PAM sites in a fashion that avoids potentially deleterious bystander mutations. To ensure that we started with nuclease-active variants, we developed a modified version of a previously reported (Kleinstiver et al., 2015b) bacterial DNA cleavage selection (FIG. 43E and 43F). In this nuclease selection, SpCas9 variants are challenged for their ability to cleave a protospacer-PAM sequence on a high-copy plasmid that also encodes a conditionally toxic gene (sacB). The surviving cells encode nuclease-active SpCas9 variants that cleave the target sequence, destroying the toxic plasmid.

[00513] Thus, we converted the dSpCas9 clones from the NAT or NAC PANCE pools into nuclease-active forms by restoring Asp 10 and His 840, then passed the resulting libraries through the nuclease selection using a TAT or CAC PAM, respectively. From this, we isolated two clones (SacB.TAT-1 and -2; FIG. 37E) that exhibited DNA cleavage activity on the TAT PAM with PID consensus mutations of D1135N, E1219V, Q1221H, P1321S, and R1335L, and a third clone that cleaved a CAC PAM with PID mutations N1135D, E1219V, D1332N, R1335Q, and T1337N (SacB.CAC; FIG. 37F). We evolved these nuclease-active TAT and CAC variants in split-dSpCas9 PACE using host cells encoding APs with either NAT (AAT or TAT) PAMs or NAC (AAC, TAC, or CAC) PAMs, respectively. These experiments resulted in the enrichment of several additional PID mutations, including R1114G, which arose independently in all three trajectories (NAA, NAT, and NAC) (Figures 37B, 37E, and 37F), suggesting that this mutation may be generally beneficial for modifying PAM recognition by the PID in a manner compatible with NA PAMs.

[00514] Next, we removed gVI from these evolved SP pools and subjected them to additional selection in split-dSpCas9 PACE using the dual-AP system (FIG. 37A and FIG. 43C). Both

protospacers contained either an AAT or TAC PAM for evolution following the NAT or NAC trajectory, respectively. Increasing stringency for the NAT-targeting SpCas9 did not improve activity despite enrichment of several mutations (TAT.P6; FIG. 37D and 44A). We therefore selected the most active NAT PAM-targeting variant from the split-dSpCas9 evolution (TAT.P5-1; Figure 37D) to move forward with. This variant contained the 11 PID mutations R1114G, D1135N, D1180G, G1218S, E1219V, Q1221H, P1249S, E1253K, P1321S, D1332G, R1335L (Figures 37E and 37G). PACE of NAC-targeting splitdSpCas9 using dual protospacers and a TAC PAM also enriched for several mutations (TAC.P9; Figure 37G). We shuffled residues 574-1368 of the surviving clones with that of SpCas9 and re-challenged the resulting library with our most stringent binding selection (TAC.P9s; Figure 37G). From the surviving SP pool, we isolated clone TAC.P9s-3 with the PID mutations R1114G, D1135N, E1219V, D1332N, R1335Q, T1337N, S1338T, and H1249R (Figures 37F and 37H).

Mutations outside of the PID

[00515] Structural studies of SpCas9 suggest that residues in the PID mediate PAM specificity (Anders et al., 2014). Indeed, most of the previous efforts to engineer or evolve SpCas9 to accept alternative PAMs have focused on modulating this region of the protein (Kleinstiver et al., 2015b; Nishimasu et al., 2018). However, because our PANCE and PACE experiments allowed mutation of either the entire SpCas9 sequence or residues 574-1368 (in the case of split-intein w–dSpCas9), we observed the enrichment of many where from 3 to 15 mutations outside of the PID. Because many of these mutations fell within the RuvC or HNH nuclease domains, we anticipated that some would negatively impact SpCas9 nuclease activity (Jiang and Doudna, 2017). However, other mutations in the helical domain consistently enriched across several independent evolving populations, suggesting that they may confer a beneficial effect on SpCas9 DNA binding/unwinding. To minimize the deleterious effects from bystander mutation accumulation in the nuclease domains while preserving beneficial mutations in the helical domain, we decided to transplant our evolved PIDs onto a fixed N-

terminal sequence that included the mutations T10A, I322V, S409I, E427G that we found to improve phage propagation in the split-dSpCas9 selection (FIG. 43D), as well as R654L and R753G, which consistently enriched across multiple independent PACE experiments (Figure 44B).

[00516] The addition of all six NTD mutations to CBEs containing the PIDs of NAA variant TAA.P4s-4 and NAC variant TAC.P9s-3 improved CBE activity in mammalian cells across several sites when compared to SpCas9 variants containing only the evolved PID mutations (Figure 44C). A smaller benefit was observed when the NTD mutations were added to the PID mutations of NAT variant TAT.P5-1 (Figure 44C). We incorporated these six N-terminal mutations into all three final variants, hereafter referred to as SpCas9-NRRH, SpCas9-NRTH, and SpCas9-NRCH, which are the addition of T10A, I322V, S409I, E427G, R654L, and R753G to the evolved PID domains from TAA.P4s-4, TAT.P5-1, and TAC.P9s-3, respectively.

Characterization of PAM specificity through bacterial depletion

[00517] To better characterize the PAM specificities of our evolved variants as nucleases, we performed bacterial PAM depletion using a NNNN PAM library (Kleinstiver et al., 2015b). For comparison, we also performed depletion experiments with SpCas9-NG in parallel. Cells were plated after 1-hour, 3-hour, or overnight expression of the SpCas9 variant from an inducible promoter to assess kinetic differences in PAM sequence preference. Consistent with the eventual cleavage of even modestly recognized PAMs, depletion scores of any given PAM (defined as the frequency of the PAM in the input library divided by the frequency of the PAM post-selection) increased with longer induction times, with the shortest induction times resulting in the most noticeable sequence preferences (Figure 45A).

[00518] For example, at shorter induction times, SpCas9-NRRH exhibited a strong preference for C at the 4th PAM position, a mixed preference for purines at positions 2 and 3 and a moderate preference for G at position 1 (Figure 38A). However, longer induction times resulted in more relaxed preferences at all PAM positions. Similarly, SpCas9-NRCH showed a strong preference for G at

position 2 and a moderate preference for pyrimidines at position 4 (Figure 38A) at shorter inductions, but only a mixed enrichment for purines at position 2 was observable at longer induction times (Figures 38A and 45A). Finally, at short induction times, SpCas9-NRTH enriched strongly for G and T at positions 2 and 3, respectively (Figure 38A), but the nucleotide preference at position 2 shifted to a mix of G and A at longer timepoints, suggesting that this variant recognizes and cleaves NAT PAMs more slowly than NGT PAMs. These results also suggest that SpCas9-NRTH may preferentially recognize NGT over NGG PAMs, as the NGT PAMs were more strongly depleted than NGG PAMs (average depletion score of 1394 for NGT compared to 223 for NGG at 1 h induction).

[00519] Interestingly, SpCas9-NG displayed a moderate preference for G at the 3rd and 4th PAM position at short induction times. This finding is consistent with the T1337R mutation in SpCas9-NG, which is also found in SpCas9 VRER and VRQR (Kleinstiver et al., 2015b) and is the basis of the increased specificity for G at the 4th PAM position in these two variants (Anders et al., 2016; Hirano et al., 2016b; Kleinstiver et al., 2015b). Similar to the evolved SpCas9s described here, SpCas9-NG's PAM sequence requirements also became more relaxed with longer induction times (Figure 45A).

Evolved SpCas9 nucleases generate indels at endogenous human genomic loci

[00520] Next, we assessed the activity of our evolved variants in HEK293T cells through indel formation at 64 endogenous target sites spanning all possible NANN PAMs. For comparison, we also tested the activity of SpCas9-NG at these sites in parallel. Generally, each of our variants displayed the highest indel formation activity on target sites containing a PAM it was evolved to recognize, with SpCas9-NRRH and -NRTH showing an average of 23±4.5% and 23±4.1% indel formation on target sites containing NAAN and NATN PAMs, respectively. Sites containing NACN PAMs were edited at slightly lower efficiencies, with SpCas9-NRCH averaging 18±3.4% indel formation. Additionally, SpCas9-NRRH displayed 23±4.3% average indel formation on sites containing a NAG PAM, even though it had not been evolved to bind this PAM sequence (Figure 3B). Indel formation activity of xCas9 was also examined at a subset of NAN sites and found to be minimal (Figure 45B).

[00521] Interestingly, we also observed indel formation with SpCas9-NG at some NANN sites. Although its average indel formation across these sites was lower than our evolved variants, SpCas9-NG displayed activity at sites with NANG PAMs (NAAG: 12±1.7%, NACG: 14±3.0%, NATG: 23±3.6%, NAGG: 20±2.5% average indel formation) (Figure 38B). In contrast, our evolved variants showed the lowest average activity at sites with PAM sequences with a G at position 4, and the highest at sites with a non-G (H) at this position (27±5.0%, 27±4.7%, 24±3.9%, and 27±4.4% average indel formation for SpCas9-NRRH, -NRTH, -NRCH, and -NRRH on NAAH, NATH, NACH, and NAGH PAMs, respectively) (Figures 38B and 38C). These results are consistent with the sequence preferences predicted by our bacterial PAM depletion experiments and suggest that our variants and SpCas9-NG exhibit complementary PAM specificities, especially with respect to non-G versus G bases at the 4th position.

[00522] We also tested the indel formation activity of our evolved variants and SpCas9-NG on a number of endogenous target sites containing NGN, rather than NAN, PAMs. While treatment with SpCas9-NG gave rise to robust indel formation on most NGN PAMs examined (48±4.4%), SpCas9-NRTH and -NRCH showed slightly higher activity than SpCas9-NG at NGT and NGC PAMs, with 68±3.9% and 42±2.5% average indel formation, respectively (Figure 45C). Consistent with the PAM depletion assay results, a preference for H at position of the PAM was observed in these experiments for SpCas9-NRTH and -NRCH.

DNA specificity of evolved SpCas9 nucleases

[00523] As broadening the PAM targeting capabilities of various Cas9 has been shown to increase the proportion of genomic off-targets edits (Kleinstiver et al., 2015a; Nishimasu et al., 2018), we performed genome-wide, unbiased identification of double-strand breaks enabled by sequencing (GUIDE-seq) using SpCas9, SpCas9-NRRH, -NRCH, and –NRTH in U2OS cells (Tsai et al., 2015). For comparison, we also analyzed xCas9, which was previously shown to possess reduced off-target activity (Hu et al., 2018). These experiments showed that, when targeting the highly promiscuous

HEK site 4 (HEK4) (Tsai et al., 2015), our evolved variants displayed comparable or better on-target activity (8.8%, 22.5%, and 7.8% on-target reads of total reads for SpCas9-NRRH, -NRTH, and -NRCH, respectively) when compared to SpCas9 (5.1% total reads) (Figure 38D and 45D). This is similar to xCas9, which also exhibited improved on-target activity (12.7% total reads) relative to SpCas9 (Figure 38D and 45D) (Hu et al., 2018). Interestingly, our variants primarily displayed off-target activity at sites containing PAMs consistent with their evolved preferences. For example, the most prominent off-target for SpCas9-NRRH occurs at a site bearing a CAA PAM (10% total reads), SpCas9-NRTH at a GGT PAM (10.2% total reads), and SpCas9-NRCH at a TGC PAM (9.9% total reads) (Figure 45D).

[00524] Various off-targets were also observed at sites with NRN PAMs, such as GAA, GAT, and CAG, for these evolved SpCas9s (Figure 45D). Taken together, these results suggest that our evolved variants may have similar or increased DNA specificity compared to SpCas9 on sites with NGG PAMs, and due to their altered PAM specificities may access a different set of off-target sequences.

Evolved SpCas9s support cytosine and adenine base editing

[00525] Since expanding the targeting scope of base editing was a major motivation behind our efforts, next we determined the ability of our evolved SpCas9s to support both cytosine and adenine base editing. We generated CBEs by incorporating our evolved variants into BE4max (Koblan et al., 2018) (hereafter referred to as "BE4") in place of SpCas9 and tested their activity at the same 64 endogenous NANN PAM sites examined above for indel formation. As with their nuclease forms, each of the three evolved CBE variants showed the highest average activity on sites containing the PAM it was evolved to recognize. BE4-NRRH and BE4-NRTH performed best on NAAN and NATN PAMs with an average of 12±2.1% and 17±2.3% C•G to T•A conversion, respectively. CBE activity on NACN PAMs was slightly less efficient, with BE4-NRCH enabling the highest editing activity at these sites at an average of 11±1.7% base conversion. Both BE4-NRRH and BE4-NG (generated from

SpCas9-NG) edit NAGN sites similarly, at 12±2.8% and 11±2.1% average base conversion (Figure 39A).

[00526] Improved editing activity was again observed on sites with NANH PAMs, where C•G to T•A conversion at NAAH, NATH, NACH, and NAGH sites increasing to 14±2.4%, 21±2.5%, 13.0±2.0%, and 14±2.3 for BE4-NRRH, -NRTH, -NRCH, and -NRRH, respectively (Figure 39A and 39B) . BE4-NG performed well at sites containing NANG PAMs, with 14±1.3% average editing (Figure 39A). Average CBE editing efficiency across all 64 sites was lower than that of indel formation, likely due to increased requirements for efficient base editing such as sequence context and position of the C within the window.

[00527] These editors also function on sites with NGN PAMs, editing at 17±2.3%, 9.1±3.0%, 19±2.9% and 20±4.0% for BE4-NRRH, -NRTH, -NRCH, and -NG, respectively (Figure 46A). Finally, we also generated ABEmax (Koblan et al., 2018) variants (hereafter referred to as "ABE") from SpCas9-NRRH, -NRTH, -NRCH, and SpCas9-NG, and tested adenine base editing at 54 endogenous loci. We observed that the newly evolved variants are also compatible with adenine base editing, exhibiting similar performance on a subset of sites containing NAN and NGN PAMs as we observed for the corresponding CBEs and nucleases. For example, ABE-NRRH, -NRTH, -NRCH, and -NRRH edited most efficiently at NAAH, NATH, NACH, and NAGH PAMs, with 16±2.6%, 24±2.9%, 13±2.2%, and 26±3.5% base conversion (Figure 39C and 46B).

[00528] The scope of base editing is limited by the requirement that the target base be located within the canonical CBE or ABE editing window (approximately protospacer positions 4-8, counting the PAM as positions 21-23). The evolved variants SpCas9- NRRH, -NRCH, and -NRTH, together with SpCas9-NG and xCas9, expand the targeting scope of SpCas9 to sites to cover the vast majority of NR PAMs, greatly increasing the fraction of known human pathogenic SNPs that can in theory be corrected by base editing.

[00529] Among all pathogenic SNPs in the ClinVar database (Landrum et al., 2014) that are corrected by C•G to T•A conversion, 95% are targetable in principle with CBEs derived from SpCas9-NRRH, -NRCH, -NRTH, or SpCas9-NG/xCas9. Likewise, 95% of pathogenic SNPs in ClinVar that are correctable via A•T to G•C conversion can now be targeted with ABEs derived from the same set of Cas9 variants (Figure 39D).

[00530] In addition, these new variants greatly increase the number of possible protospacers available for targeting a given SNP for base editing: on average, there are 2.7 protospacers per pathogenic SNP targetable with CBE and 2.7 protospacers for those targetable with ABE with NR PAMs, compared to 1.7 targetable with CBE and 1.7 targetable with ABE, respectively, when using NG PAMs, and 1.3 and 1.3 protospacers available when using NGG PAMs only to target CBE and ABE, respectively (Figure 39E).

[00531] Since many pathogenic SNPs correctable by current base editors contain multiple targetable bases within the editing window (Figure 46C), expansion to NR PAMs enables multiple targeting strategies for a given SNP to optimize editing of the desired base, as we explicitly demonstrate below.

[00532] Collectively, these findings establish that evolved Cas9 variants SpCas9-NRRH, -NRCH, and -NRTH are compatible with both CBEs and ABEs, and thereby expand the targeting scope of base editing substantially.

Characterization of evolved SpCas9s on a human cell library of 11,776 integrated target sites

[00533] To comprehensively profile the PAM preferences of these variants, we analyzed the CBE efficiencies of our three evolved variants, SpCas9-NG, and SpCas9 on a library of 11,776 unique sequences in human cells. This library was designed using 46 distinct protospacers derived from sequences found in the human genome, each with different sequence contexts surrounding a fixed C at protospacer position 6, counting the PAM as positions 21-23. Each protospacer is adjacent to a

PAM sequence of 4Ns, and is additionally flanked with designated primer binding sites for amplification for highthroughput sequencing (HTS) analysis (Figure 40A).

[00534] Due to the very large number of target sites (Figure 47A), characterization of our evolved variants in this library format revealed PAM preferences in finer detail when compared to our bacterial depletion and endogenous mammalian genomic site editingexperiments (Figure 40B). Consistent with these previous experiments, our evolved variants exhibited the highest editing activity when either A or G was present at the 2nd PAM position (Figures 40C and 47B), when the 3rd PAM base was the one on which it was evolved (Figures 40D and 47B), and when a non-G was present at the 4th position of the PAM (Figures 40E and 47B). BE4-NG also showed the highest editing activity when either A or G was present at the 2nd PAM position (Figure 40C), but, unlike our evolved variants, was most active when a G was present at the 4th position of the PAM (Figure 40E and 47B) or when G or T was in the 3rd position (Figure 40D). In contrast, we found that BE4 editing efficiency at sites containing its canonical NGG PAM or its alternate NAG/NGA PAMs showed virtually no dependence on the 4th PAM nucleotide (Figure 40B). BE4 also showed some editing at sites containing a NCGG or NTGG PAM, which could be due to PAM slippage (Jiang et al., 2013), resulting in binding to a canonical NGG sequence.

[00535] Interestingly, our evolved variants and SpCas9-NG exhibit some level of editing activity at many more non-canonical PAMs when compared to SpCas9, supporting their broadened PAM scope (Figure 40B). Finally, both SpCas9-NG and our variants (most notably BE4-NRRH) performed best when a G was present at position 1 of the PAM and worst when a T was at this position; in contrast, BE4 exhibited only a slight preference for G at position 1 (Figures 40B, 47B, and 47C). Taken together, these results strongly support the PAM preferences observed in our bacterial depletion and endogenous mammalian genome editing experiments: specifically, recognition of NRRH, NRCH, NRTH, and NRNG PAMs for SpCas9-NRRH, -NRCH, -NRTH, and -NG, respectively.

[00536] Additionally, this library allowed us to investigate the tolerance of our variants to mismatches between the sgRNA and the target DNA sequence. The U6 promoter, commonly used to express sgRNAs in mammalian cells, initiates transcription with a 5' G. If a G is not natively present at the 5' end of the protospacer, guide sequences are typically either extended to the next native G, or simply transcribed with a mismatched 5' G at position -1 of the guide sequence. However, high-fidelity (HF) SpCas9s (Chen et al., 2017; Hu et al., 2018; Kleinstiver et al., 2016; Lee et al., 2018b; Slaymaker et al., 2016), which are less tolerant of mismatches between the protospacer and sgRNA, generally exhibit decreased efficiency when using a 21 nucleotide (nt) guide with a mismatched 5' G (Kim et al., 2017b; Zhang et al., 2017). Because PACE has previously led to SpCas9s with HF properties (Hu et al., 2018), we sought to determine if our new variants shared the same characteristics.

[00537] We investigated the average base editing activity of our evolved variants across all 11,776 library sites containing either a 20 nt protospacer with a matched 5' G ("20- matched"), a 21 nt protospacer with a matched 5' G ("21-matched"), or a 21 nt protospacer with a mismatched 5'G ("21-mismatched"). Our three evolved SpCas9 variants and SpCas9 all showed the highest base editing activity with a 20-matched sgRNA (Figures 40F, 40G, and 47D-F; however, interestingly, SpCas9-NG performed best with a 21- matched sgRNA (Figures 40F and 47D-F). When examining all NRNN PAMs, our variants and SpCas9 also showed a significant decrease in base editing efficiency when the sgRNA protospacer was increased to 21 nt, regardless if the 5' G was matched with the target sequence (Figures 40F, 40G, 47D, and 47E); in contrast, for SpCas9-NG this was only true when the 21-mismatched sgRNA (Figures 40G, 47D, and 47E). The magnitude of this decrease was similar to or greater for our evolved variants (SpCas9-NRRH: 23±2.7%, SpCas9-NRTH: 12±2.9%, SpCas9-NRCH: 14±2.9%) when compared to SpCas9 (13±5.3%). In contrast, SpCas9-NG demonstrated a preference for 21-matched sgRNAs, leading to an average 18.5±5.4% increase of editing efficiency when compared to 20-matched sgRNAs (Figures 40F, 47D, and 47E); however, a decrease in editing

efficiency was still observed with 21-mismatched sgRNAs (7.3±3.2%, Figures 40G, 47D, and 47E). Interestingly, the deleterious effect of using a 21 nt protospacer on the editing efficiency of our evolved variants and SpCas9 is lessened when targeting sites with NGNN or NGGN PAMs (Figures 40F, 40G, 47D, and 47F). This is especially true for SpCas9, which shows no significantly decreased base editing activity on sites with a 21 nt matched or mismatched protospacer when the PAM is NGG (Figures 40F and 47F). Together, these results suggest that our evolved variants are somewhat sensitive to the use of 21 nt sgRNA protospacers, and that this sensitivity is exacerbated by the presence of 5'G mismatches. Additionally, these experiments suggest that the optimal sgRNA protospacer length for SpCas9-NG may be longer than 20 nt.

Evolved SpCas9s enable efficient base editing of a pathogenic SNP

[00538] To demonstrate the utility of our evolved SpCas9 variants in a disease-relevant context, we targeted the Glu to Val point mutation at amino acid 6 of β-globin (HBB), which results in the HbS allele that is the most common cause of sickle-cell anemia (Rees et al., 2010). The HbS mutation arises from a GAG (Glu) to GTG (Val) codon change that cannot be reverted through current base editing technologies. However, this SNP can be edited with ABE to a GCG (Ala) through A•T to G•C conversion on the opposite strand (Figure 41A). The resulting HBB E6A genotype, known as the hemoglobin Makassar allele (HbG), has been reported as clinically normal in homozygous and heterozygous individuals (Quentin Blackwell et al., 1970; Sangkitporn et al., 2002; Viprakasit et al., 2002).

[00539] Unfortunately, the only NGG or NGN PAMs available at this site place the target A at either protospacer position 2 or 9, respectively, which fall outside the optimal editing window for ABE (positions 4-7) (Rees and Liu, 2018). However, two alternative target protospacer sequences using a CAT or CAC PAM place the target A at either position 4 or 7, respectively, with a bystander A present at either position 6 or 9 leading to a silent CCT to CCC (Pro to Pro) mutation. Thus, we tested the ability of our evolved variants, along with SpCas9-NG, to convert the sickle-cell SNP to the

Makassar mutation using these two protospacer sites with non-G PAMs. We transfected ABE-NRRH, -NRTH, and NRCH, or ABE-NG into HEK293T cells with homozygous GAG to GTG mutations at codon 6 of HBB (Figure 48A). While ABEs derived from the SpCas9 variants evolved in this study supported substantial (14-55%) A•T-to-G•C conversion using guide RNAs targeting either the CAT PAM or the CAC PAM site, ABE-NG edited efficiently (40±0.2%) only using the protospacer sequence containing a CAT PAM (Figures 41B and 41C), perhaps due to the presence of a G at the 4th position of the CAT PAM, which improves SpCas9-NG's recognition of NAN PAMs (see above).

[00540] Unfortunately, editing using the CAT PAM protospacer occurred primarily at the silent bystander base (position 6), with the target A (position 4) showing less than 10% editing across all four ABEs tested (Figures 41B and 48B).

[00541] Target base conversion of GTG to GCG in codon 6 of HBB using the CAC PAM protospacer, however, was much more efficient. As expected, ABE-NRCH showed the highest editing activity, with 41±3.8% base conversion at the target A (position 7) and 13±3.2% at the silent bystander A (position 9). ABE-NRRH and ABE-NRTH achieved 29±4.3% and 14±2.8% conversion, respectively (Figures 41C and 48C). In comparison, ABE-NG showed negligible (1.0±0.5%) target base conversion activity at this site (Figures 41C and 48C). Collectively, these results demonstrate that our evolved SpCas9 variants enable efficient base editing of previously inaccessible disease-relevant SNPs using non-G PAMs, and furthermore highlight the utility of evaluating multiple protospacer/PAM sequences for targeting a desired SNP.

Discussion

[00542] Here we report three new variants of SpCas9, evolved using phage-assisted continuous evolution (PACE), that are capable of recognizing NRRH (SEQ ID NO: 149), NRCH (SEQ ID NO: 150), and NRTH (SEQ ID NO: 151) PAM sequences. As our initial experiments suggested that increased selection stringency may be necessary to produce SpCas9 variants that were highly active on non-G PAMs, we developed several improved selection strategies for evolving Cas9:DNA

binding. Specifically, by increasing the number of target DNA protospacer/PAM sites that must be recognized by the evolving SpCas9 through use of an additional PACE-compatible selection marker (gVI), and limiting the total concentration of full-length SpCas9 in the host cell through use of a split-intein strategy, we were able to select for variants that efficiently recognize a desired PAM while reducing the probability of evolving undesired recognition of specific protospacer sequences. These improved selection strategies should be applicable to a majority of Cas9 orthologs, and enable the further evolution of Cas9 variants capable of targeting a wide range of PAM sequences.

[00543] From our initial experiments evolving SpCas9 for binding activity on all 16 individual NAN PAMs, we were able to identify three distinct groups of consensus mutations that conferred binding activity on NAA, NAT, and NAC PAMs, respectively (Figure 1D), leading us to split our subsequent evolutionary efforts into three separate trajectories to target these specific PAMs. Accordingly, the diverging consensus mutations of our evolved variants give insight to potential modes of PAM interaction.

[00544] SpCas9-NRRH, evolved to bind HAA PAMs, acquired a mutation at R1333, which in SpCas9 contacts the 2nd guanine in its canonical PAM, but not R1335, which contacts the 3rd NGG guanine (Figure 37B and 37D). The R1333K mutation likely allows SpCas9-NRRH to accept both A and G at the 2nd PAM nucleotide, while the preservation of R1335 may explain why this variant recognizes both NAA and NAG PAMs. On the other hand, SpCas9-NRTH (evolved to bind HAT PAMs) preserves R1333 but eliminates R1335 through mutation to a Leu (Figure 37E and 37G). Interestingly, SpCas9-NRTH shows a strong preference for T in the 3rd PAM position and appears to have lost some recognition of the wild-type NGG PAM (Figure 40B). Finally, SpCas9-NRCH displays altered interactions at both R1335 and T1337 (Figure 37F and 37H); the T1337N in particular may form contacts with a 4th PAM nucleotide to compensate for weakened binding interactions with the HAC target PAM.

[00545] In addition to alterations at residues responsible for direct contacts with PAM nucleobases, we observed a number of additional mutations which we suspect modulate more general interactions with the target- and non-target DNA, including R1114G, E1219V, Q1221H, and D1135N (Figure 37B, 37D-H). Residue E1219 forms hydrogen bonds with R1335 in SpCas9, and mutations at this residue are thought to destabilize the interaction between R1335 and the 3rd PAM guanine. Mutations at residue D1135 have been previously reported and are thought to modulate interactions with the sugarphosphate backbone of the non-target DNA strand; R1114G and Q1221H may alter similar interactions. Finally, we observed mutations in the helical domain of Cas9 that arose in several independently evolving populations (Figure 44B).

[00546] These mutations, when added to the N-terminal region of NRRH and NRCH, improve their recognition of non-G PAMs in base editing experiments (Figure 44C), and may contribute to increasing the overall DNA binding/unwinding activity of these variants. Along with bacterial PAM depletion and mammalian cell genome editing on endogenous genomic sites spanning all NANN PAMs, we characterized our variants and SpCas9-NG using a 11,776-member sgRNA/protospacer/NNNN PAM library that was genomically integrated into HEK293T cells. The large number of sites examined greatly increases our ability to confidently profile the editing activity of these proteins using all NNNN PAMs in a human cell context, and illuminated the sequence preferences of these Cas9 variants, including previously uncovered activity of SpCas9-NG on NANG PAMs.

[00547] Both our bacterial PAM depletion experiments and mammalian library data demonstrated that our evolved variants display a different 4th base PAM preference (H) compared to SpCas9-NG (G), suggesting that they may have complimentary utility. While further investigation is required to explain the 4th base preferences of our mutants, crystal structures of SpCas9-NG and other previously reported evolved SpCas9s (VRER/VRQR) suggest that the T1337R mutation in these variants may

create a direct interaction with the 4th base G (Anders et al., 2016; Hirano et al., 2016b; Nishimasu et al., 2018).

[00548] Additionally, both SpCas9-NG and our variants display a moderate preference for G at the 1st PAM position, whereas this preference in this position in SpCas9 is virtually nonexistent (Figure 47B and 47C). Because of these numerous sequence preferences, we suggest screening all variants reported here along with SpCas9-NG when optimizing targeting efficiency on sites with NR PAMs, and provide a recommended list of SaCas9 and SpCas9 variants to test for targeting any given NRNN PAM (FIG. 42). However, we note that other Cas9 orthologs and related CRISPR effector proteins not included here have been also been shown to mediate genome editing in mammalian cells (Chatterjee et al., 2018; Cong et al., 2013; Edraki et al., 2019; Esvelt et al., 2013; Harrington et al., 2017; Hirano et al., 2016a; Hou et al., 2013; Kim et al., 2017a; Zetsche et al., 2015). Our evolved variants, along with SpCas9-NG, expand the utility of SpCas9 towards disease-relevant genome editing applications. Access to a broad range of PAMs is especially essential for base editing, as illustrated by our experiments targeting the sickle cell mutation of human β-globin. While ABE-NG was able to bind to this locus using a CATG PAM, the majority of base editing we observed occurred at an off-target A within the window (Figure 41B). However, we were able to achieve high levels of conversion at the correct base and lower levels of off-target editing with our evolved variants by using an adjacent CACC PAM (Figure 41C). Notably, the sickle cell SNP occurs within the optimal ABE window for both sgRNAs tested, suggesting that it may be beneficial to assay several protospacer sequences for a single target. Expanding the PAMs accessible by Cas9 variants to NR increases not only the number of targetable pathogenic SNPs (Figure 39D), but also the number of possible sgRNAs that can target an individual SNP (Figure 39E). Additionally, although only results from indel formation and base editing are shown in this work, we anticipate that our evolved variants should be compatible with the majority of Cas9-associated genome editing technologies. Access to

NR PAMs should benefit all precision genome editing applications, including other base editing applications, HDR, and predictable template-free genome editing.

Example 13 REFERENCES

The following references are incorporated herein by reference.

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAMdependent target DNA recognition by the Cas9 endonuclease. Nature 513, 569–573.

Anders, C., Bargsten, K., and Jinek, M. (2016). Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. Mol. Cell 61, 895–902.

Badran, A.H., Guzov, V.M., Huai, Q., Kemp, M.M., Vishwanath, P., Kain, W., Nance, A.M., Evdokimov, A., Moshiri, F., Turner, K.H., et al. (2016). Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. Nature 533, 58–63.

Brödel, A.K., Jaramillo, A., and Isalan, M. (2016). Engineering orthogonal dual transcription factors for multi-input synthetic promoters. Nat. Commun. 7, 13858.

Bryson, D.I., Fan, C., Guo, L.-T., Miller, C., Söll, D., and Liu, D.R. (2017). Continuous directed evolution of aminoacyl-tRNA synthetases. Nat. Chem. Biol. 13, 1253–1260.

Carlson, J.C., Badran, A.H., Guggiana-nilo, D.A., and Liu, D.R. (2014). Negative selection and stringency modulation in phage-assisted continuous evolution. Nat. Chem. Biol. 10, 216–222.

Cebrian-Serrano, A., and Davies, B. (2017). CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. Mamm.

Genome 28, 247–261.

Chatterjee, P., Jakimo, N., and Jacobson, J.M. (2018). Minimal PAM specificity of a highly similar SpCas9 ortholog. Sci. Adv. 4, eaau0766.

Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington, L.B., Sternberg, S.H., Joung, J.K., Yildiz, A., and Doudna, J.A. (2017). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature 550, 407–410.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science (80-.). 339, 819–823.

Dickinson, B.C., Leconte, A.M., Allen, B., Esvelt, K.M., and Liu, D.R. (2013). Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. Proc. Natl. Acad. Sci. 110, 9007–9012.

Dickinson, B.C., Packer, M.S., Badran, A.H., and Liu, D.R. (2014). A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. Nat. Commun. 5, 5352.

Edraki, A., Mir, A., Ibraheim, R., Gainetdinov, I., Yoon, Y., Song, C.Q., Cao, Y., Gallant, J., Xue, W., Rivera-Pérez, J.A., et al. (2019). A Compact, High-Accuracy Cas9 with a Dinucleotide PAM for In Vivo Genome Editing. Mol. Cell 73, 714-726.e4.

Esvelt, K.M., Carlson, J.C., and Liu, D.R. (2011). A system for the continuous directed evolution of biomolecules. Nature 472, 499–503.

Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yaung, S.J., and Church, G.M. (2013). Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat. Methods 10, 1116–1121.

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of T to G C in genomic DNA without DNA cleavage. Nature 551, 464–471.

Harrington, L.B., Paez-Espino, D., Staahl, B.T., Chen, J.S., Ma, E., Kyrpides, N.C., and Doudna, J.A. (2017). A thermostable Cas9 with increased lifetime in human plasma. Nat. Commun. 8, 1424.

Hirano, H., Gootenberg, J.S., Horii, T., Abudayyeh, O.O., Kimura, M., Hsu, P.D., Nakane, T., Ishitani, R., Hatada, I., Zhang, F., et al. (2016a). Structure and Engineering of Francisella novicida Cas9. Cell 164, 950–961.

Hirano, S., Nishimasu, H., Ishitani, R., and Nureki, O. (2016b). Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. Mol. Cell 61, 886–894.

Hou, Z., Zhang, Y., Propson, N.E., Howden, S.E., Chu, L.-F., Sontheimer, E.J., and Thomson, J.A. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc. Natl. Acad. Sci. 110, 15644–15649.

Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W., Chen, L., Sun, N., Zeina, C.M., Gao, X., Rees, H.A., Lin, Z., et al. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature 556, 57–63.

Hubbard, B.P., Badran, A.H., Zuris, J.A., Guilinger, J.P., Davis, K.M., Chen, L., Tsai, S.Q., Sander, J.D., Joung, J.K., and Liu, D.R. (2015). Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. Nat. Methods 12, 939–942.

Jiang, F., and Doudna, J.A. (2017). CRISPR–Cas9 Structures and Mechanisms. Annu.Rev. Biophys. 46, 505–529.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. 31, 233–239.

Jin, S., Zong, Y., Gao, Q., Zhu, Z., Wang, Y., Qin, P., Liang, C., Wang, D., Qiu, J.L., Zhang, F., et al. (2019). Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. Science (80-.). 364, 292–295.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (80-.). 337, 816–821.

Karvelis, T., Gasiunas, G., Young, J., Bigelyte, G., Silanskas, A., Cigan, M., and Siksnys, V. (2015). Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. Genome Biol. 16, 253.

Kim, E., Koo, T., Park, S.W., Kim, D., Kim, K., Cho, H.Y., Song, D.W., Lee, K.J., Jung, M.H., Kim, S., et al. (2017a). In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. Nat. Commun. 8, 14500.

Kim, S., Bae, T., Hwang, J., and Kim, J.S. (2017b). Rescue of high-specificity Cas9 variants using sgRNAs with matched 5' nucleotides. Genome Biol. 18, 218.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Topkar, V. V, Zheng, Z., and Joung, J.K. (2015a). Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat. Biotechnol. 33, 1293–1298.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V. V, Nguyen, N.T., Zheng, Z., Gonzales, A.P.W., Li, Z., Peterson, R.T., Yeh, J.R.J., et al. (2015b). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature 523, 481–485.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genomewide off-target effects. Nature 529, 490–495.

Koblan, L.W., Doman, J.L., Wilson, C., Levy, J.M., Tay, T., Newby, G.A., Maianti, J.P., Raguram, A., and Liu, D.R. (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nat. Biotechnol. 36, 843–846.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420–424.

Komor, A.C., Badran, A.H., and Liu, D.R. (2017). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell 168, 20–36.

Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 42, D980--D985.

Lee, H.K., Willi, M., Miller, S.M., Kim, S., Liu, C., Liu, D.R., and Hennighausen, L. (2018a). Targeting fidelity of adenine and cytosine base editors in mouse embryos. Nat. Commun. 9, 4804

Lee, J.K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y.-H., Lee, K., Jung, I., Kim, D., Kim, S., et al. (2018b). Directed evolution of CRISPR-Cas9 to increase its specificity. Nat. Commun. 9, 3048.

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T.,Abudayyeh, O.O., Gootenberg, J.S., Mori, H., et al. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science (80-.). 361, 1259–1262.

Packer, M.S., Rees, H.A., and Liu, D.R. (2017). Phage-assisted continuous evolution of proteases with altered substrate specificity. Nat. Commun. 8, 956.

Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., Olsen, K.M., Gregg, A., Noggle, S., and Tessier-Lavigne, M. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature 533, 125–129.

Pu, J., Zinkus-Boltz, J., and Dickinson, B.C. (2017). Evolution of a split RNA polymerase as a versatile biosensor platform. Nat. Chem. Biol. 13, 432–438.

Quentin Blackwell, R., Oemijati, S., Pribadi, W., Weng, M.I., and Liu, C.S. (1970).

Hemoglobin G Makassar: β6 Glu→Ala. BBA - Protein Struct. 214, 396–401.

Rees, H.A., and Liu, D.R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. Nat. Rev. Genet. 19, 770–778.

Rees, D.C., Williams, T.N., and Gladwin, M.T. (2010). Sickle-cell disease. Lancet 376, 45 2018–2031.

Roth, T.B., Woolston, B.M., Stephanopoulos, G., and Liu, D.R. (2019). Phage-Assisted Evolution of Bacillus methanolicus Methanol Dehydrogenase 2. ACS Synth. Biol. 8, 796–806.

Sangkitporn, S., Rerkamnuaychoke, B., Sangkitporn, S., Mitrakul, C., and Sutivigit, Y. (2002). Hb G Makassar (beta 6: Glu→ Ala) in a Thai Family. 85, 577–582.

Shen, M.W., Arbab, M., Hsu, J.Y., Worstell, D., Culbertson, S.J., Krabbe, O., Cassa,C.A., Liu, D.R., Gifford, D.K., and Sherwood, R.I. (2018). Predictable and precise template-free CRISPR editing of pathogenic variants. Nature 563, 646–651.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016).

Rationally engineered Cas9 nucleases with improved specificity. Science (80-.). 351,84–88.

Suzuki, T., Miller, C., Guo, L.T., Ho, J.M.L., Bryson, D.I., Wang, Y.S., Liu, D.R., andSöll, D. (2017). Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNAsynthetase. Nat. Chem. Biol. 13, 1261–1266.

Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V. V, Thapar, V., Wyvekens,N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. 33, 187–197.

Viprakasit, V., Wiriyasateinkul, A., Sattayasevana, B., Miles, K.L., and Laosombat, V. (2002). Hb G-Makassar [β6(A3)Glu→Ala; codon 6 (GAG→GCG)]: Molecular characterization, clinical, and hematological effects. Hemoglobin 26, 245–253

Wang, T., Badran, A.H., Huang, T.P., and Liu, D.R. (2018). Continuous directed evolution of proteins with improved soluble expression. Nat. Chem. Biol. 14, 972–980.

Xin, H., Wan, T., and Ping, Y. (2019). Off-Targeting of Base Editors: BE3 but not ABE induces substantial off-target single nucleotide variants. Signal Transduct. Target. Ther. 4, 9

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., Van Der Oost, J., Regev, A., et al. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell 163, 759–771. 46

Zhang, D., Zhang, H., Li, T., Chen, K., Qiu, J.L., and Gao, C. (2017). Perfectly matched 20-nucleotide guide RNA sequences enable robust genome editing using high-fidelity SpCas9 nucleases. Genome Biol. 18, 191.

Zuo, E., Sun, Y., Wei, W., Yuan, T., Ying, W., Sun, H., Yuan, L., Steinmetz, L.M., Li, Y., and Yang, H. (2019). Cytosine base editor generates substantial off-target singlenucleotide variants in mouse embryos. Science (80-.). 364, 289–292.

SEQUENCES

SEQ ID NO:	Sequence	Description of Sequence
	Cas9 sequences	
1	ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCG CCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTT GTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTT GGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGC GCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAA ACCCTATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTT GGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTT ATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTAT CAGCTTCAATGATTAAGCGCTACGATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAAC GGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGA TTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAAGACAATC GTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGG AATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGA GTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAGGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAA ATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCAT GATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGGGATGATTGAGGAAAG ACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAAT CTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGATATTCAAAAAGCACAGGTGTCT GGACAAGGCCATAGTTTACATGAACAGATTGCTAACTTAGCTGGCAGTCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAATGGGGCATAA GCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTCAGA TTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGT GATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAGACGATTCAATAGACAATAAGGTACTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGA AGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATAAAG CTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTT AAAGTGATTACCTTAAAATCTAAATTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCTAAATGCCGTCGT TGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAA CCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAAATT GTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACC AAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAG GGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAAA AAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAA ATATGTGAATTTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTG AGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATT CATTTATTTACGTTGACGAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCA ATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA	NC_017053.1 Cas9 wild type [Streptococcus pyogenes MGAS1882]
2	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENII HLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	NC_017053.1 Cas9 wild type [Streptococcus pyogenes MGAS1882]
3	ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGATGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCG TCATTCGATTAAAAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATAT GTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTT GGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGC CCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGA ACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTC	Cas9 wild type [Streptococcus pyogenes]

	GGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGACAATCT ACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTAT CCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAAC GGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGA TCTACTGCGAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCAAAGACAATC GTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGG AATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGA GTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCA ACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCAT GACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTGAGGAAAG ACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAA GTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCC GGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCA CAAACCGGAAAACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCC AGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTA TCTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGA GGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGACA AGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAA GTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGCTTATCTTAATGCCGT CGTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGG CTACAGCCAAATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAA ATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCT TCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGA AGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTA AAAAAGGATCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGTC TAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCGACGAAATCA TAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATT ATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCA CCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTGGGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGACGGTGATTATA AAGATCATGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA	
4	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGSPKKKRKVSSDYKDHDGDYKDHDIDYKDDDDKAAG	Cas9 wild type [Streptococcus pyogenes]
5	ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCG CCACAGTATCAAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTT GTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTT GGAAATATAGTAGATGAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGTAGATTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGC GCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAA ACCCTATTAACGCAAGTGGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGCTCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTT GGGAATCTCATTGCTTTGTCATTGGGTTTGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGATAATTT ATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTCAGATATCCTAAGAGTAAATACTGAAATAACTAAGGCTCCCCTAT CAGCTTCAATGATTAAACGCTACGATGAACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTTGATCAATCAAAAAAC GGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGA TTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAAAAGACAATC GTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGG AATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGA GTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAA ATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCAT GATTTGCTAAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAG ACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAAT CTGGCAAAACAATATTAGATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCT GGACAAGGCGATAGTTTACATGAACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAAGTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCA TAAGCCAGAAAATATCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAATCGAAGAAGGTATCAAAGAATTAGGAAGTC AGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAAAAGCTCTATCTC[illegible]AAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTA	NC_002737.2 [Streptococcus pyogenes M1GAS]

	AGTGATTATGATGTCGATCACATTGTTCCACAAAGTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGA AGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACTTCTAAACGCCAAGTTAATCACTCACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAACTTGATA AAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACGATGAAAATGATAAACTTATTCGAGAG GTTAAAGTGATTACCTTAAAATCTAAAATTAGTTTCGACTTCCGAAAGATTTCAATTTTATCAAAGAATTACCATATGCCCATGATCGCATTGATCCAATATTCGGT CGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGATTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAG CAACCGCAAAATATTTCTTTTACTCTAATAATCATGAACTTCTTCAAACAGAAATTACACTTGCAAACGCCCTCTAATCGAACTAATGGGAAACTGGAGAA ATTGTCTGGGATAAAGGGCGAGATTTTGCCACGAGTGTCCATGCCCAAGAAGTACAGAAGGCGGATTCTCCAAGGAGTCAATTT ACCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAA AAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTT AAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGGCTAGTGCCGGAGAATTACAAAAAGGAAATATGAGCTGGCTCTGCCAAG CAAATATGTGAATTTTTATTTAGCTATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACACAAAAACAAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTA TTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTAGCAGATGCCAATTAGATAAAGTTCTTAGTGCATATAACAAACATAGAGACAAAGCAGACAAGCAGAAAATATT ATTCATTTATTTACGTTGACGAATCTTTGGAGCTCCCGCTGCTTTAAATATTTTGATATCGTAAACGATATACGTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCA TCAATCCATCACTCGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA	
6	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLIKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
7	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLIKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
8	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLIKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
9	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLIKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	

10	KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHN VNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTY FPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIY QSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIE LAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSD SKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIF KEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDP QTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKL KKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	Staphylococcus aureus Cas9
11	MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRRLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARILL HLAKRRGFRSNRKSERTNKENSTMLKHIEENQSILSSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVG FCTFEPKEKRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHDVRTLLNLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVYGKGAAKS FRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLADKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQKTVLLPNIPPIANPVVMRA LTQARKVVNAIIKKYGSPVSIHIELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDHVIPYSRSLDDSYT NKVLVLTKENREKGNRTPAEYLGLGSERWQQFETFVLTNKQFSKKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVNGRITAHLRSRWNFNKNRE ESNLHHAVDAAIVACTTPSDIARVTAFYQRREQNKELSKKTDPQFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMPKRSITGAAHQETLRRYIGIDERSGKI QTVVKKKLSEIQLDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGELGPIIRTIKIIDTTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTIDMMKGI LPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIDSSNGGLSLVSHDNNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRGEKRVGVASSS HSKAGETIRPL	Geobacillus thermodenitrificans Cas9
	Cas9 circular permutants	
12	DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKK YGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD GGSGGSGGSGGSGGSGGSGGDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHR LEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKS RRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLAR GNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFD SVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ LIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYG	Cas9 varient CP1012 [Synthetic]
13	EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHY LDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGSGGSGGSG GSGGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHER HPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQK AQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD INRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ	Cas9 varient CP1028 [Synthetic]
14	NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGSGGSGGSGGSGGDKKYSIGLA IGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYH EKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGA SQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKD FLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHI ANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKL VSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS	Cas9 varient CP1041[Synthetic]
15	PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG DGGSGGSGGSGGSGGSGGSGGMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFF HRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLS KSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKORTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL	Cas9 varient CP1249 [Synthetic]

	ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIEC FDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEI TLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIM ERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS	
16	KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSGGSGGSGGSGGSGGDKKYSIGLAIGTNSVGWAVITDEYKVPSKK FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKAD LRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKD TYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKV LPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLF EDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNR GKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQ KGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRD	Cas9 varient CP1300 [Synthetic]
17	DYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKK YGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	Cas9 varient CP1012 [Synthetic]
18	EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHY LDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	Cas9 varient CP1028 [Synthetic]
19	NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSK RVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	Cas9 varient CP1041[Synthetic]
20	PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGG D	Cas9 varient CP1249 [Synthetic]
21	KPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	Cas9 varient CP1300 [Synthetic]
	Cas9 with non-canonical PAM specificity	
22	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGFNSPTAAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIGFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGVLHKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGVPAAFKYFDTTIDKKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SpCas9-NRRH variant with non- canonical PAM specificity [Synthetic]
23	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTINRKQYNTTKEVLDATLIRQSITGLYETRIDLSQLGGD	SpCas9-NRCH variant with non- canonical PAM specificity [Synthetic]

24	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASASVLHKGNELALPSKYVNFLYLASHYEKLKGSSEDNKQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIGRKIYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	SpCas9-NRTH variant with non- canonical PAM specificity[Synth etic]
	High fidelity Cas9 / other Cas9 mutants	
25	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLINGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASASVLHKGNELALPSKYVNFLYLASHYEKLKGSSEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	Cas9 mutations [Synthetic]
	cytidine deaminases	
27	MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDR	AID [Homo sapiens]
28	MDSLIMKQKKFLYHFKNVRWAKGRHETYICYVVKRRDSATSCSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPNLSLRIFTARLYFCEDR	AID [Mus musculus]
29	MDSLLMKQRKFLYHFKNVRWAKGRHETYICYVVKRRDSATSFSLDFGHLRNKSGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFAARLYFCEDR	AID [Canis familiaris]
30	MDSLLKKQRQFLYQFKNVRWAKGRHETYICYVVKRRDSPTSFSLDFGHLRNKAGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGYPNLSLRIFTARLYFCDKE	AID [Bos taurus]
31	MAVGSKPKAALVGPHWERERIWCFLCSTGLGTQQTGQTSRWLRPAATQDPVSPPRSLLMKQRKFLYHFKNVRWAKGRHETYLCYVVKRRDSATSFSLDFGYLRNKSGCHVELLFLRYIS DWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYGWGALPAGLMSPARPSDYFYCWNTFVENHERTFKAWEGLHENSVRLSRRLRRILLPLYEVDDLRDAFRTLG L	AID [Rattus]
32	MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLGYAKGRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQIVRFLATH HNLSLDIFSSRLYNVQDPETQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKRLLTNFRYQDSKLQEILRPCYIPVPSSSSSSTLSNICLTKGLPETRFCVEGRRMDPLSE EEFYSQFYNQRVKHLCYYHGRMKPYLCYQLEQFNGQAPLKGCLLSEKGKQHAEILFLDKIRSMELSQVTITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSL WQSGIIVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLRRIKESWGLQDLVNDFGNLQLGPPMS	APOBEC-3 [Mus musculus]
33	MGPFCLGCSHRKCYSPIRNLISQETFKFHFKNLRYAIDRKDTFLCYEVTRKDCDSPVSLHHGVFKNKDNIHAEICFLYWFHDKVLKVLSPREEFKITWYMSWSPCFECAEQVLRFLATH HNLSLDIFSSRLYNIRDPENQQNLCRLVQEGAQVAAMDLYEFKKCWKKFVDNGGRRFRPWKKLLTNFRYQDSKLQEILRPCYIPVPSSSSSSTLSNICLTKGLPETRFCVERRRVHLLSE EEFYSQFYNQRVKHLCYYHGVKPYLCYQLEQFNGQAPLKGCLLSEKGKQHAEILFLDKIRSMELSQVIITCYLTWSPCPNCAWQLAAFKRDRPDLILHIYTSRLYFHWKRPFQKGLCSL WQSGIIVDVMDLPQFTDCWTNFVNPKRPFWPWKGLEIISRRTQRRLHRIKESWGLQDLVNDFGNLQLGPPMS	APOBEC-3 [Rattus]

34	MVEPMDPRTFVSNFNNRPILSGLNTVWLCCEVKTKDPSGPPLDAKIFQGKVYSKAKYHPEMRFLRWFHKWRQLHHDQEYKVTWYVSWSPCTRCANSVATFLAKDPKVTLTIFVARLYYF WKPDYQQALRILCQKRGGPHATMKIMNYNEFQDCWNKFVDGRGKPFKPRNNLPKHYTLLQATLGELLRHLMDPGTFTSNFNNKPWVSGQHETYLCYKVERLHNDTWVPLNQHRGFLRNQ APNIHGFPKGRHAELCFLDLIPFWKLDGQQYRVTCFTSWSPCFSCAQEMAKFISNNEHVSLCIFAARIYDDQGRYQEGLRALHRDGAKIAMMNYSEFEYCWDTFVDRQGRPFQPWDGLD EHSQALSGRLRAI	APOBEC-3G [Macaca mulatta] Rhesus macaque
35	MKPHFRNPVERMYQDTFSDNFYNRPILSHRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSKLKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDVATFLAEDPKVTLTIF VARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTSNFNNELWVRGRHETYLCYEVERLHNDTWVLLNQR RGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLHQDYRVTCFTSWSPCFSCAQEMAKFISNNKHVSLCIFAARIYDDQGRCQEGLRTLAKAGAKISIMTYSEFKHCWDTFVDHQGCP FQPWDGLEEHSQALSGRLRAILQNQGN	APOBEC-3G [Pan troglodytes] - Chimpanzee
36	MNPQIRNMVEQMEPDIFVYYFNNRPILSGRNTVWLCYEVKTKDPSGPPLDANIFQGKLYPEAKDHPEMKFLHWFRKWRQLHRDQEYEVTWYVSWSPCTRCANSVATFLAEDPKVTLTIF VARLYYFWKPDYQQALRIILCQERGGPHATMKIMNYNEFQHCWNEFVDGQGKPFKPRKNLPKHYTLLHATLGELLRHVMDPGTFTSNFNNKPWVSGQRETYLCYKVERSHNDTWVLLNQH RGFLRNQAPDRHGFPPKGRGHAELCFLDLIPFWKLDDQQYRVTCFTSWSPCFSCAQKMAKFISNNKHVSLCIFAARIYDDQGRC37QEGLRTLHRDGAKIAVMNYSEFEYCWDTFVDRQGR PFQPWDGLDEHSQALSGRLRAI	APOBEC-3G [Chlorocebus sabaeus]
37	MK38PHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMRFFHWFSKWRKLHR39DQEYEVTWYISWSPCTKCTRDMATFLAEDPKVT LTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVL LNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDH QGCPFQPWDGLDEHSQDLSGRLRAILQNQEN	APOBEC-3G [Homo sapiens]
38	MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPRLDAKIFRGQVYSQPEHHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLAEHPNVTLTISA ARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYSEGQPFMPWYKFDDNYAFLHRTLKEILRNPMEAMYPHIFYFHFKNLRKAYGRNESWLCFTMEVVKHHSPVSWKRGVF RNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGLRSLSQEGASVEIMGYKDFKYCWENFVVNDDEPFKPWKGL KYNFLFLDSKLQEILE	APOBEC-3F [Homo sapiens]
39	MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGQVYFKPQYHAEMCFLSWFCGNQLPAYKCFQITWFVSWTPCPDCVAKLAEFLSEHPNVTLTIS AARLYYYWERDYRRALCRLSQAGARVTIMDYEEFAYCWENFVYNEGQQFMPWYKFDENYAFLHRTLKEILRYLMDPDTFTFNFNNDPLVLRRRQTYLCYEVERLDNGTWVLMDQHMGFL CNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEAILQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQ PWDGLEEHSQALSGRLRAILQNQGN	APOBEC-3B [Homo sapiens]
40	MQPQGLGPNAGMGPVCLGCSHRRPYSPIRNPLKKLYQQTFYFHFKNVRYAWGRKNNFLCYEVNGMDCALPVPLRQGVFRKQGHTHAELCFIYWFHDKVLRVLSPMEEFKVTWYMSWSPC SKCAEQVARFLAAHRNLSLAIFSSRLYYYLRNPNYQQKLCRLIQEGVHVAAMDLPEFKKCWNKFVDNDGQPFRPWMRLRINFSPYDCKLQEIFSRMNLLREDVFYLQFNNSHRVKPVQN RYYRRKSYLCYQLERANGQEPLKGYLLYKKGEQHVELLFLEKMRSMELSQVRITCYLTWSPCPNCARQLAAFKKDHPDLILRIYTSRLYFYWRKKFQKGLCTLWRSGIHVDVMDLPQFA DCWTNFVNPQRPFRPWNELEKNSWRIQRRLRRIKESWGL	APOBEC-3B [Bos taurus]
41	DGWEVAFRSGTVLKAGVLGVSMTEGWAGSGHPGQGACVWTPGTRNTMNLLREVLFKQQFGNQPRVPAPYYRRRTYLCYQLKQRNDLTLDRGCFRNKKQRHAEIRFIDKINSLDLNPSQS YKIICYITWSPCPNCANELVNFITRNNHLKLEIFASRLYFHWIKSFKMGLQDLQNAGISVAVMTHTEFEDCWEQFVDNQSRPFQPWDKLEQYSASIRRRLQRILTAPI	APOBEC-3B [Pan troglodytes]
42	MNPQIRNPMEWMTQRTFYYNFENEPILYGRSYTWLCYEVKIRRGHSNLLWDTGVFRGQMYSQPEHHAEMCFLSWFCGNQLSAYKCFQITWFVSWTPCPDCVAKLAKFLAEHPNVTLTIS AARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFVYNEGQPFMPWYKFDDNYAFLHRTLKEILRHLMDPDTFTFNFNNDPLVLRKRQTYLCYEVERLDNGTWVLMDQHMGFL CNEAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGQVRAFLQENTHVRLRIFAARIYDYDPLYKEAILQMLRDAGAQVSIMTYDEFEYCWDTFVYRQGCPFQ PWDGLEEHSQALSGRLRAILQVRASSLCMVPHRPPPPPQSPGPCLPLCSEPPLGSLLPTGRPAPASLPFLLTASFSFPPPASLPPLPSLSLSPGHLPVPSFHSLTSCSIQPPCSSRIRET EGWASVSKEGRDLG	APOBEC-3C [Homo sapiens]
43	MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETHCHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDCAGEVAEFLARHSNVNLTIF TARLYYFQYPCYQEGLRSLSQEGVAVEIMDYEDFKYCWENFVYNDNEPFKPWKGLKTNFRLLKRRLRESLQ	APOBEC-3C [Homo sapiens]
44	MNPQIRNPMKAMYPGTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTGVFRNQVDSETHCHAERCFLSWFCDDILSPNTNYQVTWYTSWSPCPECAGEVAEFLARHSNVNLTIF TARLYYFQDTDYQEGLRSLSQEGVAVKIMDYKDFKYCWENFVYNDDEPFKPWKGLKYNFRFLKRRLQEILE	Gorilla APOBEC3C [Gorilla gorilla]
45	MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGFLHNQAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTWFISWSPCFSWGCAGEVRAFLQENTH VRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN	APOBEC-3A [Homo sapiens]
46	MDGSPASRPRHLMDPNTFTFNFNNDLSVRGRHQTYLCYEVERLDNGTWVPMDERRGFLCNKAKNVPCGDYGCHVELRFLCEVPSWQLDPAQTYRVTWFISWSPCFRRGCAGQVRVFLQE NKHVRLRIFAARIYDYDPLYQEALRTLRDAGAQVSIMTYEEFKHCWDTFVDRQGRPFQPWDGLDEHSQALSGRLRAILQNQGN	APOBEC-3A [Macaca mulatta]
47	MDEYTFTENFNNQCWPSKTYLCYEMERLDGDATIPLDEYKGFVRNKGLDQPEKPCHAELYFLGKIHSWNLDRNQHYRLTCFISWSPCYDCAQKLTTFLKENHHISLHILASRIYTHNRF GCHQSGLCELQAAGARITIMTFEDFKHCWETFVDHKGKPFQPWEGLNVKSQALCTELQALILKTQQN	APOBEC-3A [Bos taurus]
48	MALLTAETFRLQFNNKRRLRRPYYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFINEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLGIFASRLYYHWCKPQ QKGLRILCGSQVPVEVMGFPKFADCWENFVDHEKPLSFNPYKMLEELDKNSRAIKRRLERIKIPGVRAQGRYMDILCDAEV	APOBEC-3H [Homo sapiens]
49	MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAEIRFINKIKSMGLDETQCYQVTCYLTWSPCPSCAGELVDFIKAHRHLNLRIFASRLYYHWRPNY QRGLLLLCGSQVPVEVMGLPEFTDCWENFVDHKEPPSFNPSEKLEELDKNSQAIKRRLERIKSRSVDVLENGLRSLQLGPVTPSSSIRNSR	APOBEC-3H [Macaca mulatta]
50	MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGVFRGPVLPKRQSNHRQEVYFRFENHAEMCFLSWFCGNRLPANRRFQITWFVSWNPCLPCVVKVTKF LAEHPNVTLTISAARLYYYRDRDWRWVLLRLHKAGARVKIMDYEDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEILRNPMEAMYPHIFYFHFKNLLKACGRNESWLCFTMEVI KHHSAVFRKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCPECAGEVAEFLARHSNVNLTIFTARLCYFWDTDYQEGLCSLSQEGASVKIMGVKDFVSCWKNFV YSDDEPFKPWKGLQTNFRLLKKRLREILQ	APOBEC-3D [Homo sapiens]
51	MTSEKGPSTGDPTIRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKNTTNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYVARL FWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLIHPSVAWR	APOBEC-1 [Homo sapiens]
52	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSVWRHTSQNTSNHVEVNFLEKFTTERYFRPNTRCSITWFLSWSPCGECSRAITEFLSRHPYVTLFIYIARL YHHTDQRNRQGLRDLISSGVTIQIMTEQEYCYCWRNFVNYPPSNEAYWPRYPHLWVKLYYVLELYCIILGLPPCLKILRRKQPQLTFFTITLQTCHYQRIPPHLLWATGLK	APOBEC-1 [Mus musculus]
53	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK	APOBEC-1 [Rattus]

54	MAQKEEAAVATEAASQNGEDLENLDDPEKLKELILPPFEIVTGERLPANFFKFQFRNVEYSSGRNKTFLCYVVEAQGKGGQVQASRGYLEDEHAAAHAEEAFFNTILPAFDPALRYNV TWYVSSSPCAACADRIIKTLSKTKNLRLLILVGRLFMWEEPEIQAALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEKLADILK	APOBEC-2 [Homo sapiens]
55	MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKTFLCYVVEVQSKGGQAQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNV TWYVSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK	APOBEC-2 [Mus musculus]
56	MAQKEEAAEAAAPASQNGDDLENLEDPEKLKELIDLPPFEIVTGVRLPVNFFKFQFRNVEYSSGRNKTFLCYVVEAQSKGGQVQATQGYLEDEHAGAHAEEAFFNTILPAFDPALKYNV TWYVSSSPCAACADRILKTLSKTKNLRLLILVSRLFMWEEPEVQAALKKLKEAGCKLRIMKPQDFEYLWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK	APOBEC-2 [Rattus]
57	MAQKEEAAAAAEPASQNGEEVENLEDPEKLKELIELPPFEIVTGERLPAHYFKFQFRNVEYSSGRNKTFLCYVVEAQSKGGQVQASRGYLEDEHAINHAEEAFFNSIMPTFDPALRYMV TWYVSSSPCAACADRIVKTLNKTKNLRLLILVGRLFMWEEPEIQAALRKLKEAGCRLRIMKPQDFEYIWQNFVEQEEGESKAFEPWEDIQENFLYYEEKLADILK	APOBEC-2 [Bos taurus]
58	MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRGNGHTL KIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV	CDA1 [Petromyzon marinus]
59	MKPHFRNTVERMYRDTFSYNFYNRPILSRRNTVWLCYEVKTKGPSRPPLDAKIFRGQVYSELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKVTLTIF VARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNNNEPWVRGRHETYLCYEVERMHNDTWVLLNQR RGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYRRQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ FQPWDGLDEHSQDLSGRLRAILQNQEN	APOBEC3G D316R_D317R [Homo sapiens]
60	MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIY DDQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ	APOBEC3G [Homo sapiens]
61	MDPPTFTFNFNNEPWVRGRHETYLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIY RRQGRCQEGLRTLAEAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQ	APOBEC3G D120R_D121R [Homo sapiens]
	adenosine deaminases	
62	EACKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQEERFNHQAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE	TadA [Bacillus subtilis]
63	MTQDELYMKEAIKEAKKAEEKGEVPIGAVLVINGEIIARAHNLRETEQRSIAHAEMLVIDEACKALGTWRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFGAFDPKGGCSGTLMNLLQE ERFNHQAEVVSGVLEEECGGMLSAFFRELRKKKKAARKNLSE	TadA [Bacillus subtilis]
64	MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMD VLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD	TadA, catalytically inactive (E59A) [E. coli]
65	MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMAIRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMD VLHYPGMNHRVETEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD	TadA 7.10 [Synthetic]
66	MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGWRNAKTGAAGSLMD VLHYPGMNHRVETEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD	TadA 7.10 (V106W) [Synthetic]
67	MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRWAKTGAAGSLMD VLHYPGMNHRVETEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD	TadA 7.10 (N108W) [Synthetic]
68	MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRQAKTGAAGSLMD VLHYPGMNHRVETEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD	TadA 7.10 (N108Q) [Synthetic]
69	MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGFRNAKTGAAGSLMD VLHYPGMNHRVETEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD	TadA 7.10 (V106F) [Synthetic]
70	MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGQRNAKTGAAGSLMD VLHYPGMNHRVETEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD	tadA 7.10 (V106Q) [Synthetic]
71	MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGMRNAKTGAAGSLMD VLHYPGMNHRVETEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD	tadA 7.10 (V106M) [Synthetic]
72	MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNFAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMD VLHYPGMNHRVETEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD	tadA 7.10 (R47F) [Synthetic]
73	MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNWAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMD VLHYPGMNHRVETEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD	tadA 7.10 (R47W) [Synthetic]
74	MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNQAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMD VLHYPGMNHRVETEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD	tadA 7.10 (R47Q) [Synthetic]
75	MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNMAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMD VLHYPGMNHRVETEGILADECAALLCYFFRMPRQVFNAQKKAQSSTD	tadA 7.10 (R47M) [Synthetic]
76	MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAQIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMD VLHHPGMNHRVETEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD	TadA (E59Q) [Synthetic]

77	MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMD VLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD	TadA, N-terminal truncated [E. coli]
78	MRRAFITGVFFLSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARD AKTGAAGSLMDVLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTD	TadA [E. coli]
79	MGSHMTNDIYFMTLAIEEAKKAAQLGEVPIGAIITKDDEVIARAHNLRETLQQPTAHAEHIAIERAAKVLGSWRLEGCTLYVTLEPCVMCAGTIVMSRIPRVVYGADDPKGGCSGSLMN LLQQSNFNHRAIVDKGVLKEACSTLLTTFFKNLRANKKSTN	TadA [Staphylococcus aureus]
80	MPPAFITGVTSLSDVELDHEYWMRHALTLAKRAWDEREVPVGAVLVHNHRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVLQNYRLLDTTLYVTLEPCVMCAGAMVHSRIGRVVFGARD AKTGAAGSLIDVLHHPGMNHRVEIIEGVLRDECATLLSDFFRMRRQEIKALKKADRAEGAGPAV	TadA [S. typhimurium]
81	MDEYWMQVAMQMAEKAEAAGEVPVGAVLVKDGQQIATGYNLSISQHDPTAHAEILCLRSAGKKLENYRLLDATLYITLEPCAMCAGAMVHSRIARVVYGARDEKTGAAGTVVNLLQHPA FNHQVEVTSGVLAEACSAQLSRFFKRRRDEKKALKLAQRAQQGIE	TadA [S. putrefaciens]
82	MDAAKVRSEFDEKMMRYALELADKAEALGEIPVGAVLVDDARNIIGEGWNLSIVQSDPTAHAEIIALRNGAKNIQNYRLLNSTLYVTLEPCTMCAGAILHSRIKRLVFGASDYKTGAIG SRFHFFDDYKMNHTLEITSGVLAEECSQKLSTFFQKRREEKKIEKALLKSLSDK	TadA [H. influenzae]
83	MRTDESEDQDHRMMRLALDAARAAAEAGETPVGAVILDPSTGEVIATAGNGPIAAHDPTAHAEIAAMRAAAAKLGNYRLTDLTLVVTLEPCAMCAGAISHARIGRVVFGADDPKGGAVV HGPKFFAQPTCHWRPEVTGGVLADESADLLRGFFRARRKAKI	TadA [Caulobacter crescentus]
84	MSSLKKTPIRDDAYWMGKAIREAAKAAARDEVPIGAVIVRDGAVIGRGHNLREGSNDPSAHAEMIAIRQAARRSANWRLTGATLYVTLEPCLMCMGAIILARLERVVFGCYDPKGGAAG SLYDLSADPRLNHQVRLSPGVCQEECGTMLSDFFRDLRRRKKAKATPALFIDERKVPPEP	TadA [G. sulfurreducens]
85	GAGGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTT TGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATTTCGACCTGGACATGCTGATTAACTCTAGATAG	TAD [Synthetic]
86	EASGSGRADALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLINSR	TAD [Synthetic]
87	TCGCCAGGGATCCGTCGACTTGACGCGTTGATATCAACAAGTTTGTACAAAAAAGCAGGCTACAAAGAGGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATAT GCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATTTCGACCTGGACATGC TGATTAACTCTAGAAGTTCCGGATCTCCGAAAAAGAAACGCAAAGTTGGTAGCCAGTACCTGCCCGACACCGACGACCGGCACCGGATCGAGGAAAAGCGGAAGCGGACCTACGAGACA TTCAAGAGCATCATGAAGAAGTCCCCCTTCAGCGGCCCCACCGACCCTAGACCTCCACCTAGAAGAATCGCCGTGCCCAGCAGATCCAGCGCCAGCGTGCCAAAACCTGCCCCCCAGCC TTACCCCTTCACCAGCAGCCTGAGCACCATCAACTACGACGAGTTCCCTACCATGGTGTTCCCCAGCGGCCAGATCTCTCAGGCCTCTGCTCTGGCTCCAGCCCCTCCTCAGGTGCTGC CTCAGGCTCCTGCTCCTGCACCAGCTCCAGCCATGGTGTCTGCACTGGCTCAGGCACCAGCACCCGTGCCTGTGCTGGCTCCTGGACCTCCACAGGCTGTGGCTCCACCAGCCCCTAAA CCTACACAGGCCGGCGAGGGCACACTGTCTGAAGCTCTGCTGCAGCTGCAGTTCGACGACGAGGATCTGGGAGCCCTGCTGGGAAACAGCACCGATCCTGCCGTGTTCACCGACCTGGC CAGCGTGGACAACAGCGAGTTCCAGCAGCTGCTGAACCAGGGCATCCCTGTGGCCCCTCACACCACCGAGCCCATGCTGATGGAATACCCCGAGGCCATCACCCGGCTCGTGACAGGCG CTCAGAGGCCTCCTGATCCAGCTCCTGCCCCTCTGGGAGCACCAGGCCTGCCTAATGGACTGCTGTCTGGCGACGAGGACTTCAGCTCTATCGCCGATATGGATTTCTCAGCCTTGCTG GGCTCTGGCAGCGGCAGCCGGGATTCCAGGGAAGGGATGTTTTTGCCGAAGCCTGAGGCCGGCTCCGCTATTAGTGACGTGTTTGAGGGCCGCGAGGTGTGCCAGCCAAAACGAATCCG GCCATTTCATCCTCCAGGAAGTCCATGGGCCAACCGCCCACTCCCCGCCAGCCTCGCACCAACACCAACCGGTCCAGTACATGAGCCAGTCGGGTCACTGACCCCGGCACCAGTCCCTC AGCCACTGGATCCAGCGCCCGCAGTGACTCCCGAGGCCAGTCACCTGTTGGAGGATCCCGATGAAGAGACGAGCCAGGCTGTCAAAGCCCTTCGGGAGATGGCCGATACTGTGATTCCC CAGAAGGAAGAGGCTGCAATCTGTGGCCAAATGGACCTTTCCCATCCGCCCCCAAGGGGCCATCTGGATGAGCTGACAACCACACTTGAGTCCATGACCGAGGATCTGAACCTGGACTC ACCCCTGACCCCGGAATTGAACGAGATTCTGGATACCTTCCTGAACGACGAGTGCCTCTTGCATGCCATGCATATCAGCACAGGACTGTCCATCTTCGACACATCTCTGTTTTGA	TAD [Synthetic] VPR compises a VP64-SV40-P65-RTA
88	SPGIRRLDALISTSLYKKAGYKEASGSGRADALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLGSDALDDFDLDMLINSRSSGSPKKKRKVGSQYLPDTDDRHRIEEKRKRTYET FKSIMKKSPFSGPTDPRPPPRRIAVPSRSSASVPKPAPQPYPFTSSLSTINYDEFPTMVFPSGQISQASALAPAPPQVLPQAPAPAPAPAMVSALAQAPAPVPVLAPGPPQAVAPPAPK PTQAGEGTLSEALLQLQFDDEDLGALLGNSTDPAVFTDLASVDNSEFQQLLNQGIPVAPHTTEPMLMEYPEAITRLVTGAQRPPDPAPAPLGAPGLPNGLLSGDEDFSSIADMDFSALL GSGSGSRDSREGMFLPKPEAGSAISDVFEGREVCQPKRIRPFHPPGSPWANRPLPASLAPTPTGPVHEPVGSLTPAPVPQPLDPAPAVTPEASHLLEDPDEETSQAVKALREMADTVIP QKEEAAICGQMDLSHPPPRGHLDELTTTLESMTEDLNLDSPLTPELNEILDTFLNDECLLHAMHISTGLSIFDTSLF	TAD [Synthetic] VPR comprises a VP64-SV40-P65-RTA
	Linkers	
89	SGSETPGTSESATPES	XTEN linker [Synthetic]
90	SGGS	linker [Synthetic]
91	(SGGS)n	linker [Synthetic]
92	(SGGS)n	linker [Synthetic]
93	(GGGS)n	linker [Synthetic]
94	GGGGS	linker [Synthetic]
95	(GGGGS)n	linker [Synthetic96]
96	(SGGS)2-SGSETPGTSESATPES-(SGGS)2	linker [Synthetic]

97	(G)n	linker [Synthetic]
98	(SGGS)n-SGSETPGTSESATPES-(SGGS)n	linker [Synthetic]
99	(EAAAK)n	linker [Synthetic]
100	(EAAAK)n	linker [Synthetic]
101	(GGS)n	linker [Synthetic]
102	(GGS)n	linker [Synthetic]
103	SGGS(GGS)n	linker [Synthetic]
104	SGGSSGGSSGGSSGGSSGGSSGGSSGGSSGGSSGGSSETPGTSESATPESSGGSSGGSSGGSSGGSSGGSSGGSSGGSSGGSSGGSSGG SSGGSSGGSSGGSSGGS	linker [Synthetic]
105	SGGSSGGSSGGSSGGSSGGSSGGSSGGSSGGSSETPGTSESATPESSGGSSGGSSGGSSGGSSGGSSGGSSGGS	linker [Synthetic]
106	SGGSSGGSSGGSETPGTSESATPESSGGSSGGSSGGSSGGS	linker [Synthetic]
107	SGGSSGGSSGGSETPGTSESATPESSGGSSGGSSGGSSETPGTSESATPESSGGSSGGS	linker [Synthetic]
108	PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATS	linker [Synthetic]
109	SGGSSGGSSGGSETPGTSESATPESSGGSSGGS	linker [Synthetic]
110	GGSGGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGGSGGS	linker [Synthetic]
111	SGGSSGGSSGGSETPGTSESATPES	linker [Synthetic]
112	GGGGS	C-terminal linker [Synthetic]
	Nuclear localization signals (NLS)	
113	PKKKRKV	NLS [Synthetic]
114	MDSLLMNRRKFLYQFKNVRWAKGRRETYLC	NLS [Synthetic]
	uracil glycosylase inhibitors (UGI)	
115	MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKML	UNGI_BPPB2 [Bacillus phage PBS2]
116	KGHLTRLGLEFFDQPAVPLARAFLGQVLVRRLPNGTELRGRIVETQAYLGPEDEAAHSRGGRQTPRNRGMFMKPGTLYVYIIYGMYFCMNISSQGDGACVLLRALEPLEGLETMRQLRS TLRKGTASRVLKDPRELCSGPSKLCQAIAINKSFDQRDIAQDEAVWLERGPLEPSEPAVVAAARVGVGHAGEWARKPLRFYVRGSPWVSVVDRVAEQDTQA	Truncated AAG nuclease (E125Q) [H. sapiens]
117	DIASLRAQQIEIASSVIREDRLDKDPPDLIAGAAVGFEQGGEVTRAAMVLLKYPSLELVEYKVARIATTMPYIPGFLSFREYPALIAAWEMLSQRPDLVFVDGHGISHPRRLGVASHFG LLVDVPTIGVAKKRLCGKFEPLSSEPGALAPLMDKGEQLAWVWRSKARCNPLFIATGHRVSVDSALAWVQRCMKGYRLPEPTRWADAVASERPAFVRYTANQP	EndoV nuclease (D35A) [Synthetic]
118	MASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRDKQTGETKEKTEWHRVVLFGKLAEVAGEYLRKGSQVYIEGALQTRKWTDQAGVEKYTTEVVVNVGGTMQMLGGRSQG GGASAGGQNGGSNNGWGQPQQPQGGNQFSGGAQQQARPQQQPQQNNAPANNEPPIDFDDDIP	SSB [Erwinia tasmaniensis]
119	MAGAQDFVPHTADIAEIAAAAGECRGCGLYRDATQAVFGAGGRSARIMMIGEQPGDKEDIAGLPFVGPAGRLLDRALEAADIDRDALYVTNAVKHFKFTRAAGGKRRIHKTPSRTEVVA CRPWLIAEMTSVEPDVVVLLGATAAKALLGNDFRVTQHRGEVLHVDDVPGDPALVATVHPSSLLRGPKEERESAFAGLVDDLRVAADVRP	UdgX [Synthetic]
120	MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAAIPAKKAPAGQEEPGTPPSSPLSAEQLDRIQRNKAAALLRIAARNVPVGFGESWKKHLSGEFGKPYFIKLMGFVA EERKHYTVYPPPHQVFTWTQMCDIKDVKVVILGQEPYHGPNQAHGLCFSVQRPVPPPPSLENIYKELSTDIEDFVHPGHGDLSGWAKQGVLLLNAVLTVRAHQANSHKERGWEQFTDAV VSWLNQNSNGLVFLLWGSYAQKKGSAIDRKRHHVLQTAHPSPLSVYRGFFGCRHFSKTNELLQKSGKKPIDWKEL	UDG [Synthetic]

	Gam protein	
121	AKPAKRIKSAAAAYVPQNRDAVITDIKRIGDLQREASRLETEMNDAIAELTEKFAARIAPIKTDIETLSKGVQGWCEANRDELTNGGKVKTANLVTGDVSWRVRPPSVSIRGMDAVMET LERLGLQRFIRTKQEINKEAILLEPKAVAGVAGITVKSGIEDFSIIPFEQEAGI	Gam [bacteriophage Mu]
	Base editors	
122	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTS ESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE KKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIF FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL YLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH	xCas9(3.7)-BE3 [Synthetic]
123	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTS ESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE KKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET ITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL YLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH	xCas9(3.6)-BE3 [Synthetic]
124	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRTHSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSETPGTS ESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGE KKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL YLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH	rAPOBEC1-XTEN-Cas9n-UGI-NLS [Synthetic]
125	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGSSSGGSS GSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYYVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSF FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL SKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLL KALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP LARGNSRFAWMTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL YLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH	xCas9 3.7-BE4 [Synthetic]

	VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEV EEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENIKIMSLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENV MLLTSDAPEYKPWALVIQDSNGENIKIMSLSGGSGPKKKRK	BE4-Gam [Synthetic]
126	MAKPAKRIKSAAAAVVPQNRDAVITDIKRIGDLQREASRLETEMNDAIAEITEKFAARIAPIKTDIETLSKGVQGWCEANRDELTNGGKVKTANLVTGDVSWRVRPPSVSIRGMDAVME TLERLIGLQRFIRTKQEINKEAILLEPKAVAGVAGITVKSGIEDFSIIPPEQEAGISGSETPGTSESATPESSSESTGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHS IWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWP RYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSGGSGGSGETPGTSESATPESSGGSGSGGSDKYSIGLAIGTNSVGWAVITDEYKV PSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLP NEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLT LTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRS DKNRGKSDNVPSEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREIN NYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKT EVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENIKIMSLSGGSGPKKKRK	xCas9(3.7)-ABE [Synthetic]
127	MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHI30YPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ VLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGGSGGSGETPGTSESATPESSGGSGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRV IGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ NAQKKAQSSTDSGGSSGGSGGSGETPGTSESATPESSGGSGSGGSDKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKN RICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS PLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLK DNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDLLF KTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRD KQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKEL GSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRK KRKV	xCas9(3.6)-ABE [Synthetic]
128	MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMD VLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGGSGGSGETPGTSESATPESSGGSGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRV IGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ KKAQSSTDSGGSSGGSGGSGETPGTSESATPESSGGSGSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC YLQEIFSNEMAKVDDSFFHRLEESFLVGEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLAKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNR EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTN RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEI GKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDEGADKRTADGSEFESPKKKRK KRKV	ABE7.10[Synthetic]
129	MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMD VLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGGSGGSGETPGTSESATPESSGGSGSGGSSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRV IGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ KKAQSSTDSGGSSGGSGGSGETPGTSESATPESSGGSGSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRIC YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNR IGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ KKAQSSTDSGGSSGGSGGSGETPGTSESATPESSGGSGSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRIC YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNR	

	EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDK AGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV	BE4 [Synthetic]
130	MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARL YHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNIILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLKSGGSSGGSS GSETPGTSESATPESSGGSSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSF FHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARL SKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLL KALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGP LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL YLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKVDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKVDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSPKKKRKV	dCas9-VPR [Synthetic]
131	ATGGACAAGAAGTACTCCATTGGGCTCGATATCGGCACAAACAGCGTCGGCTGGGCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCG CCACAGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAGACGGCCGAAGCCACGCGGCTCAAAAGAACAGCACGGCCGCAGATATACCCGCAGAAGAATCGGATCT GCTACCTGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGGTGGAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTT GGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGTACCCCACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTGCGGCTTATCTATCTGGCGCTGGC GCATATGATCAAATTTCGGGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCAGCTGGTTCAGACTTACAATCAGCTTTTCGAAGAGA ACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAAATCTCGGCGGCTCGAAAACCTCATCGCCCAGCTCCCGGGGGAGAAGAAGAACGGCCTGTTT GGTAATCTCATTGCCCTGTCCCTGGGCCTGACCCCTAACTTCAAGTCCAACTTCGACCTGGCCGAGGATGCCAAGCTCCAGCTGAGCAAGGACACCTACGACGACGACCTCGACAATCT GCTGGCCCAGATCGGCGACCAGTACGCGGACCTTTTTTGGCGCCAAAAACCTGTCCGACGCCATTCTGCTGAGTGATATCCTGCGAGTAAACACTGAAATAACCAAGGCTCCCCTGTCA GCGCTAGTATGATCAAGCGCTACGATGAGCACCACCAAGACTTGACCTTGCTGAAAGCCCTTGTCAGACAGCAACTGCCTGAGAAGTATAAAGAAATTTTCTTCGATCAGTCTAAAAAT GGCTACGCCGGATACATTGACGGCGGAGCCAGCCAAGAGGAATTTTACAAATTTATCAAGCCCATCTTGGAAAAAATGGACGGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGA TCTGTTGCGCAAGCAGCGCACTTTCGACAACGGCAGCATTCCCCACCAGATTCACCTGGGCGAGCTGCACGCTATCCTCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAGATAACGA GGGAAAAGATTGAGAAAATCCTCACATTTCGGATACCCTACTATTGTGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAAATCCAGAAGAGACCATCACTCCCTGG AACTTCGAGGAAAGTCGTGGATAAGGGCGCTTCTGCCCAGTCATCAAGAAGCGATTGACTTAATCTACCTGGCCCTGATTGGAGCCCCTAAAGCCAACTTTAAGTCTCACTTCACGCTGG GTACTTCACAGTTTATAACGACCTCACCAAGGTCAAATACGTCACCGAGGGAATGAGAAAGCCAGCATTCCTGTCTGGAGATCAGAAGAAAGCTATTGTGGACCTCCTCTTCAAGACGA ACCGGAAAGTTACCGTGAAACAGCTCAAAGAAGATTATTTCAAAAAGATTGAATGTTTCGACTCTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCAC GATCTCCATCTGCATGAGATTAAAAATTTATTACAAATTTTTAGACAAGAATCCTATTGTCCTCACCCTTACGTTGTTTGAGAATAGAGGGAGAATGCCTGAGCTGGAAAACGTATCCT CTTGAAAACTTACGTCTCATCTCTTCGACAATCAAAGTCATGAAGCAGCTCAAGAGGCGCCGATATACAGGATGGGGGCGCCTGTCTAAGAAAATCGATGGAATTATAAAGAGGTC GTGGAAAAGACAATCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAGCTGATCCATGATGATAGCCTTACCTTTAAGGAAGATATCCAGAAAGCACAAGTTTCT GGCCAAGGGGGACAGCCTTCACGAGCACATCGCTAATCTTGCAGGCAGCCCTGCAGCTGAAGTCGATAAGGTTGTGGATGAACTGGTGAAAGTAATGGGGAGGCA TAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCAAACCACTCAGAAGGGACAGAAGAACAGCCGAGAGAGAATGAAGCGAATCGAAGAAGGTATCAAAGAACTGGGGAGCCAA AAATCCTTAAGGAACACCCAGTTGAAAACACAGCTCCAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGGCGGGATATGTACGTGGATCAGGAACTGGACATCAATCGGCTC TCCGACTACGACGTGGATCATATCGTGCCCCAGTCTTTCCTTAAAGATGATTCTATTGATAATAAAGTGTTGACAAGATCCGGACAAGAACGTGATCACAAGTCGGAAAAATGTATCCT CCCGAAAAGGGAACAGCGACAAGCTGATCGCGAGCACCCAAGAAATACGGCGGATTCTCCAAGCCGAGTCGTCTACAAGTTGACTGGTTTGTCCACCAAAGAATGGAGAAGTATCCT AAGGGAAGTCTAAAAAACTCAAAGGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCTCGAAAAAAACCCCATCGACTTTCTCGAGGCGAAAGGATATAAAGAGGTC AAAAAAGACCTCATCATTAAGCTTCCCAAGTATACTCTCTTTGAGCTTGAAAACGGCCGCAAGAGAATGCTGGCTAGTGCGGGCGAACTGCAGAAGGGTAACGAGCTGGCACTGCCCTCC TAAATACGTTAATTTCTTATATCTGGCCAGCCACTATGAAAAGCTCAAGGGTCTCCCGAAGATAATGAGCAGAAGCAGCTGTTCGTGGAACAACACACACTACCTTGATGAGATCA TCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGAGAAACAAATT ATCCACTTGTTTACCTTGACCAACCTGGGCGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATAGACCGGAAGCGGTACACCAGCACCAAAGAGGTCCTGGACGCCACACTCATCCA TCAGTCCATTACGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACCCCCAAGAAGAAGAGGAAGGTGTCGCCGACGGTTCGATCCCCCAAGCCGTTGATATCAACA AGTTTGTACAAAAAAGCAGGCTACAAGGAGGCCAGCGGTCCGGACGGGCTGACGCATTTGATCTGGAATATGCTGGAAGTGACGCCCTCGATGATTTTGACCTTGACATG GCTTGGTTGATGATCCTCGGCAGTGACCTGGACATGCTGATTAACTCTGACATTCAACTCGGATCTCGGAAAGTTCCCCCCTTCAGCGGCCCGC GCAAAGTTGGTAGCCAGTACCTCCCGACACCGGACGAGCACCAGGAAAAGCGGAAGCGCACCTACGAGACATCTAGAAGAAGTCCCCCTTCAGCGGCCCC ACCGACCCTAGACCTCCACCTAGAAGAATCGCCGTGCCCAGCAGATCCAGCGCCAGCGT""""""""CCCCCCAGCCTTACCCCTTCACCAGCAGCCTGAGCACCATCAACTACGA	

	CGAGTTCCCTACCATGGTGTTCCCCAGCGGCCCAGATCTCTCAGGCCTCTGCTCTGGCTCCAGCCCCTCCTCAGGTGCTGCCTCAGGCTCCTGCTCCTGCACCAGCCTCCAGCCATGGTGT CTGCACTGGGCTCAGGCACCCAGCACCCGTGCCTGTGCTGGACCTCCACACAGGCTGTGGCTCCCACCAGCCCCTAAACCTACACAGGCCGGCGAGGGCACACTGTCTGAAGCTCTG CTGCAGCTGCAGTTCGACGACGAGGATCTGGGAGCCCTGCTGGGAAACAGCACCGATCCTGCCGTGTTCACCGACCTGGCCAGCGTGGACAACACAGCGAGTTCCAGCAGCTGCTGAACCA GGGCATCCCTGTGGCCCCTCACACCACCGAGCCCATCTGATGGAATACCCCGAGGCCATCACCCGGCTCTGATCCAGCTCCTCCCCCTCTGGGAG CACCAGGCCTGCCTAATGGACTGCTGTCTGGCGACGAGGACTTCAGCTCTATCGCCGATATGGATTTCTCAGCCTTGCTGGCTCTGGCAGCGGCATTCCAGGGAAGGGATG TTTTTGCCGAAGCCTGAGGCCGGCTCCGCTATTAGTGACGTGTTTGAGGGCCGCGAGGTGTGCCAGCCAAAACGAATCCGGCCATTTCATCCTCCAGGAAGTCCATGGCCAACCGCCC ACTCCCCGCCAGCCTCGCACCAACACCAACCGGTCACTGACCCCGGCACCATCAGCCACTGGATCCAGCGCCCGCAGTGACTCCCGAGGCCA GTCACCTGTTGGAGGATCCCGATGAAGAGACCGAGCCTGTCAAAGCCCTTCGGGAGATGGCCGATACTGTGATTCCCCAGAAGGAAGAGGCTGCAATCTGTGGCCAAATGGACCTT TCCCATCCGCCCCCAAGGGGCCATCTGGATGAGCTGACAACCACACACTTGAGTCCATGACCGAGGATCTCGACAACCCTGAACCTCAACCCTGAACCCGGAATTGAACGAGATTCTGGATACCTT CCTGAACGACGAGTGCCTCTTGCATGCCCATGCACATCTTCGACACATCTCTGTTT	xCas9(3.6)-ABE [Synthetic]
132	MSEVEFSHEYWMRHALTLAKRAWDEREVPVGAVLVHNNRVIGEGWNRPIGRHDPTAHAEIMAIRQGGLVMQNYRLIDATLYVTLEPCVMCAGAMIHSRIGRVVFGARDAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ VLHHPGMNHRVEITEGILADECAALLSDFFRMRRQEIKAQKKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALITLAKRARDEREVPVGAVLVLNNRV IGEGWNRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVVFGVRNAKTGAAGSLMDVLHYPGMNHRVEITEGILADECAALLCYFFRMPRQVFNAQ KKAQSSTDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSGSETPGTSESATPESSGGSSGGSSEVEFSHEYWMRHALITLAKRARDEREVPVGAVLVLNNRV YLQEIFSNEMAKVDDSFFHRLEESFLVGEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLAKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQLHLGELHAILRRQEDFYPFLKDNR EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTN RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH IARGNSRFAWMTRKSEETITPWNFEKVVDKGASAQQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL YLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTE ITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITI MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANL DKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEV EEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKPWALVIQDSNGENKIKMLSGGSGGSGGSTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENV MLITSDAPEYKPWALVIQDSNGENKIKMLSGGSPKKKRK	
	Guide RNA	
134	MMMMMMMNNNNNNNNNNNNXGG	Cas9 Target /guide RNA [Synthetic]
135	NNNNNNNNNNNNNXGG	Cas9 Target /guide RNA [Synthetic]
136	MMMMMMMNNNNNNNNNNNNXGG	Cas9 Target [Synthetic]

137	NNNNNNNNNNNNXGG	Cas9 Target [Synthetic]
138	MMMMMMMMMNNNNNNNNNNXXAGAAW	Cas9 Target [Synthetic]
139	NNNNNNNNNNNNNXXAGAAW	Cas9 Target [Synthetic]
140	MMMMMMMMMNNNNNNNNNNXXAGAAW	Cas9 Target [Synthetic]
141	NNNNNNNNNNNNXXAGAAW	Cas9 Target [Synthetic]
142	MMMMMMMMMNNNNNNNNNNXGGXG	Cas9 Target [Synthetic]
143	NNNNNNNNNNNNNXGGXG	Cas9 Target [Synthetic]
144	MMMMMMMMMNNNNNNNNNNXGGXG	Cas9 Target [Synthetic]
145	NNNNNNNNNNNNNXGGXG	Cas9 Target [Synthetic]

146	NNNNNNNNgttttttgtactctcaagatttaGAAAtaaatcttgcagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcatttatggcagggtgtttttcgttatttaa TTTTTT	Transcription Terminator [Synthetic]
147	NNNNNNNNNNNNNNNNNNNNgttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcatttatggcagggtgtttttcgttatttaaTTTT TT	Transcription Terminator [Synthetic]
148	NNNNNNNNNNNNNNNNNNNNgttttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccgaaatcaacaccctgtcatttatggcagggtgtTTTTTT	Transcription Terminator [Synthetic]
149	NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcTTTTTT	Transcription Terminator [Synthetic]
150	NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAATAGcaagttaaaataaggctagtccgttatcaacttgaaaaagtgTTTTTTT	Transcription Terminator [Synthetic]
151	NNNNNNNNNNNNNNNNNNNNgttttagagctagAAATAGcaagttaaaataaggctagtccgttatcaTTTTTTTT	Transcription Terminator [Synthetic]
152	guuuuagagcuagaaauagcaaguuaaaauaaggcuaguccguuaucaacuugaaaaaguggcaccgagucggugcuuuuu	guide RNA [Synthetic]
	Table 4 Additional Cas9 proteins	
153	MNKPYSIGXDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASMIQRYDEHREDIKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGNSRFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTEVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	Streptococcus agalactiae]
154	GAASMDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHER HPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFD QSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEET ITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQK AQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD	4UN5_B [Synthetic]

	INRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDK LIREVVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNG ETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKG YKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQ AENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	5AXW_A [Staphylococcus Aureus]
155	GSHMKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRR GVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMG HCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKI LTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPND IIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEEASKKGNRTPFQYLL SSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANA DFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMY HHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKAYEE AKKLKKISNQAEFIASFYNNDLIKINGELIYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	AKA60242.1 [Synthetic]
156	MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLIKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	AKS40380.1[Synthetic]
157	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLIKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGSPKKKRKVSSGGAAG	AKS40380.1[Synthetic]
158	MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRIC YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNR REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKA TAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEK GKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRPAATKKAGQAKKKK	AII16583.1 [Synthetic]
159	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLIKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGSPKKKRKVSSGGAAG	AKQ21048.1 [Synthetic]

160	MLKKDYSIGLDIGTNSVGHAVVTDDYKVPTKKMKVFGDTSKKTIKKNMLGVLLFNEGQTAADTRLKRGARRRYTRRKNRLRYLQEIFAPALAKVDPNFFYRLEESSLVAEDKKYDVYPI FGKREEELLYHDTHKTIYHLRSELANNDRPADLRLVYLALAHIIKYRGNFLLEGEIDLRNTDINKVFAEFSETLNENSDENLGKLDVADIFKDNTFSKTKKSEELLKLSGAKKNQLAHQ LFKMMVGNMGSFKKVLGTDEEHKLSFGKDTYEDDLNDLLAEAGDQYLDIFVAAKKVYDAAILASILDVKDTQTKTVFSQAMIERYEEHQKDLIELKRVFKKYLPEKCHDFFSEPKISGY AGYIDGKVSEEDFYKYTKTLKGTIPTETEEILQKIDANNYLRKQRTFDNGAIPHQVHLKELVAIVENQGKYYPFLRENKDKFEKILNFRIPYYVGPLARGNSKFAWLTRAGEGKITPYNF DEMIDKETSAEDFIKRMTINDLYLPTEPVLPKHSLLYERYTIFNELAGVRYVTENGEAKYFDAQTKRSIFELFKLDRKVSEKMVIKHLKVVMPAIRIQALKGLDNGKFNASYGTYKDLV DMGVAPELLNDEVNSEKWEDIIKTLTIFEGRKLIKRRLENYRDFIGEDILRKLSRKKYTGWGRLSAKLLDGIYDKKTHKTILDCLMTEDYSQNFMQLINDDTYSFKETIKNAQVIEKEE TLAKTVQELPGSPAIKKGILQSLEIVDEIIKVMGYKPKSIVVEMARETQKTHGTRKREDVQQIVKNLKDANELPKKLPSNAELSDERKYLYCLQNGRDMYTGAPLDYDHIQFYDVDHI IPQSFLKDDSIENKVLTIKKENVRKTNGLPSEAVIQKMGSFWKKLLDAGAMTNKKYDNLRRNLHGGLNEKLKERFIERQLVETRQITKYVAQLLDQRLNYDGNGVELDEKIAIVTLKAQ LASQFRSEFKLRKVRALNNLHHAHDAYLNAVVANLIMAKYPELEPEFVYGKYRKTKFKGLGKATAKNTLYANVLYFLKENEVVPFWDKARDLPTIKRYLYRAQVNKVRKAERQTGGFSD EMLVPKSDSGKLLPRKEGLDPVKYGGYAKAVESPAYLITADEVKKGKTKKVLTVNIPITDSKKYEADPTAYLASRGYTVNTNSFILPKYSLLEDPEGRRYLASPKEFQKANELILPQ HLVELLYWVNAKDGEQKLEDHKAEFKELFDKIMEFADKYVVAPKNSEKIRRLYEENQDATPMELGKNFVELLRYTADGAASDFKFFGENIPRKRVNSAGSLLNGTLIYQSKTGLYETRI DLGKL	KDA45870.1 [Lactobacillus animalis]
161	MAPKKKRKVGIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEE SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRL ENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNS RFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVE ISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIH DDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKDPVDFLEAKGYKEVRKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE EKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD NNIVFISQSPTGIYTKKYKL	AKE81011.1 [Synthetic]
162	IVDYCIGLDLGTGSVGWAVVDMNHRLMKRNGKHLWGSRLFSNAETAANRRASRSIRRRYNKRRERIRLLQDMVEKDPTFFIRLEHTSFLDEDKAKYLGTDYKDNYNLFIDEDF NDYTYYHKYPTIYHLRKALCESTEKADPRLIYLALAHIIVKYRGNFLIEGQFKFNMDASNIEDKLSDIFTQFTSFNNIPYEDDEKKNLEILEIILKKPLSKKAKVDEVMTLIAPEKDYKSAF KELVTGIAGNKMNVTKMILCEPIKQGDSEIKLKFSDSNYDDQFSEVEKDIGEYVEFVDALHNVYSWVELQTIMGATHTDNASISEAMVSRYNKHHDDLKLLKDCIKNNVPNKYFDMFRN DSEKSKGYYNYINRPSKAPVDEFYKFVVKKCIEKVDTPEAKQILNDIELENFLLKQNSRTNGSVPYQMQLDEMIKIIDNQAEYYPFLKEKREQLLSIILTFRIPYYVGPLNETSEHAWIKR LEGKENQRILPWNYQDIVDVDATAEGFIKRMRSYCTYFPDEEVLPKNSLIVSKYEVYNELNKIRVDDKLLEVDVKNDIYNELFMKNKTVTEKKLKNWLVNNQCCSKDAEIKGFQKENQF STSLTPWIDFTNIFGKIDQSNFDLIENIIYDLTVFEDKEKIMKRRLKKYALPDDKVKQILKLKYKDWSRLSKKLLDGIVADNRFGSSVTVLDVLEMSRINLMEIINDKDLGYAQMIEEA TSCPEDGKFTYEEVERLAGSPALKRGIWQSLQIVEEITKVMKCRPKYIYIEFERSEEAKERTESKIKKLENVYKDLDEQTKKEVKSVLEELKGFDNTKKISSDSLFLYFTQLGKCMYSG KKLDIDSLDKYQLDHIVPQSLVKDDSFDNRVLVVPSENQRKLDDLVVPFDIRDKMYRFWKLLFDHELISPKKFYSLIKTEYTERDBERFINRQLVETRQITKNVTQILEDHYSTTKVAA IRANLSHEFRVKNHIYKNRDINDYHHAHDAYIVALIGGFMRDRYPNMHDSKAVYSEYMKMFRKNKNDQKRWKDGFVINSMNNYPYEVDGKLIWNPDLINEIKKCFYYKDCYCTTKLDQKS GQLFNLITVLSNDAHADKGFSGLQYTIVAIEGQKKKGKKTELVKKISGVPLHLKAASINEKINIYIEEKEGLSDVRIIKDNIPVNQMIEMDGGEYLLTSPT EYVNARQLVLNEKQCALIADIYNAIYKQDYDNLDDIILMIQLYIELTNKMKVLYPAYRGIAEKFESMNENYVVISKEEKANIIKQMLIVMHRGPQNGNIVYDDFKISDRIGRLKTKNHNL NNIVFISQSPTGIYTKKYKL	WP_006506696.1 [Catenibacterium mitsuokai]
163	LSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SDILKEYPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVRVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDIRKMIAKSEQEIG KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKV EKGKSKKLKSVKELVGITIMERSSFEKDPVDFLEAKGYKEVRKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKCFDTTIGRNRYKSTKEVLDATLIHQSITGLYETRIDLSQLGGD	WP_033162887.1 [Sharpea azabuensis]
164	MAKNKDIRYSIGLDIGTNSVGWAVMDEHYELLKKGNHHMWGSRLFDAAEPAATRRASRSIRRRYNKRRERIRLLRDLLGDMVMEVDPTFFIRLLNVSFLDEEDKQKNLGNDYVKDNYNLF IEKDFNDKTIYYDKYPTIYHLRKELCENKEKADPRLIYLALAHHIVKYRGNFLIEGQSFTMDNSDIEERLNSAIEKFMSINEFDNRIVESDINSMIAVLSKIYQRSKRADDLLIKIMNPTKE EKAAYKEFTKALVGLKFNISKMIIAQEVKKGDTDIYLEFSNANYDSTIDELQSELGEYIEFIEMLHNIYSWVELQAILGATHTDNPSISAAMVERYEEHKKDLRVLKKVIREELPDKYN EVFRKDNRKLHNYLGYIKYPKDTPVEEFYKYIKGLLAKVDTDEARAELIERIDLEKFMLKQNSRTNGSIPYQMQKDEMIQIIDNQSVYYPFLQKENRDKLISILEFRIPYYFGPLNAHSEF AWIKKFEDKQKERILPWNYDQIVDIDATAEGFIERMNNTGTYFPDEPVMAKNSLITVSKFEVLNELNKIRINGKLIAVETKKELLISDLFMKNKTITDKKLKDWLVTHQIYDINEELKIEG YQKDLQFSTSLAPWIDFTKIFGEINASNYQLIEKIIYDISIFEDKKILKRRLKKVYQLDDLLVDKILKLNYTGWSRLSEKLLTGMTADNEFGSKATVLFVLENSNKNLMEIINDEKLGY KQIIEESNMQDIEGCPFKYDEVKKLAGSPAIKRGIWQALLVVREITKFMKHEPPSHIYIEFAREEQEKVRKESKIAKLQKIYENLNLQTKEDQQVYESLKKEDAKKRMETDALYLYYLQMG KSMYSGKPLDIDKLSTYQIDHILPQSLIKDDSFDNRVLVLPEENQWKLDDSTVPFEIRNRMIGFWQMLHENGLMSNKKFFSLITRTDFSDKDKERFINRQLVETRQITKNVAQILTDRY NTNIVTVRAELSHQFRERYKIYKNRDINDFHHAHDAYIACIVGQFMHQNFEHLDAKIIYGQYKKNYKKDVKHNNYGFILNSMMHLQSDIDTGEVMWDPAKIGKIKSCFYYKDVYVTKK LEQNSGTLFNVTVLPNDAHSEKGITAATVPLNKYRADVHKYGGFGNVQSIIVAIEGKKKKGKKKLIDVRKLTSIPLHLIKNAPVEEQLSYIASPEHEDLIDVRIVKELILKNQLIEIDGGLY YVTSPTEYVTARQLSLNEQSCKLISIEIAYQAAMLKKRYFDRAEDFKNIDVVEKCDVIKQILIIMHAGPMNGNIMYDDFKFTNRIGRFT HKNIDLNKTTFISTSVTGLFSKKYKL	WP_030126706.1 [Streptococcus pyogenes]
165	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHGIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPYQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKDPVDFLEAKGYKEVRKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKCFDTTIGRNRYKSTKEVLDATLIHQSITGLYETRIDLSQLGGD	WP_009880683.1 [Streptococcus pyogenes]

	DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEYPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV RKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	WP_011527619.1 [Streptococcus pyogenes]
166	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF	
	GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF	
	GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRLNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN	
	GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW	
	NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH	
	DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTVWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS	
	GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL	
	SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE	
	VKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMLAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE	
	IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV	
	RKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI	
	IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATFIHQSITGLYETRIDLSQLGGD	
167	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF	WP_002989955.1 - [Streptococcus multispecies]
	GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF	
	GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRLNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN	
	GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW	
	NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH	
	DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS	
	GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL	
	SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE	
	VKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMLAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE	
	IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV	
	RKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI	
	IHLFTLTNLGAPAAFKCFDTTIGRNRYKSIKEVLDATLIHQSITGLYETRIDLSQLGGD	
168	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRVTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF	WP_032464890.1 [Streptococcus pyogenes]
	GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF	
	GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRLNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN	
	GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNRKDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW	
	NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH	
	DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS	
	GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL	
	SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE	
	VKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMLAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE	
	IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKDPVDFLEAKGYKEV	
	RKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI	
	IHLFTLTNLGAPAAFKYFDTTIGRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
169	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEIAKVDDSFFHRLEESFLVEEDKKHERHPIF	WP_012560673.1 [Streptococcus pyogenes]
	GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLF	
	GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN	
	GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW	
	NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH	
	DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS	
	GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEYPVETTQLQNEKLYLYYLQNGRDMYVDQELDINRL	
	SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWKQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE	
	VRVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAVVGTALIKKYPKLESEFVYGDYKVYDIRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE	
	IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELVGITIMERSSFEKDPVDFLEAKGYKEV	
	RKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI	
	IHLFTLTNLGAPAAFKYFDTTIGRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
170	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEIAKVDDSFFHRLEESFLVEEDKKHERHPIF	WP_032460140.1 [Streptococcus pyogenes]
	GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLF	
	GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN	
	GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW	
	NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH	
	DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDLQKAQVS	
	GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEYPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL	

	SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWKQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VRVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDIRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELVGITIMERSSFEKDPVDFLEAKGYKEV RKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIGRNRYKSIKEVLDATLIHQSITGLYETRIDLSQLGGD	WP_032461047.1 [Streptococcus pyogenes]
171	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEIAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLIKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEYPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWKQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VRVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDIRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELVGITIMERSSFEKDPVDFLEAKGYKEV RKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIGRNRYKSIKEVLDATLIHQSITGLYETRIDLSQLGGD	WP_038434062.1 [Streptococcus pyogenes]
172	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENLINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLIKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEYPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWKQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV RKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIGRNRYKSIKEVLDATLIHQSITGLYETRIDLSQLGGD	WP_032462936.1 [Streptococcus pyogenes]
173	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLF GNLIALLLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLIKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEYPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWKQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VRVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV RKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	WP_038431314.1 [Streptococcus pyogenes]
174	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLIKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEYPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELVGITIMERSSFEKNPIDFLEAKGYKEV RKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	WP_038431314.1 [Streptococcus pyogenes]
175	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLIKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEYPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWKQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDPRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE	WP_031488318.1 [Streptococcus pyogenes]

	IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV RKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNFGAPAAFIYFDTTIGRNRYKSIKEVLDATLIHQSITGLYETRIDLSQLGGD	
176	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKVDLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPYQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNTSLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKKYANLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLINDDSLTFKEDIQKAQVS GQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVR KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAKNII HLFTLTNLGAPAAFKYFDTTIERNRYKSIKEVLDATLIHQSITGLYETRIDLSQLGGD	
177	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINANGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKKYANLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
178	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRLNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
179	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
180	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRLNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV	

	KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
181	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKASLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEYPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV RKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIGRNRYKSIKEVLDATLIHQSITGLYETRIDLSQLGGD	WP_030125963.1 [Streptococcus pyogenes]
182	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRIRYLQEIFSSEMSKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINKASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPEFLSGKQKEAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDKEMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLINDDSLTFKEDIQKAQVS GQGHSLHEQIANLAGSPAIKKGILQSVKVVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFIKDDSIDNKILTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAVVGTALIKKYTKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNEETGEI VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGALTNESIYARGSFDKLISRKHRFESSKYGGFGSPTVTYSVLVVAKSEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV VQIDKCIKLPKYSLFEFENGTRRMLASVMANNNSRGDLQKANEMFLPAKLVTLLYHAHKIESSKELEHEAYILDHYNDLYQLLSYIERFASLYVDVEKNISKVKELFSNIESYSISEIC SSVINLLTLTASGAPADFKFLGITTIPRKRYGSPQSTILSSTLIHQSITGLYETRIDLSQLGGD	WP_015017095.1 [Streptococcus dysgalactiae]
183	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRIRYLQEIFSSEMSKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDMDKLFIQLVQTYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPEFLSGKQKEAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDKEMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLINDDSLTFKEAIQKAQVS GQGHSLHEQIANLAGSPAIKKGILQSVKVVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNEETGEI VWNKGRDFATVRKVLSMPQVNIVKKTEVQTGALTNESIYARGSFDKLISRKHRFESSKYGGFGSPTVTYSVLVVAKSEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV VQIDKCIKLPKYSLFEFENGTRRMLASVMANNNSRGDLQKANEMFLPAKLVTLLYHAHKIESSKELEHEAYILDHYNDLYQLLSYIERFASLYVDVEKNISKVKELFSNIESYSISEIC SSVINLLTLTASGAPADFKFLGITTIPRKRYGSPQSTILSSTLIHQSITGLYETRIDLSQLGGD	WP_015057649.1 [Streptococcus dysgalactiae]
184	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRIRYLQEIFSSEMSKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDMDKLFIQLVQTYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPEFLSGKQKEAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDKEMIEERLKKYANLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLINDSSLTFKEAIQKAQVS GQGHSLHEQIANLAGSPAIKKGILQSVKVVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNEETGEI VQIDKCIKLPKYSLFEFENGTRRMLASVMANNNSRGDLQKANEMFLPAKLVTLLYHAHKIESSKELEHEAYILDHYNDLYQLLSYIERFASLYVDVEKNISKVKELFSNIESYSISEIC SSVINLLTLTASGAPADFKFLGITTIPRKRYGSPQSTILSSTLIHQSITGLYETRIDLSQLGGD	WP_015057649.1 [Streptococcus dysgalactiae]
185	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRIRYLQEIFSSEMSKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLAQSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDMDKLFIQLVQTYNQLFEENPINKASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPEFLSGKQKEAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVS GQGHSLHEQIANLAGSPAIKKGILQSVKVVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNEETGEI VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGALTNESIYARGSFDKLISRKHRFESSKYGGFGSPTVTYSVLVVAKCKVQDGKVKKIKTGKELIGITLLDKLVFEKNPLKFIEDKGYGN	WP_048327215.1 [Streptococcus dysgalactiae]

	VQIDKCIKLPKYSLFEFENGTRRMLASVMANNNSRGDLQKANEMFLPAKLVTLLYHAHKIESSKELEHEAYILDHYNDLYQLLSYIERFASLYVDVEKNISKVKELFSNIESYSISEIC SSVINLLTLTASGAPADFKFLGTTIPRKRYGSPQSILSSTLIHQSITGLYETRIDLSQLGGD	
186	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRIRYLQEIFSSEMSKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEKPINASGVDAKAILSARLSKSKRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGNQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKKYANLFDDKVMKQLKRRHYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLINDDSLTFKEAIQKAQVS GQGHSLHEQIANLAGSPAIKKGILQSVKVVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGALTNESIYARGSFDKLISRKHRFESSKYGGFGSPTVTYSVLVVAKSKVQDGKVKKITKTGKELIGITLLDKLVFEKNPLKFIEDKGYGN VQIDKCIKLPKYSLFEFENGTRRMLASVMANNNSRGDLQKANEMFLPAKLVTLLYHAHKIESSKELEHEAYILDHYNDLYQLLSYIERFASLYVDVEKNISKVKELFSNIESYSISEIC SSVINLLTLTASGAPADFKFLGTTIPRKRYGSPQSILSSTLIHQSITGLYETRIDLSQLGGD	WP_014612333.1 [Streptococcus dysgalactiae]
187	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYREIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKKYANLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK KDLIIKLPKYSLFELENGNGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENII HLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	WP_011054416.1 [Streptococcus pyogenes]
188	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKLKVLGNTDRHGIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKVDLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPYQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKVGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VRVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKDPIDFLEAKGYKEV RKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	WP_023080005.1 [Streptococcus pyogenes]
189	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKLKVLGNTDRHGIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKVDLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPYQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNTSLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDKEMIEERLKKYANLFDDKVMKQLKRRHYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLINDSSLTFKEAIQKAQVS GQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYHHADAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVR	WP_023610282.1 [Streptococcus pyogenes]

	KDLIIKLPKYSIFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAKNII HLFTLTNLGAPAAFKYFDTTIERNRYKSIKEVLDATLIHQSITGLYETRIDLSQLGGD	
191	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLIKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	WP_010922251.1 [Streptococcus pyogenes]
192	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLIKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	WP_010922251.1 [Streptococcus pyogenes]
193	MDQKYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLIGALLFDSGETAEATRLKRTARRYTRRRNRLRYLQEIFAEEMNKVDENFFQRLDDSFLVDEDKRGERHPIF GNIAAEVKYHDDFPTIYHLRKHLADISQKADLRLVYLALAHMIKFRGHFLIEGQLKAENTNVQALFKDFVEVTDKTVEESHLSEMTVDALSILITEKVSKSRRLENLIAHYPAEKKNTLF GNLIALSLGLQPNFKTNFQLSEDAKLQFSKDTYEEDLEGILGEIGDEYADLFSAKNLYDAILLSGILTVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKK GYAGYIENGVKQDEFYKYLKGILLQINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQEEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPW NFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSLLYETFTVYNELTKVKYVNEQGFAKFFDANMKQEIFDHVFKENRKVTKDKLINYLNKEFEBFRIVNLITGLDKENKAFNSSLGT YHDIRKILDKSFDDDKANEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQI IGDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQMTDKGRRNSQQRLKLLQDSIKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDINNLS QYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGNKRRIRNV KIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLKVYPQLEPEFVYGEYPKYNSYRSKSATEKFLFYSNILRFFKKEDIQTNEDGETAWNKEKHIKILRKVLSYPQV NIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWNPKKYGGFDSPVVAYSILVFADVEKGKSKKLKRKVQDMVGITIMEKKRRFEKHPVDFLEQKGYRNVRLEKIIKLPKYSLFELEN KRRRLIASARELQKGNELVIPQRFTTLLYHSYQIEKNYEPEHREYVEKHKDEFKELLEYISVFSRKVYLADNNLITKIEMLFSKNKDAEVSSLAKSFISLITFTAFGAPAAFNFFGENID RKRYTSVTECLNATLIHQSITGLYETRIDLSKLGED	WP_003030002.1 MULTISPECIES: [Streptococcus]
194	MEKEYTIGLDIGTNSVGWAVLTDDYRLVARKMSIQGDSNRKKIKKNFWGARLFEEGKTAQFRRIKRTNRRRIARRRQRVLALQDIFAEEIHKKDPNFFARLEEGDRVEADKRFAKFPVF ATLSEEKNYHRQYPTIYHLRHDLANSKEQADIRLVYLAIAHCLKYRGHFLFEGELDTENTSVTENYQQFLQAYQQFFPEPIGDLDDAVPILTERLSKAKRVEKVLAYYPSEKSTGNFAQ FLKLMVGNQANFKKTFDLEEEMKLNFTRDCYEEDLNELLEKTSDDYAELFLKAKGVYDAILLSQILSKSDDFTKAKLSANMKLRFEEHQRDLKQLKELVRRDLPKKYDDFFKNRSKNGY AGYVKGKATQEDFYKFLRTELAGLEESQSIMEKIDLEIYLLKQRTFANGVIPHQIHLVEMREIMDRQKRFYPFLKGAQGKIEKLLTFRIPYYVGPLAQEGQSPFAWIKRKSPSQITPWN FAEVVDKENSAIEFIERMTNQDTYLPKEKVLPKQSLIYQRFMIFNELTKVSYTDERGKSHYFSSEQKRKIFNELFKQHPRVTEKQLRKFLELINEQIDSTEIKGIETSFNASYSTYHDLL KLSDQMDTLLIDDPDMTTMFEEIIKILTIFEDREMIRERLQKYETVLGLIPAIKKLAKKHYTGWGRLSEKMIQGMREKQSRKTTILDYLIDDGIANRNFMQLINDHLISFKETIANEL IMSDSNVLLDQVKAIPGSPAVKKGIWQSIKIVEEIIGIGKAPKNIVIEMARENQRTSRSRPRLKALEEALKNIDSPLLKDYPTDNQALQKDRLYLYYLQNGKDMYTGEPLEIHRLSEY DIDHIIPRSFIVDNSLDNKVLVSSKVNRGKLDNAPDPLVVKRMRSHWEKLHQAKLISDKKLANLTKQNLTEADKARFIQRQLVETRQITKHVANLLHQHFNLPEEVSATEKTSIITLKS TLTSQFRQMFDIYKVREINNHHHAHDAYLNGVVAMTLLKKYPKLAPEFVYGSYIKGDINQINKATAKKEFYSNIMKFFESEEIICDEQGEVIWNKKRDLSTIKKTIGAHQVNIVKKVEK QKGGFYKETINSKANPEKLIPRKASLDPLKYGGYGSPLVAYTVIFIFEKGKQKKVTKGIEQLTVMEQLRFEDDPREFLKTKGYEGVKQWLILPKYILFEAQGGYRRMIASHQETQKANS LILPENLVTLLYHARHYDEINHKVSFDYVNAHKEGFNDIFDFISDFGVRYILAPQHLEKIKVAYEKNKEVDLKEMIDAILSLLKFTLFGASVEFKFFDIKILKRKYKSLTDIWEATIIYQ SVTGLYERRVEVRKLWDGERL	
195	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLKRTARRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESFLVEEDKKNERHPIF GNLADEVAYHRNYPTIYHLRKKLASDPEKADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLF GNIIAIALGLTPNFKSNFDLTEDAKLQLSKDTYDDDLDNLLGQIGDQYADLFSAAKNLSDAILLSDIIRSNSEVTKAPLSASMVKRYDEHHQDLALLKTIVRQQFPEKYAEIFKDDTKN GYAGYVGIGIKHRKRTTKLATQEEFYKFIKPILEKMDGAEELLAKLNRDDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPLARGNSRFAWLT RKSEEAITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRHYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFSNRNFMQLIHDDSLTF KEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKR DKNDKPIIREVKVITLKSKLVSDFRKDFQLYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNFFKTEVKLANGEIRKRPL IETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSKESILSKRESAKLIPRKKGWDTRKYGGFSPTVAYSILVVAKVEKGKAKKLKSVKELVGITIMEKGSYEKDPIGF	WP_003043819.1 [Streptococcus canis]

	LEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASATELQKANELVLPQHLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFTFLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD	
196	MEKSYSIGLDIGTNSVGWAVITDDYKVPAKKMKVLGNTDKKYIKKNLIGALLFDSGETAEVTRLKRTARRRYTRRKNRLRYLQEIFAKEMTKVDESFFQRLEESFLTDDDKTFDSHPIF GNKAEEDAYHQKFPTIYHLRKYLADSQEKADLRLVYLALAHMIKYRGHFLIEGQLNAENTDVQKLFNVFVETYDKIVDESHLSEIEVDASSILTEKVSKSRRLENLIKQYPTEKKNTLF GNLIALALGLQPNFKTNFKLSEDAKLQFSKDTYEEDLEELLGKVGDDYADLFISAKNLVDAILLSGILTVDDNSTKAPLSASMIKRYVEHHEDLEKLKEFIKINKLKLYHDIFKDKTKN GYAGYIDNGVKQDEFYKYLKTILTKIDDSDYFLDKIERDDFLRKQRTFDNGSIPHQIHLQEMHSILRRQGDYYPFLKENQAKIEKILTFRIPYYVGPLARKDSRFAWAMYHSDEPITPW NFDEVVDKEKSAEKFITRMTLNDLYLPEEKVLPKHSHVYETFTVYNELTKIKYVNEQGESFFFDANMKQEIFDHVFKENRKVTKAKLLSYLNNEFEEFRINDLIGLDKDSKSFNASLGT YHDLKKILDKSFDDKTNGQIIEDIVLTLTLFEDRDMIHERLQKYSDFFTSQQLKKLERRHYTGWGRLSYKLINGIRNKENNKTILDFLIDDGHANRNFMQLINDESLSFKTIIQEAQV VGDVDDIEAVVHDLPGSPAIKKGILQSVKIVDELVKVMGDNPDNIVIEMARENQTTGYGRNKSNQRLKRLQDSLKEFGSDILSKKKPSYVDSKVENSHLQNDRLFLYYIQNGKDMYTGE ELDIDRLSDYDIDHIIPQAFIKDNSIDNKVLTSSAKNRGKSDDVPSIEIVRNRRSYWYKLVKSGLISKRKFDNLTKAERGGLTEADKAGFIKRQLVETRQITKHVAQILDARFNTKRDE NDKVIRDVKVITLKSNLVSQFRKEFKFYKVREINDYHHAHDAYLNAVVGTALLKKYPKLIPEFVYGEYKKYDVRKLIAKSSDDYSEMGKATAKYFFYSNIMNFFKTEVVKYADGRVFERP DIETNADGEVVWNKQKDFDIVRKVLSYPQVNIVKKVEAQTGGFSKESILSKGDSDKLIPRKTKKVYWNTKKYGGFDSPTVAYSVLVVADIEKGKAKKLKTVKELVGISIMERSFFEENP VSFLEKKGYHNVQEDKLIKLPKYSLFEFEGGRRRLLASATELQKGNEVMLPAHLVELLYHAHRIDSFNSTEHLKYVSEHKKEFEKVLSCVENFSNLYVVDVEKNLSKVRAAAESMTNFSL EEISASFINLLITFTALGAPAAFKFFDKDVDRKRYTSTKECLSATLIHQSVTGLYETRIDLSKLGEE	
197	MEKTYSIGLDIGTNSVGWAVITDDYKVPAKMKVLGNTDKKYIKKNLIGALLFDSGETAEATRLKRAARRYTRRKNRLRYLQEIFAKEMAKVDESFFQRLEESFLTDDDKTFDSHPIF GNKAEEDTYHQEFPTIYHLRKHLADSPEKVDLRLVYLALAHMIKFRGHFLIEGQLNAENTDVQKIFADFVGVYDRTFDDSHLSEITVDAASILTEKISKSRRLENLIKQYPTEKKNTLF GNIVALALGLQPNFKTNFKLSEDAKLQFSKDTYDEDLEELLGKIGDDYADLIFTAAKNLVDAILLSGILTVDDNSTKAPLSASMIKRYEEHHEDLEKLKTFIKVNNFDKYHEIFKDVKSKN GYAGYIENGVKQDIFYKHLKSIISEKNGGQYFLDKIEREDFLRKQRTFDNGSIPYQIHLQEMRTILRRQGEVYPFLKENQAKIEKILTFRIPYYVGPLARKNSRFAWAKYHSDEPITPW NFDEVVDKEKSAEKFITRMTLNDLYLPEEKVLPKHSVYVETFAVYNELTKIKYVNEQGKSFFFDANMKQEIFDHVFKENRKVTKAKLLSYLNNEFEEFRINDLIGLDKDSKSFNASLGT YHDLKKILDKSFDLDDRDMIHERLQKYSDIFTSQQLKKLERRHYTGWGRLSYKLINGIRNKENNKTILDFLIDDGDANRNFMQLINDDSLSFKTTIQEAQV VGDVDDIEAVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPQNIVIEMARENQITGYGRNRSNQRLKRLQDSLKEFGSDILSKKKPSYVDSKVENSHLQNDRLFLYYIQNGKDMYTGE ELDIDHLSDYDIDHIIPQAFIKDNSILTSSAKNRGKSDDVPSIEIVRNRKSYWYKLVKSGLISKRKFDNLTKAERGGLTETDKAGFIKRQLVETRQITKHVAQILDARFNTKCDE NDKVIRDVKVITLKSSLVSQFRKEFKFYKVREINDYHHAHDAYLNAVVGTALLKKYPKLAPEFVYGEYKKYDVRKLIAKSSDNHSEIGKATAKYFFYSNIMNFFKTEVVKYADGRVFERP DIETNADGEVVWNKQRDFNIVRKVLSYPQVNIVKKVEVQTGGFSKESILPKGDSDKLIPRKTKKLQWETQRYGGFDSPTVAYSVLVVADVEKGKTRKLKTVKELVGISIMERSSFEENP VSFLEKKGYHNVQEDKLIKLPKYSLFELENGSRRLLASATELQKGNEVFLPQYMVNLLYHSQHVNNSHKPEHLNYVKQHKDEFKDIFNLIISIARINILKPKVVDNLINEFTEYGQEDI SSLSESFINLLKFISFGAPGAFKFLKLDVKQSNLRYKSTTEALSATLIHQSVTGLYETRIDLSKLGEE	
198	MENKNYSIGLDIGTNSVGWAVITDDYKVPSKKMKVLGNTDKRFIKKNLIGALLFDEGTTAEARRLKRTARRRYTRRKNRLRYLQEIFAEEMSKVDDSFFHRLDDSFLIPEDKKGSKYPI FATLIEEKEYHKQFPTIYHLRKQLADSKEKTDLRLIYLALAHMIKYRGHFLYEEDTFDIKNNDIQKIFNEFISIYNNTFEGNSLSGQNVQVEAIFTDKISKSAKRRVLKLFPDDEKSTGL FSEFLKLIVGNQADFKKHFDLEEKAPLQFSRDTYDEDLENLLGQIGDDFADLFVAAKKLYDAILLSGILTVTDPSTKAPLSASMIERYENHQKDLASLKQFIKNNLPEKYDEVFSDQSK DGYAGYIDGKTTQESFYKYIKNLLSKFEGADYFLEKIERDDFLRKQRTFDNGSIPHQIHLQEMNAILRRQGEHYPFLKENKEKIEKILTFRIPYYVGPLARGNRDFAWLTRNSDQAIRP WNFEEIVDKASSAESFINKMTNYDLYLPEEKVLPKHSLLYETFAVYNELTKVKFIAEGLRDYQFLDSRQKKDIFYTLFKAEDKRKVTEKDIIQYLHTVDGYDGIELKGIEKQFNASLSTYH YHDLLKIIKDKEFMDDPNNEEILENIVHTLTIFEDREMIKQRLAQYDSLFDEKVIKALTRRHYTGWGKLSAKLINGISDKQTGNTILDYLIDDGEINRNFMQLINDDGLSFKEIIQKAQVV VVGKTDDVKQVVQELPGSPAIKKGILQSIKLVDELVKVMGHEPESIVIEMARENQTTARGKKNSQQRYKRIEDSLKIILASGLNAKILKEHPTDNIQLQNDRLFLYYLQNGRDMYTGKPL DINQLSSYDIDHIVPQAFIKDDSLDNRVLTSLKDNRGKSDNVPSLEVVEKMKTFWQQLLDSKLISYRKFNNLTKAERGGLDERDKVGFIKRQLVETRQITKHVAQILDARYNTEVNEKD KKNRTVKIITLKSNLVSNFRKEFRLYKIREINDYHHAHDAYLNAVVAKALIKKYPKLESEFVYGDYQKYDLKRYISRSKDPKEIGKATAKYFFYSNLLNFFKEEVHYAQGTIVKRENIEYS YSKDTGEIAWNKEKDFATIKKVLSLPQVNIVKKREVQTGGFSKESILPKGNSDKLIPRKTKDILWDTTKYGGFDSPVIAYSILLIADIEKGKAKKLKTVKTLVGITIMEKDAFEKNPIAFL FLENKGYHNVRKENILCLPKYSLFELKNGRRRMLASAKELQKGNEIVLPVHLTTLLYAHKNIHRLDEPEHLEYIQKHRNEFKDLINLVSEFSQKYVLADANLEKIKNLYADNEQADIEILA LANSFINLLTFTALGAPAAFKFFGKDVDRKRYTTVSEILNATLIHQSITGLYETRIDLSKLGED	
199	MENKNYSIGLDIGTNSVGWSVITDDYKVPSKKMKVLGNTDKHFIKKNLIGALLFDEGTTAEARRLKRTARRRYTRRKNRLRYLQEIFSEEIGKVDDSFFHRLDDSFLIPEDKRGSKYPI FATLAEEKKYHKQFPTIYHLRKQLADSKEKTDLRLIYLALAHMIKYRGHFLYEEEAFDIKNNDIQKIFNEFISIYNNTFEGNSLSGQNVQVEAIFTDKISKSAKRRVLKLFPDDEKRGSKYPI FSEFLKLIVGNQADFKKHFDLEEKAPLQFSRDTYDEDLENLLSGILTVTNPSTKAPLSASMIERYENHQKDLAVLKQFIKNNLPEKYDEVFSDQSK DGYAGYIDGKTTQETFYKYIKNLLSKFEGADYFLDKIERDDFLRKQRTFDNGSIPHQIHLQEMNAILRRQGEHYPFLKENKEKIEKILTFRIPYYVGPLARGNRDFAWLTRNSDEAIRP WNFEEIVDKASSAEDFINKMTNYDLYLPEEKVLPKHSLLYETFAVYNELTKVKFIAEGLRDYQFLDSGQKKQIVTQLFKEKRKVTEKDIIQFLHNVDGYDGIELKGIEKQFNASLSTYH DLLKIIKDKAFMDDAKNEAILENIVHTLTIFEDREMIKQRLAQYDSLFDEKVIKALTRRHYTGWGKLSAKLINGISDKQTGNTILDYLIDDGLSFKEIIQKAQVV GKTNDVKQVVQELPGSPAIKKGILQSIKIVDELVKVMGHAPESIVIEMARENQTTARGKKNSQQRYKRIEDSLKILKEHPTDNIQLQNDRLFLYYLQNGKDMYTGEALDI NQLSSCDIDHIIPQAFIKDDSLDNRVLTSSKDNRGKSDNVPSLEVVDKMKVFWQQLLDSKLISYRKFNNLITKAERGGLNELDKVGFIKRQLVETRQITKHVAQILDARFNKEVTEKDKK NRTVKIITLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVVAKALIKKYPKLESEFVYGDYQKYDLKRYISRSKDPKEIGKATAKYFFYSNLLNFFKEEVHYADGTIVKRENIEYS KDTGEIAWNKEKDFATIKKVLSLPQVNIVKKREVQTGGFSKESILPKGNSDKLIPRKTKDILWETTKYGGFDSPVIAYSILLIADIEKGKAKKLKTVKTLVGITIMEKAAFEENPITFL ENKGYHNVRKENILCLPKYSLFELENGNRRRLLASAKELQKGNEIVLPVYLTTLLYHSKNVHKLDEPEHLEYIQKHRNEFKDLINLVSEFSQKYVLADANLEKIQNLYADNEQADIEILA NSFINLLTFTALGAPAAFKFFGKDIDRKRYTTVSEIINATLIHQSITGLYETRIDLSKLGED	
200	MENKNYSIGLDIGTNSVGWSVITDDYKVPSKKMKVLGNTDKRFIKKNLIGALLFDEGTTAEARRLKRTARRYTRRKNRLRYLQEIFAEEMSKVDSSFFHRLDDSFLIPEDKRGSKYPI FATLAEEKEYHKQFPTIYHLRKQLADSKEKTDLRLIYLALAHMIKYRGHFLYEDTFDIKNNDIQKIFSEFISIYNDNTFEGSSLSGQNAQVEAIFTDKISKSAKRERVLKLFPDEKSTGL FSEFLKLIVGNQADFKKHFDLEEKAPLQFSKDTYDEDLADLIFSVAKKLYDAILLSGILTVTDPSTKAPLSASMIERYENHQKDLAVLKQFIKNNLPEKYDEVFSDQSK DGYAGYIDGKTTQEAFYKYIKNLLSKFEGTDYFLEKIERDDFLRKQRTFDNGSIPHQIHLQEMNAILRRQGEHYPFLKDNKEKIEKILTFRIPYYVGPLARGNRDFAWLTRNSDEAIRP WNFEEIVDKASSAEDFINKMTNYDLYLPEEKVLPKHSLLYETFAVYNELTKVKFIAEGLRDYQFLDSGQKKQIVTQLFKEKRKVTEKDIIQFLHNVDGYDGIELKGIEKQFNASLSTYH DLLKIIKDKEFMDDSKNEEILENIVHTLTIFEDREMIKQRLAQYDSLFDEKVIKALIRRHYTGWGKLSAKLIDGICDKQTGNTILDYLIDDGKNNRNFMQLINDDGLSFKEIIQKAQVV GKTDDVKQVVQELPGSPAIKKGILQSIKIVDELVKVMGHTPESIVIEMARENQTTARGKKNSQQRYKRIEDALKNLASGLDSNILKEHPTNNIQLQNDRLFLYYLQNGRDMYTGKPLDI NQLSSYDIDHIVPQAFIKDDSLDNRVLTSLKDNRGKSDNVPSIEVVQKRKAFWQQLLDSKLISERKFNNLITKAERGGLDERDKVGFIKRQLVETRQITKHVAQILDARFNKEVNEKDKK NRTVKIITLKSNLVSNFRKEFRLYKVREINDYHHADAYLNAVVAKALIKKYPKLESEFVYGDYQKYDLKRYISRSRNPKEVEKATEKYFFYSNILNFFKEEVHYADGTIVKRENIEYS KDTGEIAWNKEKDFATIKKVLSYPQVNIVKKREVQTGGFSKESILPKGNSDKLIPRKTKDILWETTKYGGFDSPVIAYSILLIADIEKGKAKKLKTVKTLVGITIMEKAAFEENPITFL ENKGYHNVRKENILCLPKYSLFELENGNRRRLLASAKELQKGNEIVLPVYLTTLLYHSKNVHKLDEPEHLEYIQKHRNEFKDLINLVSEFSQKYVLADANLEKIQNLYADNEQADIEILA NSFINLLTFTALGAPAAFKFFGKDIDRKRYTTVSEIINATLIHQSITGLYETRIDLSKLGED	

201	MGKEYTIGLDIGTNSVGWAVLQEDLDLVRRKMKVYGNTEKNYLKKNFWGVDLFDEGMTAKDTRLKRTTRRRYFRRRQRISYLQTFFSQEEMNRIDPNFFNRLDESFLIEEDKLSERHPIF GTIEEEVAYHKNYATIYHLRKELADAEEKADLRLVYLALAHIIKYRGHFLIEGRLSTENTSTEETFKTFLQKYNQTFNPVDETISIGSIFADKVSRAKKAEGVLALFPDEKRNGTFDQF LKMIVGNQGNFKKTFELEEDAKLQFSKEEYDESLEALIGDEYADVFEAAKNVYNAVELSGILTVTDNSTKAKLSASMIKRYEDHKTDLKLFKEFIRKNLPEKYHEIFNDKNTDGYA GYIDNSKKTSQEKFYKYITNLIEKIDGAEYFLKKIENEDFLRKQRTFDNGIIPHQIHLELKAILHHQAMYYPFLQEKFSNFVDLITFRIPYYVGPLANGNSRFSWLSRKSDEPIRPWN LAEVVDLSKSAELFIERMTNFDLYLPSEKVLPKHSMLYEKYTVYNELTKVTYKDEQGKVQNFSSEEKERIFIDLFKQHRKVTKKDLSNFLRNEYNLDDVIIDGIENKFNASFNTYHDFL KIKIDPKVLDDPANEPMFEEIVKILTIFEDRKMLRERLSKFSDRLSEKTIKDLERRHYTGWGRLSAKLINGIHDKQSNKTILDYLINDDAPKKNINRNFMQLINDNRLTFKEEIEKEQL KANSEESLIEIVQNLAGSPAIKKGILQSLKIVDELVKIMGYAPTNIVVEMARENQTTANGRRNSRPRLKNLEKAIDDLDSEILKKHPVDNKALQKDRLYLYYLQNGKDMYTNEELDIHK LSTYDIDHIIPQSFIVDNSLDNRVLVSSSKNRGKLDDVPSKEVVKKMKAFWESLYRSGLISKKKFDNLVKAESGGLSEDDKAGFIHRQLVETRQITKNVARILEQRFNSEKDEEGNLIR KVRIITLKSALTSQFRKNYGIPKIREINDYHHAHDAYLNGVVATALLKIYPQLEPEFVYGEFHRFNAFKENKATAKKQFYSNIMNFFSKSDKVIIDENGEILWNQKKIVTVKKVMNYRQM NIVKKVEIQKGGFSKESILPKGDSDKLISRKKEWDTTKYGGFDSPNVAYSVVVIRYEKGKTRKLVKTIVGITIMERAAFEKNEREFLIKNKGYQNPQICMKLPKYSLYEFEDGRRRLLASA KEAQKGNQMVLPAHLVTFLYHAKHCNEKPDSLKYVTEHQSGFSEIMAHVKDFAEKYTLVDKNLEKILSLYAKNMDSEVKEIAQSFVDLMQLNAFGAPADFKFFGETIPRKRYTSVNELL EATIINQSITGLYETRRRLGD	WP_048604708.1 [Enterococcus sp. AM1]
202	MGKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTGEMNKVDENFFQRLDDSFLVDEDKRGEHHPIF GNIAAEVKYHDDFPTIYHLRRHLADTSKKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVSKSRRLENLIAHYPTEKKNTLF GNLIALSLDLHPNFKTNFQLSEDAKLQFSKDTYEEDLEGFLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNKK GYASYIESGVKQDEFYKYIKGILLKINGSGDFLDKIDREDFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPW NFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEAFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLINYLNKEFEEFRIVNLTGLDKENKAFNSSLGT YHDLRKILDKSFLDDKANEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIAKAQI IDDVVDIANVVHDLPGPSGKAIKKGILQSVKIVDELVKIMGYAPKRIVVEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGEELTINNLS QYDIDHIIPQAFIKDNSLDNRVLVTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEKDENKRRIRNV KIITLKSNLVSNFRKEFELYKVRETNDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQV NIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWDPKKYGGFDSPVVAYSILVFAIVEKGKSKKLKRVKDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELEN KRRRLLASAKELQKGNELVIPQRFTTLLYHSYRIEKDYEPEHRREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSKNKDAEVSSLAKSFISLITFTAFGAPAAAFNFFGENID RKRYTSVTECLNATLIHQSITGLYETRIDLSKLGED	WP_010824395.1 [Enterococcus faecalis]
203	MKKDYVIGLDIGSNSVGWAVMTEDYQLVKKKMPIYGNTEKKHKIKKNFWGVRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTDLDENFFARLQESFLVPEDKKWHRHPIF AKLEDEVAYHETVPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHFLIEGKLSTENTSVKDQFQQFMVIYNQTFVNGESRLVSAPLPESVLIEELEILTEKASRTKKSEKVLQQFPQE KANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEYSDVFIAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKNFKRFIRENCPDEYDNLF KNEQKDGYAGYITHAGKVSQLKFYQVVKKIIQDIAGAEYFLEKIAQENFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQKKIEQLVTFRIPYYVGPLSKGDASTFAWLKR QSEEPIRPWNLQETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTKIKYVTEGMRKPAFLSGEQKKAIVDLFFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFN NASFSTYQDLLKCGLITRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFKGQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLIKDDGVSKHYNRNFMQLINDSQL SFKNAIQKAQSSEHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEKAMAEIGSKNLLKEQPTTNEQLRDTRLFLYYMQNGKDM YTGDELSLHRLSHYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKKMKAYWEKLYAAGLISQRKFQRLTKGEQGGLTELEDKAHFIQRQLVETRQITKNVAGILDQRYNA KSKEKKVQIITLKASLTSQFRSIFGLPKVREVNDYHHGQDAYLNCVVATTLLKVYPNLAPEFVYGEYPKFQTFKENKATAKTIYTNLMRFFTEDEPRFTKDGEILWSNSYLKTIKKEL NYHQMNIVKKVEVQKGGFSKESIKPKGPSNKLIPVKNGLDPQKYGGFDSPVVAYTVLFTHEKGKKPLIKQEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRR LLASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSBSLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFNAMGAPASTFKFFQKDIERARY TSIKEIFDATIIYQSTTGLYETRRKVVD	WP_033624816.1 [Enterococcus faecalis]
205	MKKDYVIGLDIGTNSVGWAVMTEDYQLVKKKMPIYGNTEKKHKIKKNFWGVRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTALDENFFARLQESFLVPEDKKWHRHPIF AKLEDEVAYHETVPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHFLIEGKLSTENISVKEKFQQFMIIYNQTFVNGEGRLVSAPLPESVLIEELETEKASRTKKSEKVLQQFPQE KANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEYSDVFIAAKNVYDAVELSSILTLADSDKKSYAKLSSSMIVRFTEHQEDLKKFFKRFIRENCPDEYDNLF KNEQKDGYAGYIAHAGKVSQLKFYQVVKKIIQDIAGAEYFLEKIAQENFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQEKIEQLVTFRIPYYVGPLSKGDASTFAWLKR QSEEPIRPWNLQETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTKIKYVTDDRGIKANFSGKEKEKIFDTLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQF NASFSTYQDLLKCGLITRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFKGQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLVKDDGVSKHYNRNFMQLINDSQL SFKNAIQKAQSSEHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEKAMAEIGSNLLKEQPTTNEQLRDTRLFLYYMQNGKDM YTGDELSLHRLSHYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKDMKAYWEKLYAAGLISQRKFQRLTKGEQGGLTELEDKAHFIQRQLVETRQITKNVAGILDQRYNA KSKEKKVQIITLKASLTSQFRSIFGLPKVREVNDYHHGQDAYLNCVVATTLLKVYPNLAPEFVYGEYPKFQTFKENKATAKAIYTNLLRFFTEDEPRFTKDGEILWSNSYLKTIKKEL NYHQMNIVKKVEVQKGGFSKESIKPKGPSNKLIPVKNGLDPQKYGGFDSPIVAYTVLFTHEKGKKPLIKQEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRR LLASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSBSLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFNAMGAPASTFKFFQKDIERARY TSIKEIFDATIIYQSTTGLYETRRKVVD	WP_016622645.1 [Enterococcus faecalis]

206	MKKDYVIGLDIGTNSVGWAVMTEDYQLVKKKMPIYGNTEKKKIKKNFWGVRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTALDENFFARLQESFLVPEDKKWHRHPIF AKLEDEVAYHETYPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHFLIEGKLSTENISVKEQFQQFMIIYNQTFVNGESRLVSAPLPESVLIEEELTEKASRTKKSEKVLQQFPQE KANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEYSDVFLAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKNFKRFIRENCPDEYDNLF KNEQKDGYAGYIITHAGKVSQLKFYQVVKKIIQDIAGAEYFLEKIAQENFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQEKIEQLVTFRIPYYVGPLSKGDANTFAWLKR QSEEPIRPWNLQETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTKISYTDDRGIKANFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQF NASFSTYQDLLKCGLTRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFKGQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLIKDDGVSKHYNRNFMQLINDSQL SFKNAIQKAQSSEHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEKAMAEIGSNLLKEQPTTNEQLRDTRLFLYYMQNGKDM YTGDELSLHRLSHYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKKMKAYWEKLYAAGLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITKNVAGILDQRYNAKSKEKKVQIITL KSKEKKVQIITLKASLTSQFRSIFGLYKVREVNDYHHGQDAYLNCVVATTLLKVYPNLAPEFVYGEYPKFQTFKENKATAKAIIYTNLLRFFTEDEPRFTKDGEILWSNSYLKTIKKEL NYHQMNIVKKVEVQKGGFSKESIKPKGPSNKLIPVKNGLDPQKYGGFDSPVVAYTVLFTHEKGKKPLIKQEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRR LLASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFNAMGAPSTFKFFQKDIERARY TSIKEIFDATIIYQSTTGLYETRRKVVD	WP_002378009.1 [Enterococcus faecalis]
207	MKKDYVIGLDIGTNSVGWAVMTEDYQLVKKKMPIYGNTEKKKIKKNFWGVRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTALDENFFARLQESFLVPEDKKWHRHPIF AKLEDEVAYHETYPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHFLIEGKLSTENISVKEQFQQFMIIYNQTFVNGESRLVSAPLPESVLIEEELTEKASRTKKSEKVLQQFPQE KANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEYSDVFLAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKNFKRFIRENCPDEYDNLF LFKNEQKDGYAGYIAHAGKVSQLKFYQVVKKIIQDIAGAEYFLEKIAQENFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQEKIEQLVTFRIPYYVGPLSKGDASTFAWLKR KRQSEEPIRPWNLQ ETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTKISYTDDRGIKANFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQFNASFSTYQDLLK CGLTRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFKGQFSEEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLIKDDGVSKHYNRNFMQLINDSQLSFKNAIQKAQSS EHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEKAMAEIGSNLLKEQPTTNEQLRDTRLFLYYMQNGKDMYTGDELSLHRLS HYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKKMKAYWEKLYAAGLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITKNVAGILDQRYNAKSKEKKVQIITL KASLTSQFRSIFGLYKVREVNDYHHGQDAYLNCVVATTLLKVYPNLAPEFVYGEYPKFQAFKENKATAKAIIYTNLLRFFTEDEPRFTKDGEILWSNSYLKTIKKELNYHQMNIVKKVE VQKGGFSKESIKPKGPSNKLIPVKNGLDPQKYGGFDSPVVAYTVLFTHEKGKKPLIKQEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRRLLASAKEAQKGN QMVLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFNAMGAPSTFKFFQKDIERARYTSIKEIFDATII YQSTTGLYETRRKVVD	WP_010775580.1 [Enterococcus faecalis]
208	MKKDYVIGLDIGTNSVGWAVMTEDYQLVKKKMPIYGNTEKKKIKKNFWGVRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTALDENFFARLQESFLVPEDKKWHRHPIF AKLEDEVAYHETYPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHFLIEGKLSTENISVKEQFQQFMIIYNQTFVNGESRLVSAPLPESVLIEEELTEKASRTKKSEKVLQQFPQE KANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEYSDVFLAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKKFKRFIRENCPDEYDNLF KNEQKDGYAGYIAHAGKVSQLKFYQVVKKIIQDIAGAEYFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQEKIEQLVTFRIPYYVGPLSKGDASTFAWLKR QSEEPIRPWNLQETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTKISYTDDRGIKANFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQF NASFSTYQDLLKCGLTRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFKGQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLIKDDGVSKHYNRNFMQLINDSQL SFKNAIQKAQSSEHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEKAMAEIGSNLLKEQPTTNEQLRDTRLFLYYMQNGKDM YTGDELSLHRLSHYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKKMKAYWEKLYAAGLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITKNVAGILDQRYNA KSKEKKVQIITLKASLTSQFRSIFGLYKVREVNDYHHGQDAYLNCVVATTLLKVYPNLAPEFVYGEYPKFQAFKENKATAKAIIYTNLLRFFTEDEPRFTKDGEILWSNSYLKTIKKEL NYHQMNIVKKVEVQKGGFSKESIKPKGPSNKLIPVKNGLDPQKYGGFDSPVVAYTVLFTHEKGKKPLIKQEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRR LLASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFNAMGAPSTFKFFQKDIERARY TSIKEIFDATIIYQSTTGLYETRRKVVD	WP_010818269.1 [Enterococcus faecalis]
209	MKKDYVIGLDIGTNSVGWAVMTEDYQLVKKKMPIYGNTEKKKIKKNFWGVRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTALDENFFARLQESFLVPEDKKWHRHPIF AKLEDEVAYHETYPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHFLIEGKLSTENISVKEQFQQFMIIYNQTFVNGESRLVSAPLPESVLIEEELTEKASRTKKSEKVLQQFPQE KANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEYSDVFLAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKKFKRFIRENCPDEYDNLF KNEQKDGYAGYIAHAGKVSQLKFYQVVKKIIQDIAGAEYFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQEKIEQLVTFRIPYYVGPLSKGDASTFAWLKR QNEKPIRPWNLQETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTKISYTDDRGIKANFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQF NASFSTYQDLLKCGLTRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFKGQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLIKDDGVSKHYNRNFMQLINDSQL SFKNAIQKAQSSEHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEKAMAEIGSNLLKEQPTTNEQLRDTRLFLYYMQNGKDM YTGDELSLHRLSHYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKKMKAYWEKLYAAGLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITKNVAGILDQRYNA KSKEKKVQIITLKASLTSQFRSIFGLYKVREVNDYHHGQDAYLNCVVATTLLKVYPNLAPEFVYGEYPKFQAFKENKATAKAIIYTNLLRFFTEDEPRFTKDGEILWSNSYLKTIKKEL NYHQMNIVKKVEVQKGGFSKESIKPKGPSNKLIPVKNGLDPQKYGGFDSPVVAYTVLFTHEKGKKPLIKQEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRR LLASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFNAMGAPSTFKFFQKDIERARY TSIKEIFDATIIYQSTTGLYETRRKVVD	WP_033625576.1 [Enterococcus faecalis]
210	MKKDYVIGLDIGTNSVGWAVMTEDYQLVKKKMPIYGNTEKKKIKKNFWGVRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTALDENFFARLQESFLVPEDKKWHRHPIF AKLEDEVAYHETYPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHFLIEGKLSTENISVKEQFQQFMIIYNQTFVNGESRLVSAPLPESVLIEEELTEKASRTKKSEKVLQQFPQE KANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEYSDVFLAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKKFKRFIRENCPDEYDNLF KNEQKDGYAGYIAHAGKVSQLKFYQVVKKIIQDIAGAEYFLEKIAQENFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQKKIEQLVTFRIPYYVGPLSKGDASTFAWLKR QSEEPIRPWNLQETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTKISYTDDRGIKANFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQF NASFSTYQDLLKCGLTRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFKGQFSEEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLIKDDGVSKHYNRNFMQLINDSQL SFKNAIQKAQSSEHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEKAMAEIGSNLLKEQPTTNEQLRDTRLFLYYMQNGKDM YTGDELSLHRLSHYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKKMKAYWEKLYAAGLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITKNVAGILDQRYNA KSKEKKVQIITLKASLTSQFRSIFGLYKVREVNDYHHGQDAYLNCVVATTLLKVYPNLAPEFVYGEYPKFQAFKENKATAKAIIYTNLLRFFTEDEPRFTKDGEILWSNSYLKTIKKEL NYHQMNIVKKVEVQKGGFSKESIKPKGPSNKLIPVKNGLDPQKYGGFDSPVVAYTVLFTHEKGKKPLIKQEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRR LLASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFNAMGAPSTFKFFQKDIERARY TSIKEIFDATIIYQSTTGLYETRRKVVD	WP_033789179.1 [Enterococcus faecalis]

211	MKKDYVIGLDIGTNSVGWAVMTEDYQLVKKKMPIYGNTEKKKIKKNFWGVRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTALDENFFARLQESFLVPEDKKWHRHPIF AKLEDEVAYHETYPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHFLIEGKLSTENISVKEQFQQFMIIYNQTFVNGESRLVSTPLPESVLIEEELTEKASRTKKSEKVLQQFPQE KANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEYSDVFLAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKNFKRFIRENCPDEYDNLF KNEQKDGYAGYITHAGKVSQLKFYQVVKKIIQDIAGAEYFLEKIAQENFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQEKIEQLVTFRIPYYVGPLSKGDASTFAWLKR QSEEPIRPWNLQETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTKISYTDDRGIKANFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQF NASFSTYQDLLKCGLITRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFKGQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLVKDDGVSKHYNRNFMQLINDSQL SFKNAIQKAQSSEHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEKAMAEIGSKNLLKEQPTTNEQLRDTRLFLYYMQNGKDM YTGDELSLHRLSHYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKDMKAYWEKLYAAGLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITKNVAGILDQRYNA KSKEKKVQIITLKASLTSQFRSIFGLYKVREVNDYHHGQDAYLNCVVATTLLKVYPNLAPEFVYGEYPKFQTFKENKATAKAIIYTNLLRFFTEDEPRFTKDGEIIWSNSYLKTIKKEL NYHQMNIVKKVEVQKGGFSKESIKPKGPSNKLIPVKNGLDPQKYGGFDSPIVAYTVLFTHEKGKKPLIKQEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRR LLASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFNAMGAPSTFKFFQKDIERARY TSIKEIFDATIIYQSTTGLYETRRKVVD	WP_002407324.1 [Enterococcus faecalis]
212	MKKDYVIGLDIGTNSVGWAVMTEDYQLVKKKMPIYGNTEKKKIKKNFWGVRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTDLDENFFARLQESFLVPEDKKWHRHPIF AKLEDEVAYHETYPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHFLIEGKLSTENISVKEQFQQFMIIYNQTFVNGESRLVSAPLPESVLIEEELTEKASRTKKSEKVLQQFPQE KANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEYSDVFLAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKNFKRFIRENCPDEYDNLF KNEQKDGYAGYITHAGKVSQLKFYQVVKKIIQDIAGAEYFLEKIAQENFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQEKIEQLVTFRIPYYVGPLSKGDASTFAWLKR QSEEPIRPWNLQETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTKISYTDDRGIKANFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQF NASFSTYQDLLKCGLITRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFKGQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLVKDDGVSKHYNRNFMQLINDSQL SFKNAIQKAQSSEHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEKAMAEIGSKNLLKEQPTTNEQLRDTRLFLYYMQNGKDM YTGDELSLHRLSHYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKDMKAYWEKLYAAGLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITKNVAGILDQRYNA NSKEKKVQIITLKASLTSQFRSIFGLYKVREVNDYHHGQDAYLNCVVATTLLKVYPNLAPEFVYGEYPKFQTFKENKATAKAIIYTNLLRFFTEDEPRFTKDGEIIWSNSYLKTIKKEL NYHQMNIVKKVEVQKGGFSKESIKPKGPSNKLIPVKNGLDPQKYGGFDSPIVAYTVLFTHEKGKKPLIKQEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRR LLASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFNAMGAPSTFKFFQKDIERARY TSIKEIFDATIIYQSTTGLYETRRKVVD	WP_002364836.1 MULTISPECIES: [Enterococcus]
213	MKKDYVIGLDIGTNSVGWAVMTEDYQLVKKKMPIYGNTEKKKIKKNFWGVRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTDLDENFFARLQESFLVPEDKKWHRHPIF AKLEDEVAYHETYPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHFLIEGKLSTENISVKEQFQQFMIIYNQTFVNGESRLVSAPLPESVLIEEELTEKASRTKKSEKVLQQFPQE KANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEYSDVFLAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKKFKRFIRENCPDEYDNLF KNEQKDGYAGYITHAGKVSQLKFYQVVKKIIQDIAGAEYFLEKIAQENFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQEKIEQLVTFRIPYYVGPLSKGDASTFAWLKR QSEEPIRPWNLQETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTKISYTDDRGIKANFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQF NASFSTYQDLLKCGLITRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFKGQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLVKDDGVSKHYNRNFMQLINDSQL SFKNAIQKAQSSEHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEKAMAEIGSKNLLKEQPTTNEQLRDTRLFLYYMQNGKDM YTGDELSLHRLSHYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKDMKAYWEKLYAAGLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITKNVAGILDQRYNA NSKEKKVQIITLKASLTSQFRSIFGLYKVREVNDYHHGQDAYLNCVVATTLLKVYPNLAPEFVYGEYPKFQTFKENKATAKAIIYTNLLRFFTEDEPRFTKDGEIIWSNSYLKTIKKEL NYHQMNIVKKVEVQKGGFSKESIKPKGPSNKLIPVKNGLDPQKYGGFDSPIVAYTVLFTHEKGKKPLIKQEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRR LLASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFNAMGAPSTFKFFQKDIERARY TSIKEIFDATIIYQSTTGLYETRRKVVD	WP_002413717.1 [Enterococcus faecalis]
214	MKKDYVIGLDIGTNSVGWAVMTEDYQLVKKKMPIYGNTEKKKIKKNFWGVRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTDLDENFFARLQESFLVPEDKKWHRHPIF AKLEDEVAYHETYPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHFLIEGKLSTENTSVKEQFQQFMVIYNQTFVNGESRLVSAPLPESVLIEEELTEKASRTKKSEKVLQQFPQE KANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAKVGDEYSDVFLAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKKFKRFIRENCPDEYDNLF KNEQKDGYAGYITHAGKVSQLKFYQVVKKIIQDIAGAEYFLEKIAQENFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQEKIEQLVTFRIPYYVGPLSKGDASTFAWLKR QSEEPIRPWNLQETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVFNELTKISYTDDRGIKANFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQF NASFSTYQDLLKCGLITRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLSTFKGQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLVKDDGVSKHYNRNFMQLINDSQL SFKNAIQKAQSSEHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEKAMAEIGSKNLLKEQPTTNEQLRDTRLFLYYMQNGKDM YTGDELSLHRLSHYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKKMKAYWEKLYAAGLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITKNVAGILDQRYNA RSKEKKVQIITLKASLTSQFRSIFGLYKVREVNDYHHGQDAYLNCVVATTLLKVYPNLAPEFVYGEYPKFQAFKENKATAKTIIYTNLMRFFTEDEPRFTKDSEILWSNSYLKTIKKEL NYHQMNIVKKVEVQKGGFSKESIKPKGPSNKLIPVKNGLDPQKYGGFDSPVVAYTVLFTHEKGKKPLIKQEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRR LLASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFNAMGAPSTFKFFQKDIERARY TSIKEIFDATIIYQSTTGLYETRRKVVD	WP_002373311.1 [Enterococcus faecalis]
215	MKKEYTIGLDIGTNSVGWAVLTENYDLVKKKMKVYGNTETKYLKKNLWGVRLFDEGETAADRRLKRTTRRRYSRRRNRICRLQDLFTEEMNQVDANFFHRLQESFLVPDEKEFERHAIF GRMEEEVSYYREFPTIYHLRKHLADTSEQADLRLVYLALAHIVKYRGHFLIEGELNTENSSVSETFRTFIQVYNQIFRENEVPLAVPDNIEELFSEKVSRARKVEAILSVYSEEKSTGI LAQFLKLMVGNQGRFKKTFDLEEDGIIQIPKEEYEEELETLLAIIGDEYAEIFSATKSVVDAVALSGILSVTDGDTKAKLSASMVERYEAHQKDLVQFKQFIRKELPEMYAPIFRDNSV SGYAGYVENSKVVTQAEFYKYIKKAIEKVPGAEYFLEKIEQETFLDKQRTFNNGVIPHQIHLEELEAIIQKQATYYPFLADNKEEMKQLVTFRIPYYVGPLADGNSPFAWIERISSEPI RPGNIAEVVDIKKSATKFIERMTNFDIYLPKHSMIYEKYMVFNELTKVSYKEDRGMNQRFSGEEKKQIVEELFKNEPTGKVTKKLLEKFLNEFGLVDVAIKGIETISFNAGYGTY HDFLKIGITREQLDKEENSETLEEIVKILTVFEDRRMIREQLKKYTYLFDEEVLKKLERRHYTGWGRLSAKLLIGIKEKRTHKTILDYLIEDDGKQPINRNLMQLINDSDLSFKSEIA EAQSDMNTEDLHEVVQNLAGSPAIKKGILQSLKIVDELVDIMGSLPKYVVEMARENQTTSRGRTNSNPRMKALEEEAMRNLRSNLLKEYPTDNQALQNDRLYLYYLQNGKDMYTGLDLS LHNLSSYDIDHIVPQSFTKDDSLDNRVLVSSKENRGKKDDVPSKEVVQKNITLWETILKNGLISQKKYDNLTKAERGGLSETDKAGFIKRQLVETRQITKHVARILDQRFNSQKDEEGK TIRAVRVVTLKSSLTSQFRKQFAIHKVREINDYHHGHDAYLNGVVANSLLRVYPQLQPEFVYGDYPKFNAYKANKATAKKQLYTNIMKFFAEDAVIIDENGEILWDKKNIATVKKVMSY PQMNIVKKPEIQTGSFSKETIKPKGDSDKLISRKTNWSPKLYGGFDSPQVAYSVIIIYEKGKKKVRAKAIVGITIMEQSLFKKDPVSLLEEKGYANPEVLIHLPKYTLYELENGRRRLL ASANEAQKGNQLVLPASLVITLLYHAKQVDEDSGKSEEYVREHRAEFAELINYVQAFSETKIILANKNLQTILKLYEENKEADIKEIAESFVNLMKFSAYGAPDMDFKFFGKTIPRSRYTSV GELLSATIINQSITGLYETRRKLVD	WP_010770040.1 [Enterococcus phoenicuricola]

216	MKKEYTIGLDIGTNSVGWSVLTDDYRLVSKKMKVAGNTEKSSTKKNFWGVRLFDEGQTAEARRSKRTARRLARRRQRILELQKIFAPEILKIDEHFFARLNESFLVLDEKKQSRHPVF ATIKQEKSYHQTVPTIYHLRQALADSSEKADIRLVYLAMAHLLKYRGHFLIEGELNTENSSVTETFRQFLSTYNQQFSEADDKQTEKLDEAVDCSFVFTEKMSKTKKAETLLKYFPHEK SNGYLSQFIKLMVGNQGNFKNVFGLEEAKLQFSKETYEEDLEELLEKIGDDYIDLFVQAKNVYDAVLLSEILSDSTKNTRAKLSAGMIRRYDAHKRDLVLLKRFVKENLPKKYRAFFGD NSVNGYAGYIEGHATQEDFYKFVKKELTGIRGSEVFLTKIEQENFLRKQRTFDNGVIPHQIHLTELRAIIANQKKHYPFLKEEQEKLESLLTFKIPYYVGPLAKGQNSPFAWLIRKSE EKIKPWNLPEIVDMEGSAVRFIERMINTDMYMPHNKVLPKNSLLYQKFSIYNELTKVRYQDERGQMNYFSSIEKKEIFHELFEKNRKVTKKDLQEFLYLKYDIKHAELSGIEKAFNASY TTYHDFLTMSENKREMKQWLEDPELASMFEEIIKTLTVFEDREMIKTRLSHHEATLGKHIIKKLTKKHYTGWGRLSKELIQGIRDKQSNKTILDYLINDDDFPHHRNRNFMQLINDDSL SFKKEIKKAQMITDTENLEEIVKELTGSPAIKKGILQSLKIVDEIVGIMGYEPANIVVEMARENQTTGRGLKSSRPRLKALEESLKDFGSQLLKEYPTDNSSLQKDRLYLYYLQNGRD YTGAPLDIHRLSDYDIDHIIPRSFTTDNSIDNKVLVSSKENRLKKDDVPSEKVVKKMKSFWYDLYSSKLISKRKLDNLTKIKLTEEDKAGFIKRQLVETRQITKHVAGILHHRFNKAED TNEPIRKVRIITLKSALVSQFRNRFGIYKVRETNEYHHAHDAYLNGVVGALALLKKYPQLAPEFVYGEYLKFNAHKANKATVKKEFYSNIMKFFESDTPVCDENGEIFWDKSKSIAQVKK VINHHHMNIVKKTEIQKGGFSKETVEPKKDSSKLLPRKNNWDPAKYGGLGSPNVAYTVAFTYEKGKAKRTNALEGITIMERERAFEQSPVLFLKNKGYEQAEIEMKLPKYALFELENG KRMVASNKEAQKANSFLLPEHLVTLLYHAKQYDEISHKESFDYVNEHHKEFSEVFARVLEFAGKYTLAEKNIEKLEKIYKENQTDDLAKLASSFVNLMQFNAMGAPADFKFFDVTIPRK RYTSLITEIWQSTIIHQSITGLYETRIRMGK	WP_002310644.1 [Enterococcus faecium]
217	MKKEYTIGLDIGTNSVGWSVLTDDYRLVSKKMKVAGNTEKSSTKKNFWGVRLFDEGQTAEARRSKRTARRLARRRQRILELQKIFAPEILKIDEHFFARLNESFLVLDEKKQSRHPVF ATIKQEKSYHQTVPTIYHLRQALADSSEKADIRLVYLAMAHLLKYRGHFLIEGELNTENSSVTETFRQFLSTYNQQFSEADDKQTEKLDEAVDCSFVFTEKMSKTKKAETLLKYFPHEK SNGYLSQFIKLMVGNQGNFKNVFGLEEEAKLQFSKETYEEDLEELLEKIGDDYIDLFVQAKNVYDAVLLSEILSDSTKNTRAKLSAGMIRRYDAHKRDLVLLKRFVKENLPKKYRAFFG DNSVNGYAGYIEGHATQEDFYKFVKKELTGIRGSEVFLTKIEQENFLRKQRTFDNGVIPHQIHLTELRAIIANQKKHYPFLKEEQEKLESLLTFKIPYYVGPLAKGQNSPFAWLIRKS EEKIKPWNLPEIVDMEGSAVRFIERMINTDMYIPHNKVLPKNSLLYQKFSIYNELTKVRYQDERGQMNYFSSIEKKEIFHELFEKNRKVTKKDLQEFLYLKYDIKHAELSGIEKAFNAS YTTYHDFLTMSENKREMKQWLEDPELASMFEEIIKTLTVFEDREMIKTRLSHHEATLGKHIIKKLTKKHYTGWGRLSKELIQGIRDKQSNKTILDYLINDDDFPHHRNRNFMQLINDDS LSFKKEIKKAQMITDTENLEEIVKELTGSPAIKKGILQSLKIVDEIVGIMGYEPANIVVEMARENQTTGRGLKSSRPRLKALEESLKDFGSQLLKEYPTDNSSLQKDRLYLYYLQNGRD MYTGAPLDIHRLSDYDIDHIIPRSFTTDNSIDNKVLVSSKENRLKKDDVPSEKVVKKMRSFWYDLYSSKLISKRKLDNLTKIKLTEEDKAGFIKRQLVETRQITKHVAGILHHRFNKAE DTNEPIRKVRIITLKSALVSQFRNRFGIYKVRETNEYHHAHDAYLNGVVGALALLKKYPQLAPEFVYGEYLKFNAHKANKATVKKEFYSNIMKFFESDTPVCDENGEIFWDKSKSIAQVK KVINHHHMNIVKKTEIQKGGFSKETVEPKKDSSKLLPRKNNWDPAKYGGLGSPNVAYTVAFTYEKGKAKRTNALEGITIMERERAFEQSPVLFLKNKGYEQAEIEMKLPKYALFELENG RKRMVASNKEAQKANSFLLPEHLVTLLYHAKQYDEISHKESFDYVNEHHKEFSEVFARVLEFAGKYTLAEKNIEKLEKIYKENQTDDLAKLASSFVNLMQFNAMGAPADFKFFDVTIPR KRYTSLITEIWQSTIIHQSITGLYETRIRMGK	WP_002314015.1 [Enterococcus faecium]
218	MKKEYTIGLDIGTNSVGWSVLTDDYRLVSKKMKVAGNTEKSSTKKNFWGVRLFDEGQTAEARRSKRTARRLARRRQRILELQKIFAPEILKIDEHFFARLNESFLVLDEKKQSRHPVF ATIKQEKSYHQTVPTIYHLRQALADSSEKADIRLVYLAMAHLLKYRGHFLIEGELNTENSSVTETFRQFLSTYNQQFSEADDKQTEKLDEAVDCSFVFTEKMSKTKKAETLLKYFPHEK SNGYLSQFIKLMVGNQGNFKNVFGLEEEAKLQFSKETYEEDLEELLEKIGDDYIDLFVQAKNVYDAVLLSEILSDSTKNTRAKLSAGMIRRYDAHKRDLVLLKRFVKENLPKKYRAFFG DNSVNGYAGYIEGHATQEDFYKFVKKELTGIRGSEVFLTKIEQENFLRKQRTFDNGVIPHQIHLTELRAIIANQKKHYPFLKEEQEKLESLLTFKIPYYVGPLAKGQNSPFAWLIRKS EEKIKPWNLPEIVDMEGSAVRFIERMINTDMYIPHNKVLPKNSLLYQKFSIYNELTKVRYQDERGQMNYFSSIEKKEIFHELFEKNRKVTKKDLQEFLYLKYDIKHAELSGIEKAFNAS YTTYHDFLTMSENKREMKQWLEDPELASMFEEIIKTLTVFEDREMIKTRLSHHEATLGKHIIKKLTKKHYTGWGRLSKELIQGIRDKQSNKTILDYLINDDDFPHHRNRNFMQLINDDS LSFKKEIKKAQMITDTENLEEIVKELTGSPAIKKGILQSLKIVDEIVGIMGYEPANIVVEMARENQTTGRGLKSSRPRLKALEESLKDFGSQLLKEYPTDNSSLQKDRLYLYYLQNGRD MYTGAPLDIHRLSDYDIDHIIPRSFTTDNSIDNKVLVSSKENRLKKDDVPSEKVVKKMKSFWYDLYSSKLISKRKLDNLTKIKLTEEDKAGFIKRQLVETRQITKHVAGILHHRFNKAE DTNEPIRKVRIITLKSALVSQFRNRFGIYKVRETNEYHHAHDAYLNGVVGALALLKKYPQLAPEFVYGEYLKFNAHKANKATVKKEFYSNIMKFFESDTPVCDENGEIFWDKSKSIAQVK KVINHHHMNIVKKTEIQKGGFSKETVEPKKDSSKLLPRKNNWDPAKYGGLGSPNVAYTVAFTYEKGKAKRTNALEGITIMERERAFEQSPVLFLKNKGYEQAEIEMKLPKYALFELENG RKRMVASNKEAQKANSFLLPEHLVTLLYHAKQYDEISHKESFDYVNEHHKEFSEVFARVLEFAGKYTLAEKNIEKLEKIYKENQTDDLAKLASSFVNLMQFNAMGAPADFKFFDVTIPR KRYTSLITEIWQSTIIHQSITGLYETRIRMGK	WP_002320716.1 [Enterococcus faecium]
219	MKKEYTIGLDIGTNSVGWSVLTDDYRLVSKKMKVAGNTEKSSTKKNFWGVRLFDEGQTAEARRSKRTARRLARRRQRILELQKIFAPEILKIDEHFFARLNESFLVLDEKKQSRHPVF ATIKQEKSYHQTVPTIYHLRQALADSSEKADIRLVYLAMAHLLKYRGHFLIEGELNTENSSVTETFRQFLSTYNQQFSEADDKQTEKLDEAVDCSFVFTEKMSKTKKAETLLKYFPHEK SNGYLSQFIKLMVGNQGNFKNVFGLEEAKLQFSKETYEEDLEELLEKIGDDYIDLFVQAKNVYDAVLLSEILSDSTKNTRAKLSAGMIRRYDAHKRDLVLLKRFVKENLPKKYRAFFGD NSVNGYAGYIEGHATQEDFYKFVKKELTGIRGSEVFLTKIEQENFLRKQRTFDNGVIPHQIHLTELRAIIANQKKHYPFLKEEQEKLESLLTFKIPYYVGPLAKGQNSPFAWLIRKSE EKIKPWNLPEIVDMEGSAVRFIERMINTDMYIPHNKVLPKNSLLYQKFSIYNELTKVRYQDERGQMNYFSSIEKKEIFHELFEKNRKVTKKDLQEFLYLKYDIKHAELSGIEKAFNASY TTYHDFLTMSENKREMKQWLEDPELASMFEEIIKTLTVFEDREMIKTRLSHHEATLGKHIIKKLTKKHYTGWGRLSKELIQGIRDKQSNKTILDYLINDDDFPHHRNRNFMQLINDDSL SFKKEIKKAQMITDTENLEEIVKELTGSPAIKKGILQSLKIVDEIVGIMGYEPANIVVEMARENQTTGRGLKSSRPRLKALEESLKDFGSQLLKEYPTDNSSLQKDRLYLYYLQNGRDM YTGAPLDIHRLSDYDIDHIIPRSFTTDNSIDNKVLVSSKENRLKKDDVPSEKVVKKMRSFWYDLYSSKLISKRKLDNLTKIKLTEEDKAGFIKRQLVETRQITKHVAGILHHRFNKAED TNEPIRKVRIITLKSALVSQFRNRFGIYKVRETNEYHHAHDAYLNGVVGALALLKKYPQLAPEFVYGEYLKFNAHKANKATVKKEFYSNIMKFFESDTPVCDENGEIFWDKSKSIAQVKK VINHHHMNIVKKTEIQKGGFSKETVEPKKDSSKLLPRKNNWDPAKYGGLGSPNVAYTVAFTYEKGKAKRTNALEGITIMERERAFEQSPVLFLKNKGYEQAEIEMKLPKYALFELENG KRMVASNKEAQKANSFLLPEHLVTLLYHAKQYDEISHKESFDYVNEHHKEFSEVFARVLEFAGKYTLAEKNIEKLEKIYKENQTDDLAKLASSFVNLMQFNAMGAPADFKFFDVTIPRK RYTSLITEIWQSTIIHQSITGLYETRIRMGK	WP_002330729.1 [Enterococcus faecium]
220	MKKEYTIGLDIGTNSVGWSVLTDDYRLVSKKMKVAGNTEKSSTKKNFWGVRLFDEGQTAEARRSKRTARRLARRRQRILELQKIFAPEILKIDEHFFARLNESFLVLDEKKQSRHPVF ATIKQEKSYHQTVPTIYHLRQALADSSEKADIRLVYLAMAHLLKYRGHFLIEGELNTENSSVTETFRQFLSTYNQQFSEADDKQTEKLDEAVDCSFVFTEKMSKTKKAETLLKYFPHEK SNGYLSQFIKLMVGNQGNFKNVFGLEEEAKLQFSKETYEEDLEELLEKIGDDYIDLFVQAKNVYDAVLLSEILSDSTKNTRAKLSAGMIRRYDAHKRDLVLLKRFVKENLPKKYRAFFG DNSVNGYAGYIEGHATQEDFYKFVKKELTGIRGSEVFLTKIEQENFLRKQRTFDNGVIPHQIHLTELRAIIANQKKHYPFLKEEQEKLESLLTFKIPYYVGPLAKGQNSPFAWLIRKS EEKIKPWNLPEIVDMEGSAVRFIERMINTDMYMPHNKVLPKNSLLYQKFSIYNELTKVRYQDERGQMNYFSSIEKKEIFHELFEKNRKVTKKDLQEFLYLKYDIKHAELSGIEKAFNAS YTTYHDFLTMSENKREMKQWLEDPELASMFEEIIKTLTVFEDREMIKTRLSHHEATLGKHIIKKLTKKHYTGWGRLSKELIQGIRDKQSNKTILDYLINDDDFPHHRNRNFMQLINDDS LSFKKEIKKAQMITDTENLEEIVKELTGSPAIKKGILQSLKIVDEIVGIMGYEPANIVVEMARENQTTGRGLKSSRPRLKALEESLKDFGSQLLKEYPTDNSSLQKDRLYLYYLQNGRD MYTGAPLDIHRLSDYDIDHIIPRSFTTDNSIDNKVLVSSKENRLKKDDVPSEKVVKKMRSFWYDLYSSKLISKRKLDNLTKIKLTEEDKAGFIKRQLVETRQITKHVAGILHHRFNKAE DTNEPIRKVRIITLKSALVSQFRNRFGIYKVRETNEYHHAHDAYLNGVVGALALLKKYPQLAPEFVYGEYLKFNAHKANKATVKKEFYSNIMKFFESDTPVCDENGEIFWDKSKSIAQVK KVINHHHMNIVKKTEIQKGGFSKETVEPKKDSSKLLPRKNNWDPAKYGGLGSPNVAYTVAFTYEKGKAKRTNALEGITIMERERAFEQSPVLFLKNKGYEQAEIEMKLPKYALFELENG RKRMVASNKEAQKANSFLLPEHLVTLLYHAKQYDEISHKESFDYVNEHHKEFSEVFARVLEFAGKYTLAEKNIEKLEKIYKENQTDDLAKLASSFVNLMQFNAMGAPADFKFFDVTIPR KRYTSLITEIWQSTIIHQSITGLYETRIRMGK	WP_002335161.1 [Enterococcus faecium]

221	MKKEYTIGLDIGTNSVGWSVLTDDYRLVSKKMKVAGNTEKSSTKKNFWGVRLFDEGQTAEARRSKRTARRRLARRRQRILELQKIFAPEILKIDEHFFARLNESFLVLDEKKQSRHPVF ATIKQEKSYHQTVPTIYHLRQALADSSEKADIRLVYLAMAHLLKYRGHFLIEGELNTENSSVTETFRQFLSTYNQQFSEADDKQTEKLDEAVDCSFVFTEKMSKTKKAETLLKYFPHEK SNGYLSQFIKLMVGNQGNFKNVFGLEEEAKLQFSKETYEEDLEELLEKIGDDYIDLFVQAKNVYDAVLLSEILSDSTKNTRAKLSAGMIRRYDAHKEDLIVLLKRFVKRENLPKKYRAFFG DNSVNGYAGYIEGHATQEDFYKFVKKELTGIRGSEVFLTKIEQENFLRKQRTFDNGVIPHQIHLTELRAIIANQKKHYPFLKEEQEKILESLLTFKIPYYVGPLAKKQENSPFAWLIRKS EEKIKPWNLPEIVDMEGSAVRFIERMINTDMYIPHNKVLPKNSLLYQKFSIYNELTKVRYQDERGQMNYFSSIEKKEIFHELFEKNRKVTKKDLQEFLYLKYDIKHAELSGIEKAFNAS YTTYHDFLTMSENKREMKQWLEDPELASMFEEIKTLTVFEDREMIKTRLSHHEATLGKHIIKKLTKKHYTGWGRLSKELIQGIRDKQSNKTILDYLINDDDFPHHRNRNFMQLINDDS LSFKKEIKKAQMTTDTENLEEIVKELTGSPAIKKGILQSLKIVDEIVGIMGYEPANIVVEMARENQTTGRGLKSSRPRLKALEESLKDFGSQLLKEYPTDNSSLQKDRLYLYYLQNGRD MYTGAPLDIHRLSDYDIDHIIPRSFTTDNSIDNKVLVSSKENRLKKDDVPSEKVVKKMKSFWYDLYSSKLISKRKLDNLTKIKLTEEDKAGFIKRQLVETRQITKHVAGILHHRFNKAE DINEPIRKVRIITLKSALVSQFRNRFGIYKVRELNEYHHAHDAYLNGVVAALLKKYPQLAPEFVYGEYLKFNAHKANKATVKKEFYSNIMKFFESDTPVCDENGEIFWDKSKSIAQVK KVINHHHMNIVKKTEIQKGGFSKETVEPKKDSSKLLPRKNNWDPTKYGGLGSPNVAYTVAFTYEKGKARKRTNALEGITIMEREAFEQSPVLFLKNKGYEQAEIEMKLPKYALFELENG RKRMVASNKEAQKANSFLLPEHLVTLLYHAKQYDEISHKESFDYVNEHHKEFSEVFARVLELEFAGKYTLAEKNIEKLEKIYKENQTDDLAKLASSFVNLMQFNAMGAPADFKFFDVTIPR KRYTSLTEIWQSTIIHQSITGLYETRIRMGK	WP_002345439.1 [Enterococcus faecium]
222	MKKEYTIGLDIGTNSVGWSVLTDDYRLVSKKMKVAGNTEKSSTKKNFWGVRLFDEGQTAEARRSKRTARRRLARRRQRILELQKIFAPEILKIDEHFFARLNESFLVLDEKKQSRHPVF ATIKQEKSYHQTVPTIYHLRQALADSSEKADIRLVYLAMAHLLKYRGHFLIEGELNTENSSVTETFRQFLSTYNQQFSEADDKQTEKLDEAVDCSFVFTEKMSKTKKAETLLKYFPHEK SNGYLSQFIKLMVGNQGNFKNVFGLEEEAKLQFSKETYEEDLEELLEKIGDDYIDLFVQAKNVYDAVLLSEILSDSTKNTRAKLSAGMIRRYDAHKEDLIVLLKRFVKRENLPKKYRAFFG DNSVNGYAGYIEGHATQEDFYKFVKKELTGIRGSEVFLTKIEQENFLRKQRTFDNGVIPHQIHLTELRAIIANQKKHYPFLKEEQEKILESLLTFKIPYYVGPLAKKQENSPFAWLIRKS EEKIKPWNLPEIVDMEGSAVRFIERMINTDMYMPHNKVLPKNSLLYQKFSIYNELTKVRYQDERGQMNYFSSIEKKEIFHELFEKNRKVTKKDLQEFLYLKYDIKHAELSGIEKAFNAS YTTYHDFLTMSENKREMEQWLEDPELASMFEEIKTLTVFEDREMIKTRLSHHEATLGKHIIKKLTKKHYTGWGRLSKELIQGIRDKQSNKTILDYLINDDDFPHHRNRNFMQLINDDS LSFKKEIKKAQMTTDTENLEEIVKELTGSPAIKKGILQSLKIVDEIVGIMGYEPANIVVEMARENQTTGRGLKSSRPRLKALEESLKDFGSQLLKEYPTDNSSLQKDRLYLYYLQNGRD MYTGAPLDIHRLSDYDIDHIIPRSFTTDNSIDNKVLVSSKENRLKKDDVPSEKVVKKMKSFWYDLYSSKLISKRKLDNLTKIKLTEEDKAGFIKRQLVETRQITKHVAGILHHRFNKAE DINEPIRKVRIITLKSALVSQFRNRFGIYKVRELNEYHHAHDAYLNGVVAALLKKYPQLAPEFVYGEYLKFNAHKANKATVKKEFYSNIMKFFESDTPVCDENGEIFWDKSKSIAQVK KVINHHHMNIVKKTEIQKGGFSKETVEPKKDSSKLLPRKNNWDPTKYGGLGSPNVAYTVAFTYEKGKARKRTNALEGITIMEREAFEQSPVLFLKNKGYEQAEIEMKLPKYALFELENG RKRMVASNKEAQKANSFLLPEHLVTLLYHAKQYDEISHKESFDYVNEHHKEFSEVFARVLELEFAGKYTLAEKNIEKLEKIYKENQTDDLAKLASSFVNLMQFNAMGAPADFKFFDVTIPR KRYTSLTEIWQSTIIHQSITGLYETRIRMGK	WP_002312694.1 [Enterococcus faecium]
223	MKKEYTIGLDIGTNSVGWSVLTDDYRLVSKKMKVAGNTEKSSTKKNFWGVRLFDEGQTAEARRSKRTARRRLARRRQRILELQKIFAPEILKIDEHFFARLNESFLVPDEKKQSRHPVF ATIKQEKSYHQTVPTIYHLRQALADSSEKADIRLVYLAMAHLLKYRGHFLIEGELNTENSSVTETFRQFLSTYNQQFSEAGDKQTEKLDEAVDCSFVFTEKMSKTKKAETLLKYFPHEK SNGYLSQFIKLMVGNQGNFKNVFGLEEEAKLQFSKETYEEDLEELLEKIGDDYIDLFVQAKNVYDAVLLSEILSDSTKNTRAKLSAGMIRRYDAHKEDLIVLLKRFVKRENLPKKYRAFFG DNSVNGYAGYIEGHATQEAFYKFVKKELTGIRGSEVFLTKIEQENFLRKQRTFDNGVIPHQIHLSELRAIIANQKKHYPFLKEEQEKLESLLTFKIPYYVGPLARGNRDFAWLTRNSDEAIRPW EEKIKPWNLPEIVDMEGSAVRFIERMINTDMYMPHNKVLPKNSLLYQKFSIYNELTKVRYQDERGQMNYFSSIEKKEIFHELFEKNRKVTEKDIIHYLHNVDGYDGIELKGIEKQFNASLSTYHD YTTYHDFLTMSENKREMKQWLEDPELASMFEEIKTITVFEDREMIKTRLSHHEATLGKHIIKKLTKKHYTGWGRLSKELIQGIRDKQSNKTILDYLINDDDFPHHRNRNFMQLINDDS LSFKKEIKKAQMITDTENLEEIVKELTGSPAIKKGILQSLKIVDEIVGIMGYEPANIVVEMARENQTTGRGLKSSRPRLKALEESLKDFGSQLLKEYPTDNSSLQKDRLYLYYLQNGRD MYTGAPLDIHRLSDYDIDHIIPRSFTTDNSIDNKVLVSSKENRLKKDDVPSEKVVKKMKSFWYDLYSSKLISKRKLDNLTKIKLTEEDKAGFIKRQLVETRQITKHVAGILHHRFNKAE DINDPIRKVRIITLKSALVSQFRNRFGIYKVRELNEYHHAHDAYLNGVVAALLKKYPQLAPEFVYGEYLKFNAHKANKATVKKEFYSNIMKFFESDTPVCDENGEIFWDKSKSIAQVK KVINHHHMNIVKKTEIQKGGFSKETVEPKKDSSKLLPRKNNWDPAKYGGLGSPNVAYTVAFTYEKGKARKRTNALEGITIMEREAFEQSPVLFLKNKGYEQAEIEMKLPKYALFELENG RKRMVASNKEAQKANSFLLPEHLVTLLYHAKQYDEISHKESFDYVNEHHKEFSEVFARVLELEFAGKYTLAEKNIEKLEKIYKENQTDDLAKLASSFVNLMQFNAMGAPADFKFFDVTIPR KRYTSLTEIWQSTIIHQSITGLYETRIRMGK	WP_045635197.1 [Streptococcus mitis]
224	MKKGYSIGLDIGTNSVGFAVITDDYKVPSKKMKVLGNTDKRFIKKNLIGALLFDEGTTAEARRLKRTARRRYTRRKNRLRYLQEIFSEEMSKVDSSFFHRLDDSFLIPEDKRESKYPIF ATLTEEKEYHKQFPTIYHLRKQLADSKEKTDLRLIYLALAHMIKYRGHFLYEEAFDIKNNDIQKIFNEFISIYDNTFEGSSLSGQNAQVEAIFTDKISKSAKRERVLKLFPDEKSTGLF SEFLKLIVGNQADFKKHFDLEDKAPLQFSKDTYDEDLENLLGQIGDDFTDLFVSAKKLYDAILLSGILTVTDPSTKAPLSASMIERYENHQNDLAALKQFIKNNLPEKYDEVFSDQSKD GYAGYIDGKTTQEDFYKYIKNLLSKFEGTDYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMNAILRRQGEYYPFLKDNKEKIEKILTFRIPYYVGPLARGNRDFAWLTRNSDEAIRPW NFEEIVDKASSADFINKMTNYDLYLPEEKVLPKHSLLYETFAVYNELTKVKFIAEGLRDYQFLDSGQKKQIVNQLFKENRKVTEKDIIHYLHNVDGYDGIELKGIEKQFNASLSTYHD LLKIIKDKEFMDDAKNEAILENIVHTLTIFEDREMIKQRLAQYDSLFDEKVIKALTRRHYTGWGKLSAKLINGICDKQTGNTILDYLIDDGKINRNFMQLINDDGLSFKEIIQKAQVIG KTDDVKQVVQELSGSPAIKKGILQSIKIVDELVKVMGHAPESIVIEMARENQTTARGKKNSQQRVKRIEDSLKILASGLDSNILKENPTDNNQLQNDRLFLYYLQNGKDMYTGEALDIN QLSSYDIDHIIPQAFIKDDSLDNRVLTSSKDNRGKSDNVPSIEVVQKRKAFWQQLLDSKLISERKFNNLTKAERGGLDERDKVGFIKRQLVETRQITKHVAQILDARYNTEVNEKDKKN RIVKIITLKSNLVSNFRKEFRLYKVRETNDYHHAHDAYLNAVVAKAILKKYPKLEPEFVYGEYQKYDLKRYISRSKDPKEVEKATEKYFFYSNLLNFFKEEVHYADGTIVKRENIEYSK DTGEIAWNKEKDFAIIKKVLSLPQVNIVKKREVQTGGFSKESILPKGNSDKLIPRKTKDILLDTKYGGFDSPVIAYSILLIADIEKGKAKKLKTVKTLVGITIMEKAAFEENPITFLE NKGYHNVRKENILCLPKYSLFELENGRRRLLASAKELQKGNEIVLPVYLTTLLYHSKNVHKLDEPGHLEYIQKHRNEFKDLLNIVSEFSQKYVLADANLEKIKSLYADNEQADIEILAN SFINLLTFTALGAPAAFKFFGKDIDRKRYTTVSEIINATLIHQSITGLYETWIDLSKLGED	WP_045635197.1 [Streptococcus mitis]
225	MKKGYSIGLDIGTNSVGFAVITDDYKVPSKKMKVLGNTDKRFIKKNLIGALLFDEGTTAEARRLKRTARRYTRRKNRLRYLQEIFSEEMSKVDSSFFHRLDDSFLIPEDKRESKYPIF ATLTEEKEYHKQFPTIYHLRKQLADSKEKTDLRLIYLALAHMIKYRGHFLYEEAFDIKNNDIQKIFNEFISIYDNTFEGSSLSGQNAQVEAIFTDKISKSAKRERVLKLFPDEKSTGLF SEFLKLIVGNQADFKKHFDLEDKAPLQFSKDTYDEDLENLLGQIGDDFTDLFVSAKKLYDAILLSGILTVTDPSTKAPLSASMIERYENHQNDLAALKQFIKNNLPEKYDEVFSDQSKD GYAGYIDGKTTQEDFYKYIKNLLSKFEGTDYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMNAILRRQGEYYPFLKDNKEKIEKILTFRIPYYVGPLARGNRDFAWLTRNSDEAIRPW NFEEIVDKASSADFINKMTNYDLYLPEEKVLPKHSLLYETFAVYNELTKVKFIAEGLRDYQFLDSGQKKQIVNQLFKENRKVTEKDIIHYLHNVDGYDGIELKGIEKQFNASLSTYHD LLKIIKDKEFMDDAKNEAILENIVHTLTIFEDREMIKQRLAQYDSLFDEKVIKALTRRHYTGWGKLSAKLINGICDKQTGNTILDYLIDDGKINRNFMQLINDDGLSFKEIIQKAQVIG KTDDVKQVVQELSGSPAIKKGILQSIKIVDELVKVMGHAPESIVIEMARENQTTARGKKNSQQRVKRIEDSLKILASGLDSNILKENPTDNNQLQNDRLFLYYLQNGKDMYTGEALDIN QLSSYDIDHIIPQAFIKDDSLDNRVLTSSKDNRGKSDNVPSIEVVQKRKAFWQQLLDSKLISERKFNNLTKAERGGLDERDKVGFIKRQLVETRQITKHVAQILDARYNTEVNEKDKKN RIVKIITLKSNLVSNFRKEFRLYKVRETNDYHHAHDAYLNAVVAKAILKKYPKLEPEFVYGEYQKYDLKRYISRSKDPKEVEKATEKYFFYSNLLNFFKEEVHYADGTIVKRENIEYSK DTGEIAWNKEKDFAIIKKVLSLPQVNIVKKREVQTGGFSKESILPKGNSDKLIPRKTKDILLDTKYGGFDSPVIAYSILLIADIEKGKAKKLKTVKTLVGITIMEKAAFEENPITFLE NKGYHNVRKENILCLPKYSLFELENGRRRLLASAKELQKGNEIVLPVYLTTLLYHSKNVHKLDEPGHLEYIQKHRNEFKDLLNIVSEFSQKYVLADANLEKIKSLYADNEQADIEILAN SFINLLTFTALGAPAAFKFFGKDIDRKRYTTVSEIINATLIHQSITGLYETWIDLSKLGED	WP_045635197.1 [Streptococcus mitis]

226	MKKKYAIGIDIGTNSVGWSVVTDDYKVPSKKMKVFGNTEKRYIKKNLIGTLLFDEGNTAENRRLKRTARRYTRRRNRILYLQEIFAEEINKIDDSFFQRLDDSFLIVEDKQGSKHPIF GTLQEEKEYHKQFPTIYHLRKQLADSSQKADIRLIYLALAHIIKYRGHFLFEGDLKSENKDVQHLFNDFVEMFDKTVEGSYLSENLPNVADVLVEKVSKSKRRLENILHYFPPNEKKNGLF GNFLALAIGLQPNFKTNFELAEDAKIQFSKETYEEDLEELLGKIGDDYADLFIATKSLYDGILLAGILSTTDSITKAPLSASMVNRYEEHQKDLALLKNFIHQNLSDSYKEVFNDKLKD GYAGYIEGKTTQENFYRFIKKAIEKIEGSNYFIDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAIIRRQAEFYPFLVENQDKIEKILTFRIPYYVGPLARGKSEFAWLNRKSDEKITPWN NFDEMVDKETSAENFITRMTNYDQYLPDQKVLPKHSLLYEKFAVYNELTKVRYVTEQGKSFFFDANMKQEIFDGVFKVYRKVTKEKLMDFLGKEFDEFRIVDLLGLDKDNKSFNASLGT YHDLKKIVSKDLLDNPENEDILENVVLTLTLFEDREMIRKRLEKYKDVLTEEQRKKLERRHYTGWGRLSAKLINGIRDKVTRKTILDYLIDDGTSNRNFMQLINDTLSFVDEIRLAQG SGEAEDYRAEVQNLAGSPAIKKGILQSLKIVDELIEVMGYDPEHIVVEMARENQFTNQGRRNSQQRYKKIENAIKNLDSNLNSKILIKEYPTNNQALQNDRLFLYYLQNGKDMYTDEELD IDQLSQYDIDHIIPQAFIKDDSLDNKVLTKSAKNRGKSDDVPSLEIVKHKKKNFWKQLLDSQLISQRKFDNLTKAERGGLTNEDKARFIQRQLVETRQITKHVARILDTRFNTKLDEAGN RIRDPKVNIITLKSNLVSQFRKDYQLYKVREINNYHHAHDAYLNAVVATALLKKYPQLAPEFVYGDYPKYNSYKSRKSATEKVLFYSNIMNFFRRVLVYSKTGEVRIRPVIEVNKETGE IVWDKKSDFRTVRKVLSYPQVNVVKKVEMQTGGFSKESILQHGDSDKLIPRKTEKFYLDTKKYGGFDSPTIAYSVLLIADIEKGKAKKLKRVKELIGITIMERMAFEKNPIEFLEHKGY KNILEKNIIKLPKYSLFELENGRRRLLASAKELQKGNEMILPPHLVTLLYHSSNIHKITTEPIHLNYVNKNKHEFKELLRHISDFSTRYIILAQDRLSKIEELYDKNDGDDISDLTSSFVN LLTFTAIGAPAAFKFLGSVIDRKRYTSIAEILEATLIHQSVTGLYETRIDLSKLGGD	WP_044680361.1 [Streptococcus suis]
227	MKKKYAIGIDIGTNSVGWSVVTDDYKVPSKKMKVFGNTEKRYIKKNLIGTLLFDEGNTAENRRLKRTARRYTRRRNRILYLQEIFAEEINKIDDSFFQRLDDSFLIVEDKQGSKHPIF GTLQEEKEYHKQFPTIYHLRKQLADSSQKADIRLIYLALAHIIKYRGHFLFEGDLKSENKDVQHLFNDFVEMFDKTVEGSYLSENLPNVADVLVEKVSKSKRRLENILHYFPPNEKKNGLF GNFLTLAIGLQPNFKTNFELAEDAKIQFSKETYEEDLEELLGKIGDDYADLFIATKSLYDGILLAGILSTTDSITKAPLSSSMVNRYEEHQKDLALLKNFIHQNLSDSYKEVFNDKLKD GYAGYIEGKTTQENFYRFIKKAIEKIEGSNYFIDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAIIRRQAEFYPFLVENQDKIEKILTFRIPYYVGPLARGKSEFAWLNRKSDEKITPWN NFDEMVDKETSAENFITRMTNYDQYLPDQKVLPKHSLLYEKFAVYNELTKVRYVTEQGKSFFFDANMKQEIFDGVFKVYRKVTKEKLMDFLGKEFDEFRIVDLLGLDKDNKSFNASLGT YHDLKKIVSKDLLDNPENEDILENVVLTLTLFEDREMIRKRLEKYKDVLTEEQRKKLERRHYTGWGRLSAKLINGIRDKVTRKTILDYLIDDGTSNRNFMQLINDTLSFVDEIRLAQG SGEAEDYRAEVQNLAGSPAIKKGILQSLKIVDELIEVMGYDPEHIVVEMARENQFTNQGRRNSQQRYKKIENAIKNLDSNLNSKILIKEYPTNNQALQNDRLFLYYLQNGKDMYTDEELD IDQLSQYDIDHIIPQAFIKDDSLDNKVLTKSAKNRGKSDDVPSLEIVKHKKKNFWKQLLDSQLISQRKFDNLTKAERGGLTNEDKARFIQRQLVETRQITKHVARILDTRFNTKLDEAGN RIRDPKVNIITLKSNLVSQFRKDYQLYKVREINNYHHAHDAYLNAVVATALLKKYPQLAPEFVYGDYPKYNSYKSRKSATEKVLFYSNIMNFFRRVLVYSKTGEVRIRPVIEVNKETGE IVWDKKSDFRTVRKVLSYPQVNVVKKVEMQTGGFSKESILQHGDSDKLIPRKTEKFYLDTKKYGGFDSPTIAYSVLLIADIEKGKAKKLKRVKELIGITIMERMAFEKNPIEFLEHKGYKN KNILEKNIIKLPKYSLFELENGRRRLLASAKELQKGNEMILPPHLVTLLYHSSNIHKITTEPIHLNYVNKNKHEFKELLRHISDFSTRYIILAQDRLSKIEELYDKNDGDDISDLTSSFVNLL LLTFTAIGAPAAFKFLGSVIDRKRYTSIAEILEATLIHQSVTGLYETRIDLSKLGGD	WP_044676715.1 [Streptococcus suis]
228	MKKKYAIGIDIGTNSVGWSVVTDDYKVPSKKMKVFGNTEKRYIKKNLIGTLLFDEGNTAENRRLKRTARRYTRRRNRILYLQEIFAEEINKIDDSFFQRLDDSFLIVEDKQGSKHPIF GTLQEEKKYHKQFPTIYHLRKQLADSSQKADIRLIYLALAHIIKYRGHFLFEGDLKSENKDVQHLFNDFVEMFDKTVEGSYLSENLPNVADVLVEKVSKSKRRLENILHYFPPNEKKNGLF GNFLALAIGLQPNFKTNFELAEDAKIQFSKETYEEDLEELLGKIGDDYADLFIATKSLYDGILLAGILSTTDSITKAPLSSSMVNRYEEHKKDLALLKNFIHQNLSDSYKEVFNDKLKD GYAGYIEGKTTQENFYRFIKKAIEKIEGSDYFIDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAIIRRQAEFYPFLVENQDKIEKILTFRIPYYVGPLARGKSEFAWLNRKSDEKITPWN NFDEMVDKETSAENFITRMTNYDQYLPDQKVLPKHSLLYEKFAVYNELTKVKFIAEGMRDYQFLDSGQKKDIVKTLFKTKRKVTAKDIKAYLENSNGYAGVELKGLEEQFNASLPTYHD LLKIILRDKAFIDAEENQEILEDIVLTLTLFEDREMIRKRLEKYKDILTEEQRKKLERRHYTGWGRLSAKLINGILDKVTRKTILGYLIDDGTSNRNFMQLINDTLSFVDEIRLAQGSG EAEDYRAEVQNLAGSPAIKKGILQSLKIVDELIEVMGYDPEHIVVEMARENQFTNQGRRNSQQRYKKIENAIKNLDSNLNSKILIKEYPTNNQALQNDRLFLYYLQNGKDMYTDEELDID QLSQYDIDHIIPQAFIKDDSLDNKVLTKSAKNRGKSDDVPSLEIVKHKKKNFWKQLLDSQLISQRKFDNLTKAERGGLTNEDKARFIQRQLVETRQITKHVARILDTRFNTKLDEAGNRI RDPKVNIITLKSNLVSQFRKDYQLYKVREINNYHHAHDAYLNAVVATALLKKYPQLAPEFVYGDYPKYNSYKSRKSATEKVLFYSNIMNFFRRVLVYSKTGEVRIRPVIEVNKETGEIV WDKKSDFRTVRKVLSYPQVNVVKKVEMQTGGFSKESILQHGDSDKLIPRKTEKFYLDTKKYGGFDSPTIAYSVLLIADIEKGKAKKLKRVKELIGITIMERMAFEKNPIEFLEHKGYKN ILEKNIIKLPKYSLFELENGRRRLLASAKELQKGNEMILPPHLVTLLYHSSNIHKITTEPIHLNYVNKNKHEFKELLRHISDFSTRYIILAQDRLSKIEELYDKNDGDDISDLTSSFVNLL TFTAIGAPAAFKFLGSVIDRKRYTSIAEILEATLIHQSVTGLYETRIDLSKLGGD	WP_044681799.1 [Streptococcus suis]
229	MKKKYAIGIDIGTNSVGWSVVTDDYKVPSKKMKVFGNTEKRYIKKNLIGTLLFDEGNTAENRRLKRTARRYTRRRNRILYLQEIFAEEINKIDDSFFQRLDDSFLIVEDKQGSKHPIF GTLQEEKKYHKQFPTIYHLRKQLADSSQKADIRLIYLALAHIIKYRGHFLFEGDLKSENKDVQHLFNDFVEMFDKTVEGSYLSENLPNVADVLVEKVSKSKRRLENILHYFPPNEKKNGLF GNFLALAIGLQPNFKTNFELAEDAKIQFSKETYEEDLEELLGKIGDDYADLFIATKSLYDGILLAGILSTTDSITKAPLSSSMVNRYEEHKKDLALLKNFIHQNLSDSYKEVFNDKLKD GYAGYIEGKTTQENFYRFIKKAIEKIEGSDYFIDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAIIRRQAEFYPFLVENQDKIEKILTFRIPYYVGPLARGKSEFAWLNRKSDEKITPWN NFDEMVDKETSAENFITRMTNYDQYLPDQKVLPKHSLLYEKFAVYNELTKVKFIAEGMRDYQFLDSGQKKDIVKTLFKTKRKVTAKDIKAYLENSNGYAGVELKGLEEQFNASLPTYHD LLKIILRDKAFIDAEENQEILEDIVLTLTLFEDREMIRKRLEKYKDILTEEQRKKLERRHYTGWGRLSAKLINGILDKVTRKTILGYLIDDGTSNRNFMQLINDTLSFVDEIRLAQGSG EAEDYRAEVQNLAGSPAIKKGILQSLKIVDELIEVMGYDPEHIVVEMARENQFTNQGRRNSQQRYKKIENAIKNLDSNLNSKILIKEYPTNNQALQNDRLFLYYLQNGKDMYTDEELDID QLSQYDIDHIIPQAFIKDDSLDNKVLTKSAKNRGKSDDVPSLEIVKHKKKNFWKQLLDSQLISQRKFDNLTKAERGGLTNEDKARFIQRQLVETRQITKHVARILDTRFNTKLDEAGNRI RDPKVNIITLKSNLVSQFRKDYQLYKVREINNYHHAHDAYLNAVVATALLKKYPQLAPEFVYGDYPKYNSYKSRKSATEKVLFYSNIMNFFRRVLVYSKTGEVRIRPVIEVNKETGEIV WDKKSDFRTVRKVLSYPQVNVVKKVEMQTGGFSKESILQHGDSDKLIPRKTEKFYLDTKKYGGFDSPTIAYSVLLIADIEKGKAKKLKRVKELIGITIMERMAFEKNPIEFLEHKGYKN ILEKNIIKLPKYSLFELENGRRRLLASAKELQKGNEMILPPHLVTLLYHSSNIHKITTEPIHLNYVNKNKHEFKELLRHISDFSTRYIILAQDRLSKIEELYDKNDGDDISDLTSSFVNLL TFTAIGAPAAFKFLGSVIDRKRYTSIAEILEATLIHQSVTGLYETRIDLSKLGGD	WP_044674937.1 [Streptococcus suis]
230	MKKPYSIGLDIGTNSVGWAVITDDYKVPSKKMKVLGNTDKHFIKKNLIGALLFDSGETAEDRRLKRTSRRRYTRRKNRLRYLQEIFSEEISKLDSSFFHRLDDSFLVPEDKRGSKYPIF ATLEEEKEYHKKFPTIYHLRKHLADSKEKTDLRLIYLALAHMIKYRGHFLYEESFDIKNNDIQKIFNEFISIYDNTFEGSSLSGQNAQVEAIFTDKISKSTKRERVLKLFSDEKSTGLF SEFLKLIVGNQADFKKHFDLEEKAPLQFSKDTYDEDLENLLGQIGDGFADLFIVAKKLYDAVLLSGILTVTDPSTKAPLSASMIERYENHQEDLAALKQFIKNNLSEKYAEVFSDQSKD GYAGFIDGKTTQEAFYKYIKNLLSKLEGADYFLNKIEREDFLRKQRTFDNGSIPHQIHLQEMNAIIRRQAEFYPFLDSGGKKQVNQLFKDKRKVTEKDIIHYLHNVDGYDGIELKGIEKQFNASISTYHD LLKIIKDKEFMDNPKNGEIILENIIHTLTIFEDREMIKQRLAQYDTLFDEKVIKALTRRHYTGWGKLSAKLINGIRDKQTGKTILEYLIDDGDCNRNFMQLINDDGLSFKEIIQKAQVVG KTDDVKQVVQEIPGSPAIKKGILQSIKIVDELVKVMGHNPESIVVEMARENQTTAKGKKNSQQRYKRIEDALKNLAGQILKEHPVDNQQLQNDRLYLYYLQNGKDMYTGKALDIN QLSNYDIDHIIPQAFIKDDSLDNKVLTRSDKNRGKSDNVPSEEVVQKMKAFWQQLLDSKLISERKFDNLTKAERGGLDERDKVGFIKRQLVETRQITKHVAQLLDTRFNTEVENEQKI RTVKIITLKSNLVSNFRKEFGLYKVREINDYHHAHDAYLNAVVAKALIKKYPKLESEFVYGDYQKYDLKRYILKSKASNTIDKATEKYFFYSNLLNFFKEKVRYADGTIKKRENEIEYSN DTGEIAWNKEKDFATIKKVLSLPQVNIVKKTEEQTVGQNGGLFDNNIVSKKKVVDASKLIPKSSLSPEKYGGYARPTIAYSVLVIADIEKGKAKKLKRIKEIVGITIQDKKKFESN PVTYLEECGYKNINSNIILPKYSLFELENGQRRLLASSIEIQKGNELILPYHLTALLYHAQRINKISEPIHKQYVEAHQNEFKELLTTIISLSKKYIQKPNVELLLQQAFDQADKDI YQLSESFISLLKLTSFGAPGAFKFLGVEISQSSVRYKPNSQFLDTTLIHQSITGLYETRIDLSKLGED	WP_002906454.1 [Streptococcus sanguinis]

231	MKKPYSIGLDIGTNSVGWAVVMEDYKVPSKKMKVLGNTDKQSIKKNLIGALLFDSGETAVERRINRTTSRRYDRRRNRIRYLQHIFAEEMNRADENFFHRLKESFFVEEDKTYSKYPIF GTLEEEKNYHKNYPTIYHLRKTLADTPDKMDIRLIYLALAHIIKYRGHFLIEGDLDIENIGIQDSFKSFIEEYNTQFGTKLDSTTKVEAIFTENSSKAKRVETILGLFPDETAAGNLDK FLKIMLGNQADFKKVFDLEEKITLQFSKDSYEEDLELLLSKIDEEYAALFDLAKKVYDAVLLSNILTVKEKNTKAPLSASMIKRYEEHHKDDLKAFKRFFRERLPEKYETMFKDLITKPSY AAYVSGLYKKDAKRGLVPTSKRVTEDDFYKFSKGLLIDVEGAEYFLEKIEREDFLRKQRTFDNGAIPNQVHVKELQAIILNQSKYYPFLAENKEKIEKILITFRIPYYVGPLARGNSSFA WLQRKSDEAIRPWNFEQVVDMETSASRFIERMTLHDLYLPDEKVLPRHSLIYEKYTVFNELTKVRFTPEGGKEVYFSKTDKENIFDSLFKRYRKVTKRKLKDFIEKELGYGYIDIDNIK GVEQFNASYTTYQDLLKIIGDKEFLDNEENKDLLEEIIYILTVFEDRKMIEKRLSELNIPFENKIIKKLARKKYTGWGNLSRKLIDGIRNRETNRTILGHLIDDGFSNRNLMQLINDD GLDFKEIIRKAQTIENIDTNQALVSSLPGSPAIKKGILQSLNIVDEIIAIMGYAPTNIVIEMARENQTTQKGRDNSAQRLKKIEDGIKILGSDLLKQNPIQDNKDLQKEKLFLYYMQNG IDLYTGQPLNCDPDSLAFYDVDHIVPRSYIKNDSFDNKVLTTSKGNRKKLDDVPAKEVVEKMENTWRRLHAAGLISDIKLSYLMKGELTEEDKAGFIRRQLVETRQITKHVARILDERFNTETDENNKKIRQ NRKKNENGEKLRTTKIITLKSVFASRFRANFDLYKLRELNHYHHAHDAYINAVVAQALLKVYPKFERELVYGSYVKESIFSRKISATERMRMYNNILKFISKDKKVDQETGEIVWDKKE IENIVKKVIYSSPVNIVKKREEQSGALFKQSNMAVGTNNKLIPRKDWSVDKYGGFIEPAESYSLAIFYTDINGKKPKKKSTIIAISRMEKKDYEKEPERFLAQKGFERVEKTIKLPKY SLFEMEKGRRRLIASSGELQKGNQVLIPEHLIRLLSYAKKVDVLVKSKDDDYDLEEHRAEFAELLDCIKKFNDMYILASSNMSKIEEIYQKNIDAPIEEVARSFVALLNFTMMGAATDF KFFGQIIPRKRYPSTTECLKSTLHQSVTGLYETRIDLSKLGEN	WP_018372492.1 [Streptococcus massiliensis]
232	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIEKNLLGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSIILSGILTADDSSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKLLTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGETAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLIDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKPYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RRRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_019313659.1 [Streptococcus mutans]
233	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIEKNLLGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIIGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSIILSGILTVTDVGTQAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKLLTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLITGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLITDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RRRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002263549.1 [Streptococcus mutans]
234	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIEKNLLGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSIILSGILTVTDVGTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKLLTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLITGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLITDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RRRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002263887.1 [Streptococcus mutans]
235	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIEKNLLGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSIILSGILTVTDVGTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKLLTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLITGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLITDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RRRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002263887.1 [Streptococcus mutans]

236	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIEKNLLGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKVPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVGTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKLITFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLNKLGGD	WP_019803776.1 [Streptococcus mutans]
237	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIEKNLLGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVGTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKLITFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLNKLGGD	WP_024783594.1 [Streptococcus mutans]
238	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLLGALLFDSGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMNKVDDSFFHRLDESFLTDDDKNFDSHPIF GNKAEEDAYHQKFPTIYHLRKHLADSTEKADLRLVYLALAHMIKFRGHFLIEGELNAENTDVQKLFADFVGVYDRTFDDSHLSEITVDASSILTEKISKSRRLEKLINNYPKEKKNTLF GNLIALSLGLQPNFKTNFKLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTLAKNLYDAILLSGILTADDSSTKAPLSASMIKRYAEHHEDLEKLKEFIKANKPEIYHDIFKDETKN GYAGYIENGVKQDEFYKYLKNTLSKIAGSDYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGDYYPFLKENQDRIEKILTFRIPYYVGPLARKNSRFAWAEYHSDEAVTPW NFDQVIDKESSAQAFIEHMTNNDLYLLPNEKVLPKHSPLYEKYTVYNELTKIKYVTEIGEAKFFDANLKQEIFDGLFKHERKVTKKKLRTFLDKNFDEFRIVDIQGLDKETETFNASYAT YQDLLKVIKDKVFMDNPENAEILENIVLTLTLFEDREMIKQRLAKYADVFDKKVIDQLARRHYTGWGRLSAKLLNGIRDKQSGKTIMDYLIDDAQSNRNLMQLITDDNLTFKDDIVKAQ VVDNSDDLHQVVQSLAGSPAIKKGILQSLKIVDELVKVMGKEPEQIVVEMARENQTTAKGRRNSQQRYKRLKEAIKSLDRDLNHKILKEHPTDNQALQNNRLFLYYLQNGRDMYTGESL DINRLSDYDIDHVIPQAFIKDNSIDNRVLTSSKANRGKSDDVPSEDVVNRMRPFWNKLLSSGLISQRKYNNLITKKELTLDDKAGFIKRQLVETRQITKHVARMLDERFNKEFDDNNKRI RRVKIVTLKSNLVSSFRKEFELYKVRETNDYHHAHDAYLNAVVVKALLVKYPKLESEFVYGEYPKYNSYRERKTATQKMFFYSNIMNMFKSKVKLADDQIVERPMIETVNDETGEIAWDK TKHITTVKKVLSYPQVNIVKKVEEQTIGQNGGLFDDNPKSPLEVIPSKLVPLKKALNNPEKYGGYQKPTTAYPILLIVDTKQLIPISVMDKKRFEQNPVKFLKDKGYQQIEKNNFVKLPK YTLVDIGNGIKRLWASSKEVHKGNQLVVSKKSQDLLYHAHHLDNDYSNEYVKNHYQQFDILFNEITSFSKKCKLGKEHIQKIEEAYSKERDSASIEELADGFIKLLGFTQLGATSPFSF LGIKINQKQYTGKKDYILLPCMEATLIHQSITGLYETRIDLSKLGGD	WP_024786433.1 [Streptococcus mutans]
239	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLLGALLFDSGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLTQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKLITFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTLLDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKPYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFHTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRETNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPAAFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002275430.1 [Streptococcus mutans]
240	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLLGALLFDSGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKLITFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVESQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKNVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFHTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRETNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002277364.1 [Streptococcus mutans]

241	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002280230.1 [Streptococcus mutans]
242	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKNVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_012997688.1 [Streptococcus mutans]
243	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKNVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_019312892.1 [Streptococcus mutans]
244	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKNVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_024784666.1 [Streptococcus mutans]
245	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLVQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKNVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_024784666.1 [Streptococcus mutans]

246	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFAEEMMQVDESFFQRLDDSFLVEEDKRGSRYPIF GTLKEEKKYHKEFPTIYHLREKLANSTEKADLRLVYLSLAHMIKFRGHFLIEGQLKAENTNVQALFKDFVEVYDKTVEESHLSEMTVDALSILTEKVSKSRRLENLVECYPTEKKNTLF GNLIALSLGLQPNFKTNFQLSEDAKLQFSKDTYEEDLEGLLGEIGDEYADLFASAKNLYDAILLSGILAVDDNTTKAPLSASMVKRYKEHKEELAAFKRFIKEKLPKKYEEIFKDDTKN GYAGYVGADKKLKRKRSGKLATEEEFYKYVKGILNKVEGADVWLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLVRKGSRFAWAE YKADEKITPWNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSLLYETFTVYNELTKVKYVNEQGEAKFFDANMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKE NKVFNSSLGTYHDLRKILNKSFLDNKENEQIIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHTGWGRLSYKLINGIRDKQSNKTILGYLIDDGYSNRNFMQLINDDALS FKEEIAKAQVIGEMDGLNQVVSDIAGSPAIKKGILQSLKIVDELVKVMGHNPANIVIEMARENQTTAKGRRSSQKRYKRLEEAIKNLDHDILNHKILKEHPTDNQALQNDRLFLYYLQNG RDMYTEDPLDINRLSDYDIDHIIPQAFIKDNSIDNRVLTRSDKNRGKSDDVPSEEVVHKMKPFWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDER FNTETDENNKKIRQVKIVTLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDE HISNIKKVLSYPQVNIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEEN IIKLPKYSLFKLENGRKRLLASARELQKGNEIVLPNHLETLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAI GAPAAFKFFDKNIDRKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002304487.1 [Streptococcus mutans]
247	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLDESFLTDDDKNFDSHPIF GNKAEEDAYHQKFPTIYHLRKHLADSTEKADLRLVYLALAHMIKFRGHFLIEGELNAENTDVQKLFADFVGVYDRTFDDSHLSEITVDASSILTEKISKSRRLEKLINNYPKEKKNTLF GNLIALSLGLQPNFKTNFKLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTLAKNLYDAILLSGILTADDSSTKAPLSASMIKRYAEHHEDLEKLKEFIKANKPEIYHDIFKDETKN GYAGYIENGVKQDEFYKYLKNTLSKITGSDYFLDQIERENFDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGDYYPFLKENQDRIEKILTFRIPYYVGPLARKNSRFAWAEYHSDEAVTPW NFDQVIDKESSAQAFIERHMTNNDLYLPNEKVLPKHSPLYEKYTVYNELTKIKYVTEIGEAKFFDANLKQEIFDGLFKHERKVTKKKLKRTFLDKNFDEFRIVDIQGLDKETETFNASYAT YQDLLKVIKDKVFMDNPENAEILENIVLTLTLFEDREMIKQRLAKYADVFDKKVIDQLARRHYTGWGRLSAKLINGIRDKQSCKTIMDYLIDDAQSNRNLMQLITDGNLTFKDDIVKAQ YVDNSDDLHQVVQSLAGSPAIKKGILQSLKIVDELVKVMGKEPEQIVVEMARENQTTAKGRRNSQQRYKRLKEAIKSLDRDLNHKILKEHPTDNQALQNNRLFLYYLQNGRDMYTGESL DINRLSDYDIDHVIPQAFIKDNSIDNRVLTSSKANRGKSDDVPSEDVVNRMRPFWNKLLSSGLISQRKYNNLTKKELTLDDKAGFIKRQLVETRQITKHVARMLDERFNKEFDDNNKRI RRVKIVTLKSNLVSSFRKEFELYKVREINDYHHAHDAYLNAVVVKALLVKYPKLEPEFVYGEYPKYNSYRERKTATQKMFFYSNIMNMFKSKVKLADDQIVERRPMIEVNDETGEIAWDK TKHITTVKKVLSYPQVNIVKKVEEQTIGQNGGLFDDNPKSPLEVIPSKLVPLKKALNNPEKYGGYQKPTTAYPILLIVDTKQLIPISVMDKKRFEQNPVKFLKDKGYQQIEKNNFVKLPK YTLVDIGNGIKRLWASSKEVHKGNQLVVSKKSQDLLYHAHHLDNDYSNEYVKNHYQQFDILFNEITSFSKKCKLGKEHIQKIEEAYSKERDFASIEELADGFIKLLGFTQLGATSPFSF LGIKLNQKQYTGKKDYFLLPCMEATLIHQSITGLYETRIDLSKLGGD	WP_024784288.1 [Streptococcus mutans]
248	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLDESFLTDDDKNFDSHPIF GNKAEEDAYHQKFPTIYHLRKHLADSTEKADLRLVYLALAHMIKFRGHFLIEGELNAENTDVQKLFADFVGVYDRTFDDSHLSEITVDASSILTEKISKSRRLEKLINNYPKEKKNTLF GNLIALSLGLQPNFKTNFKLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTLAKNLYDAILLSGILTADDSSTKAPLSASMIKRYAEHHEDLEKLKEFIKANKPEIYHDIFKDETKN GYAGYIENGVKQDEFYKYLKNTLSKITGSDYFLDQIERENFDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGDYYPFLKENQDRIEKILTFRIPYYVGPLARKNSRFAWAEYHSDEAVTPW NFDQVIDKESSAQAFIEHMTNNDLYLPNEKVLPKHSPLYEKYTVYNELTKIKYVTEIGEAKFFDANLKQEIFDGLFKHERKVTKKKLKRTFLDKNFDEFRIVDIQGLDKETETFNASYAT YQDLLKVIKDKVFMDNPENAEILENIVLTLTLFEDREMIKQRLAKYADVFDKKVIDQLARRHYTGWGRLSAKLINGIRDKQSCKTIMDYLIDDAQSNRNLMQLITDGNLTFKDDIVKAQ YVDNSDDLHQVVQSLAGSPAIKKGILQSLKIVDELVKVMGKEPEQIVVEMARENQTTAKGRRNSQQRYKRLKEAIKSLDRDLNHKILKEHPTDNQALQNNRLFLYYLQNGRDMYTGESL DINRLSDYDIDHVIPQAFIKDNSIDNRVLTSSKANRGKSDDVPSEDVVNRMRPFWNKLLSSGLISQRKYNNLTKKELTLDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ RRVKIVTLKSNLVSSFRKEFELYKVREINDYHHAHDAYLNAVVVKALLVKYPKLEPEFVYGEYPKYNSYRERKTATQKMFFYSNIMNMFKSKVKLADDQIVERRPMIEVNDETGEIAWDK TKHITTVKKVLSYPQVNIVKKVEEQTIGQNGGLFDDNPKSPLEVIPSKLVPLKKALNNPEKYGGYQKPTTAYPILLIVDTKQLIPISVMDKKRFEQNPVKFLKDKGYQQIEKNNFVKLPK YTLVDIGNGIKRLWASSKEVHKGNQLVVSKKSQDLLYHAHHLDNDYSNEYVKNHYQQFDILFNEITSFSKKCKLGKEHIQKIEEAYSKERDFASIEELADGFIKLLGFTQLGATSPFSF LGIKLNQKQYTGKKDYFLLPCMEATLIHQSITGLYETRIDLSKLGGD	WP_002305844.1 [Streptococcus mutans]
249	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMDKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEELEVLIAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYVTEQGKTAFFDANMKQEIFDGVFKVYRKVTDKDIMDFLEKEEDFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTKQGRRNSQQRLKGLITDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEYISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002279839.1 [Streptococcus mutans]
250	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLDESFLTDDDKNFDSHPIF GNKAEEDAYHQKFPTIYHLRKHLADSTEKADLRLVYLALAHMIKFRGHFLIEGELNAENTDVQKLFADFVGVYDRTFDDSHLSEITVDASSILTEKISKSRRLEKLINNYPKEKKNTLF GNLIALSLGLQPNFKTNFKLSEDAKLQFSKDTYEEDLEVLIAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSVEAFINRMTNYDLYLPNEKVLPKHSLLYEKFTVYNELTKVKYVTEQGKTAFFDANMKQEIFDGVFKVYRKVTDKDIMDFLEKEEDFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLKNYSDLLTKEQLKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLITDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLITSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEYISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGDSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002279839.1 [Streptococcus mutans]

251	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLDESFLTDDDKNFDSYPIF GNKAEEDAYHQKFPTIYHLRKHLADSTEKADLRLVYLALAHMIKFRGHFLIEGELNAENTDVQKLFADFVGVYDRTFDDSHLSEITVDASSILTEKISKSRRLEKLINNYPKEKKNTLF RNLVALSLGLQPNFKTNFKLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTLAKNLYDAILLSGILTADDSSTKAPLSASMIKRYAEHHEDLEKLKEFIKANKPELYHDIFKDETKN GYAGYIENGVKQDEFYKYILKNTLSKIAGSDYFLDKIERENFDNGSIPHQIHLQEMHAILRRQGDYYPFLKENQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSVEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLKNYSDLLTKEQLKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEYISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGDSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPAAFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002264920.1 [Streptococcus mutans]
252	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGECHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKTDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLTQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGNGYFLDKIERENFDYLQKSIEGQILKFDSIKEFGSQILKEHPVENSQLQNDRLFLYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKPYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPAAFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_019315370.1 [Streptococcus mutans]
253	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKTDLRLVYLALAHIIKFGGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDIYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENFDYLQKSIEGQILKFDSIKEFGSQILKEHPVENSQLQNDRLFLYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSVEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLKNYSDLLTKEQLKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKNVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPAAFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002296423.1 [Streptococcus mutans]
254	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKTDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDIYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSVEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLKNYSDLLTKEQLKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002269043.1 [Streptococcus mutans]
255	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKTDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDIYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENFDYLQKSIEGQILKFDSIKEFGSQILKEHPVENSQLQNDRLFLYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSVEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLKNYSDLLTKEQLKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKNVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFHTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPAAFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002283846.1 [Streptococcus mutans]

256	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKTDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDIYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSVEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLKNYSDLLTKEQLKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSYKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPAAFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002288990.1 [Streptococcus mutans]
257	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKTDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDIYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLKNYSDLLTKEQLKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSLAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGDSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002290427.1 [Streptococcus mutans]
258	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKTDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDIYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLKNYSDLLTKEQLKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSLAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKHQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEYISNIKKVLSYPQV NIVKKVEEQTGGFKESILPKGDSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002310390.1 [Streptococcus mutans]
259	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKTDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDIYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTLLDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFKESILPKGDSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002352408.1 [Streptococcus mutans]
260	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKTDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDIYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLKNYSDLLTKEQLKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSLAIKKGILQNLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFKESILPKGDSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_019805234.1 [Streptococcus mutans]

261	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQKLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQLSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSIILSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSVEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEYISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPIAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPAAFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002307203.1 [Streptococcus mutans]
262	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQKLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQLSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSIILSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSVEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEYISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPIAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPAAFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_019314093.1 [Streptococcus mutans]
263	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSIILSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRQNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSEDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFYTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPAAFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002269448.1 [Streptococcus mutans]
264	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEDLEELLGKIGDDYADLFTLAKNLYDAILLSGILTADSSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002271977.1 [Streptococcus mutans]
265	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEDLEELLGKIGDDYADLFTLAKNLYDAILLSGILTADSSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKPYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHEENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002272766.1 [Streptococcus mutans]

266	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTKQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFYTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002273241.1 [Streptococcus mutans]
267	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTKQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFYTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002276448.1 [Streptococcus mutans]
268	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVNDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLASGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTKQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFYTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPAAFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002295753.1 [Streptococcus mutans]
269	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKTDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGCF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLTQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKLLTFRIPYYVGPLASGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002289641.1 [Streptococcus mutans]
270	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTTRRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKTDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLTQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKLLTFRIPYYVGPLASGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRELNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDENGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_024784894.1 [Streptococcus mutans]

271	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMRVFGDTDRSHIKKNLIGTLLFDDGNTAESRRLKRTARRYTRRRNRILYLQEIFTESMNEIDESFFHRLDDSFLVPEDKRGSKYPIF ATLQEEKEYHKQFPTIYHLRKQLADSKEKSDVRLIYLALAHMIKYRGHFLYEEETFDIKNNDIQKIFNEFINIYDNTFEGSSLSQQNAQVEAIFTDKISKSAKRERVLKLFPDEKSTGLF SEFLKLIVGNQADFKKHFDLEEKAPLQFSKDTYDEELENLLGQIGDDFADLFLIAKKLYDAILLSGILTVTDPSTKAPLSASMIERYENHQKDLAALKQFIKNNLPEKYVEVFSDQSKD GYAGYIDGKTTQEAFYKYIKNLLSKFEGADYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMNAILRRQGEEYPFLKENREKIEKILTFRIPYYVGPLARDNRDFSWLTRNSDEPIRPW NFEEVVDKARSAEDFIHRMTNYDLYLPEEKVLPKHSLLYETFAVYNELTKVKFIAEGLRDYQFLDSGQKKQLVNQLFKEKRKVTEKDIIHYLHNVDGYDGIELKGIEKQFNANLSTYHD LLKITKDKEFMDDPKNEEILENIVHTLTIFEDREMIKQRLAQYDTLFDEKVIKALTRRHYTGWGKLSAKLINGIRDKQSGKTILDYLIDDGKINRNFMQLINDDGLSFKEIIQKAQVVG KTDDVKQVVQELPGSPAIKKGILQSIKIVDELVKVMGYAPESIVIEMARENQTTARGKKNSQQRYKRIEDALKNLAPGLDSNILKENPTDNIQLKNDRLFLYYLQNGKDMYTGKPLDIN QLSSYDIDHIIPQAFIKDDSIDNRVLTSSKDNRGKSDNVPSLEVVQKRKAFWQQLLDSKLISERKFNNLTKAERGGLDERDKVGFIRRQLVETRQITKNVAQILDARFNTEVKEKNQKI RIVKIITLKSNLVSNFRKEFGLYKVRETNNYHHAHDAYLNAVVAKALIKKYPKLEPEFVYGDYQKYDLKKYISRFKPSKEIEKATEKYFFYSNLINFFKEEVLYADGTIRKRENIEYSK DTGEIAWDKEKDFATIKKVLSYPQVNIVKKREVQTGGFSKESILPKGNSDKLIPRKTKDLIWDTTKYGGFDSPVIAYSILLIADIEKGKAKKLKTVKTLVGITIMERAAFEENPITFLE NKGYHNVRKENILCLPKYSLFELENGRRRLLASAKELQKGNEIVLPVCLTTLLYHSKNLHKLDDEPEHLEYIQKHRNEFKDLLNLVSEFSQKVILADANLEKIKDLYADNEQADIEILAN SFINLLTFTALGAPAAFKFFGKDVDRKRYTTVSEILNATLIHQSITGLYETRIDLSKLGGE	WP_002897477.1 [Streptococcus sanguinis]
272	MKKPYSIGLDIGTNSVGWAVVTDDYKVPDKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKTDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDIYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEEYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLKNYSDLLTKEQLKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRFNTRFDDNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRETNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002287255.1 [Streptococcus mutans]
273	MKKPYSIGLDIGTNSVGWAVVTDDYKVSAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEEYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDDNNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRETNDYHHTHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	WP_002282906.1 [Streptococcus mutans]
274	MKKPYSIGLDIGTNSVGWSVVTDDYKVPAKKMKVLGNTDKSHIEKNLLGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVGTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEEYPFLADNQDRIEKLLTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDDNNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRETNDYHHAHDAYLNAVVGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRIDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPVVAFSLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDSFNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILNATLIHQSITGLYETRIDLSQLGEN	WP_037581760.1 [Streptococcus equi]
275	MKKPYTIALDIGTNSVGWVVVTDDYRVPTKKMKVLGNTERKTIKKNLIGALLFDSGDTAEGTRLKRTARPRYTRRKNRLRFLKEIFTEEMAKVDDGFFQRLEDSFYVLEDKEGNKHPIF ANLADEVAYHKKYPTIYHLRKELVDNPQKADLRLIYLAVAHIIKFRGHFLIEGTLSSKNNNIQKSFDHLVDTYNLLFEEQRLLTEGINAKELLSAALSKSKRLENLISLIPGQKKTGIF GNIIALSLGLTPNFKANFGLSKDVKLQLAKDTYADDLDSLLAQIGDQYADLFLAAKNLSDAILLSDILTESDEITRAPLSASWVKRYREHHKDLVTLKTLIKDQLPEKYQEIFLDKTKN GYAGYIEGQVSQEEFYKFLKPILARLDGSEPLLLKIDREDFLRKQRTFDNGSIPHQIHLEELHAILRRQEDFYPFLKDNRKKIESILTFRIPYYVGPLARGHSRFAWVKRKFDGAIRPW NFEEIVDEEASAQIFIEKMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMTRPQFLSAQQKAIVDLLLFKTNRKVTVKQLKEDYFKKIECWDSVEITGVEDSFNASLGTYH DLLKIIQDKDFLDNPDNQKIIEDIILTLTLFEDKKMISKRLDQYAHLFDKVVLNKLERRHYTGWGRLSGKLINGIRDKQSGKTILDFLKADGFANRNFMQLIHDSELSFIDEIAKAQVI GKTEYSKDLVGNLAGSPAIKKGISQTIKIVDELVKIMGYLPQQIVIEMARENQTTAQGIKNSRQRMRKLEETAKKIGSDNLKEYIEKIQLKDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPQSFIKDNSFNNKVLVTSQGANRGKLDNVPSEAIVKKMKGYWQSLLRAGAISKQKFDNLTKAERGGLTQVDKAGFIQLQLVETRQITKHVAQILDSRFNTEFDDHNKRIRKV HIITLKSKLVSDFRKEFGLYKIRDINHYHHAHDAYLNAVVAKAILGKYPQLAPEFVYGDYPKYNSFKERQKATQKTLFYSNILKFFKDQESLHVNSDGEEIWNANKHLPIIKNVLSIPQ VNIVKKTEVQTGGFYKESILSKGNSDKLIPRKNNWDTRKYGGFDSPTVAYSVLVIAKMEKGKAKVLKPVKEMVGITIMERIAFEENPVVFLEAKGYREIQEHLIIKLPKYSLFELENGR RRLLASASEILQKGNELFLPVDYMPTFIYLEATRIDLSQLGEN	WP_014677909.1 [Streptococcus mutans]

276	MKKPYTIALDIGTNSVGWVVVTDDYRVPTKKMKVLGNTERKTIKKNLIGALLFDSGDTAEGTRLKRTARRRYTRRKNRLRYLKEIFTEEMAKVDDGFFQRLEDSFYVLEDKEGNKHPIF ANLADEVAYHKKYPTIYHLRKELVDNPQKADLRLIYLAVAHIIKFRGHFLIEGTLSSKNNNLQKSFDHLVDTYNLLIFEEQRLLTEGINAKELLSAALSKKSKRLENLISLIPGQKKTGIF GNIIALSLGLTPNFKANFGLSKDVKLQLAKDTYADDLDSLLAQIGDQYADLFLAAKNLSDAILLSDILTESDEITRAPLSASMVKRYREHHKDLVTLKTLIKDQLPEKYQEIFLDKTKN GYAGYIEGQVSQEEFYKYLKPIIAKLDGSEDLLLKIDREDFLRKQRTFDNGSIPHQIHLEELHAILRRQEVFYPFLKDNRKKIESLLTFRIPYYVGPLARGHSRFAWVKRKFDGAIRPW NFEEIVDEEASAQLIFIEKMTKNDLYLPNEKVLPKHSLLYETFTVYNELTKVKYATEGMTRPQFLSADQKQAIVDLLFKTNRKVTVKQLKENYFKKIECFDSVEITGVEDSFNASLGTYH DLIKIIQDKDFLDNPDNQKIIEDILTLTLFEDKRMISKRLDQYAHLFDKVVLNKLERHHYTGWGRLSGKLINGIRDKQSGKTILDFLKADGFANRNFMQLIHDSELSFIDEIAKAQVI GKTEYSKDLVGNLAGSPAIKKGISQTIKIVDELVKIMGYLPQQIVIEMARENQTTAQGIKNSRQRMRKLEEIAKKIGSNILKEHPVDNSQLQNDRRYLYYLQNGKDMYTGDDLDIDYLS SYDIDHIIPQSFIKNNSIDNKVLTSQGANRGKLDNVPSEAIVRKMKGYWQSLLRAGAISKQKFDNLTKAERGGLTQVDKAGFIQRQLVETRQITKHVAQILDSRFNTEFDDHNKRIRKV HIITLKSKLVSDFRKEFGLYKIRDINHYHHAHDAYLNAVVAKAILGKYPQLAPEFVYGDYPKYNSFKERQKATQKMILFYSNILKFFKDQESLHVNSDGEEIWNANKHLPTIKNVLSIPQ VNIVKKTEVQTGGFYKESILSKGNSDKLIPRKNNWDTRKYGGFDSPTVAYSVLVIAKMEKGKAKVLKPVKEMVGITIMERTAFEENPVVFLEARGYREIQEHLIIKLPKYSLFELENGR RRLLASASELQKGNELFLPVDYMTFLYLAAHYHELTGSSEDVLRKKYFVDRHLHYFDDIIQMINDFAERHILASSNLEKINHTYHNNSDLPVNERAENIINVFTFVALGAPAAFKFFDA TIDRKRYTSTKEVLNATLIHQSVTGLYETRIDLSQLGEN	WP_012515931.1 [Streptococcus equi]
277	MKKPYTIALDIGTNSVGWVVVTDDYRVPTKKMKVLGNTERKTIKKNLIGALLFDSGDTAEGTRLKRTARRRYTRRKNRLRYLKEIFTEEMAKVDDGFFQRLEDSFYVLEDKEGNKHPIF ANLADEVAYHKKYPTIYHLRKELVDNPQKADLRLIYLAVAHIIKFRGHFLIEGTLSSKNNNLQKSFDHLVDTYNLLIFEEQRLLTEGINAKELLSAALSKKSKRLENLISLIPGQKKTGIF GNIIALSLGLTPNFKANFGLSKDVKLQLAKDTYADDLDSLLAQIGDQYADLFLAAKNLSDAILLSDILTESDEITRAPLSASMVKRYREHHKDLVTLKTLIKDQLPEKYQEIFLDKTKN GYAGYIEGQVSQEEFYKYLKPIIAKLDGSEDLLLKIDREDFLRKQRTFDNGSIPHQIHLEELHAILRRQEVFYPFLKDNRKKIESLLTFRIPYYVGPLARGHSRFAWVKRKFDGAIRPW NFEEIVDEEASAQLIFIEKMTKNDLYLPNEKVLPKHSLLYETFTVYNELTKVKYATEGMTRPQFLSADQKQAIVDLLFKTNRKVTVKQLKENYFKKIECFDSVEITGVEDSFNASLGTYH DLIKIIQDKDFLDNPDNQKIIEDILTLTLFEDKRMISKRLDQYAHLFDKVVLNKLERHHYTGWGRLSGKLINGIRDKQSGKTILDFLKADGFANRNFMQLIHDSELSFIDEIAKAQVI GKTEYSKDLVGNLAGSPAIKKGISQTIKIVDELVKIMGYLPQQIVIEMARENQTTAQGIKNSRQRMRKLEEIAKKIGSNILKEHPVDNSQLQNDRRYLYYLQNGKDMYTGDDLDIDYLS SYDIDHIIPQSFIKNNSIDNKVLTSQGANRGKLDNVPSEAIVRKMKGYWQSLLRAGAISKQKFDNLTKAERGGLTQVDKAGFIQRQLVETRQITKHVAQILDSRFNTEFDDHNKRIRKV HIITLKSKLVSDFRKEFGLYKIRDINHYHHAHDAYLNAVVAKAILGKYPQLAPEFVYGDYPKYNSFKERQKATQKMILFYSNILKFFKDQESLHVNSDGEEIWNANKHLPTIKNVLSIPQ VNIVKKTEVQTGGFYKESILSKGNSDKLIPRKNNWDTRKYGGFDSPTVAYSVLVIAKMEKGKAKVLKPVKEMVGITIMERIAFEENPVVFLEAKGYREIQEHLIIKLPKYSLFELENGR RRLLASASELQKGNELFLPVDYMTFLYLAAHYHELTGSSEDVLRKKYFVERHLHYFDDIIQMINDFAERHILASSNLEKINHTYHNNSDLPINERAENIINVFTFVALGAPAAFKFFDA TIDRKRYTSTKEVLNATLIHQSVTGLYETRIDLSQLGEN	WP_021320964.1 [Streptococcus equi]
278	MKKPYTIGLDIGTNSVGWAALITDQYDLVKRKMKVAGNSEKKQIKKNLWGVRLVDEGKTAAHRRVNRTTRRRLIERRRNRISYLQEIFTAEMFEVDANFFYRLEDSFYIESEKRQSRHPFF ATIEEEVAYHENYRTIYHLRKELVNSSDKADLRLVYLALAHIIKYRGNFLIEGKLDTKNTSVDEVFKQFIKTYNQVFASDIEEGSLTRIEENNEVAKIFSEKLTKRRKLDKILNLIYPNE KSTDLFAQFISLIIGSKGNFKKFFNLTEKTDIECAKDSYEEDLEVLLARVGDEYAEIFVAAKNAYNAVVLSSIITVSNTETKAKLSASMIERFDTHEDLGELKAFIKLHLPKHYEEIF NDVEKDGYAGYIEGKTKQEAFYKYMKKMLEHVEGADYFINQIEEENFLRKQRTFDNGAIPHQLHLEELEAILHQQAKYYPFLKEVDYEKIKSLVTFRIPYFVGPLANGQSEFSWLTRKAD GEIRPWNIEEKVDFGKSAIDFIEKMTNKDTYLPKENVLPKHSMCYQKMVYNELTKIRYTDDQGKTSYFSGQEKQQIFNDLFKQKRKVKKKDLELFLRNMSHVESPTFEGLEDSFNSSF TTYHDLQKVGVPQEILDDPLNTEMLEEIIKILTVFEDKRMINERLQEFSNVLDEAVLKKLERRHYTGWGRLSAKLLIGIRDKQSHLITILDYLMNDDGLNRNLMQLINDSNLSFKSIIEK EQVSTADKDIQSIVADLAGSPAIKKGILQSLKIVDELVGIMGYPPQTIVVEMARENQTTGKGKNNSRPRYKSLEKAIKELGSQILKEHPTDNQELRNNRLYLYYLQNGKDMYTGQELDI HNLSNYDIDHVVPQSFITDNSIDNRVLASSAANRGKGDNVPSLEVVRKRKVYWEKLYQAKLMSKRKFDYLTKAERGGLTEADKARFIHRQLVETRQITKNVANILHQRFNYEKDDHGNI IEQVRIVTLKAALVSQFRKQFQLYKVREVNDYHHAHDAYLNCVVANTLLKVYPQLEPEFVYGDYHQFDWFKANKATAKKQFYTNIMLFFAKKDRIIDENGEILWDKKYLDTIKKVLNYR QMNIVKKTEIQKGEFFSNATANPKGNSSKLIPRKADWDPIKYGGFDSPNMAYAVVIEYAKGKNKLVFEKKIIRVTIMERKAFEKDEKAFLEEQGYRQPKVLAKLPKYTLYECEEGRRRML ASANEAQKGNQQVLPNHLVTLLHHAANCEVSDGKSLDYIESNREMFAELLAHVSEFAKRVTLAEANLNKINQLFEQNKEGDIKAIAQSFVDLMAFNAMGAPASFKFFETTIERKRYNNL KELLNSTIIYQSITGLYESRKRLDD	
279	MKKPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKIAGDSEKKQIKKNFWGVRLFDEGQTAADRRMARTARRRIERRRNRISYLQGIFAEEMSKTDANFFCRLSDSFYVDNEKRNSRHPFF ATIEEEVEYHKNYPTIYHLREELVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTQNTSVDGIYKQFIQTYNQVFASGIEDGSLKKLEDNKDVAKILVEKVTRKEKLERILKLYPGE KSAGMFAQFISLIVGSKGNFQKPFDLIEKSDIECAKDSYEEDLESLLALIGDEYAELFVAAKNAYSAVVLSSIITVAETETNAKLSASMIERFDTHEEDLGELKAFIKLHLPKHYEEIF SNTEKHGYAGYIDGKTKQADFYKYMKMTLENIEGADYFIAKIEKENFLRKQRTFDNGAIPHQLHLEELEAILHQQAKYYPFLKENYDKIKSLVTFRIPYFVGPLANGQSEFAWLTRKAD GEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKVIVYNELTKVRYINDQGKTSYFSGQEKEQIFNDLFKQKRKVKKKDLELFLRNMSHVESPTIEGLEDSFNSSY STYHDLLKVGIKQEIIDNPVNTEMLENIVKILTVFEDKRMIKEQLQQFSDVLDGVVLKKLERRHYTGWGRLSAKLIMGIRDKQSHLITILDYLMNDDGLNRNLMQLINDSNLSFKSIIEK EQVTTADKDIQSIVADLAGSPAIKKGILQSLKIVDELVSVMGYPPQTIVVEMARENQTTGKGKNNSRPRYKSLEKAIKEFGSQILKEHPTDNQELRNNRLYLYYLQNGKDMYTGQDLDI HNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGDDVPPLEIVRKRKVFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKARFIHRQLVETRQITKNVANILHQRFNYEKDDHGNI MKQVRIVTLKSALVSQFRKQFQLYKVRDVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGDYHQFDWFKANKATAKKQFYTNIMLFFAQKDRIIDENGEILWDKKYLDTVKKVMSYR QMNIVKKTEIQKGEFFSKATIKPKGNSSKLIPRKTNWDPMKYGGLDSPNMAYAVVIEYAKGKNKLVFEKKIIRVTIMERKAFEKDEKAFLEEQGYRQPKVLAKLPKYTLYECEEGRRRML ASANEAQKGNQQVLPNHLVTLLHHAANCEVSDGKSLDYIESNREMFAELLAHVSEFAKRVTLAEANLNKINQLFEQNKEGDIKAIAQSFVDLMAFNAMGAPASFKFFETTIERKRYNNL KELLNSTIIYQSITGLYESRKRLDD	WP_010991369.1 [Listeria innocua]
280	MKKPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKIAGDSEKKQIKKNFWGVRLFDEGQTAADRRMARTARRRIERRRNRISYLQGIFAEEMSKTDANFFCRLSDSFYVDNEKRNSRHPFF ATIEEEVEYHKNYPTIYHLREELVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTQNTSVDGIYKQFIQTYNQVFASGIEDGSLKKLEDNKDVAKILVEKVTRKEKLERILKLYPGE KSAGMFAQFISLIVGSKGNFQKPFDLIEKSDIECAKDSYEEDLESLLALIGDEYAELFVAAKNAYSAVVLSSIITVAETETNAKLSASMIERFDTHEEDLGELKAFIKLHLPKHYEEIF SNTEKHGYAGYIDGKTKQADFYKYMKMTLENIEGADYFIAKIEKENFLRKQRTFDNGAIPHQLHLEELEAILHQQAKYYPFLKENYDKIKSLVTFRIPYFVGPLANGQSEFAWLTRKAD GEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKVIVYNELTKVRYINDQGKTSYFSGQEKEQIFNDLFKQKRKVKKKDLELFLRNMSHVESPTIEGLEDSFNSSY STYHDLLKVGIKQEIIDNPVNTEMLENIVKILTVFEDKRMIKEQLQQFSDVLDGVVLKKLERRHYTGWGRLSAKLIMGIRDKQSHLITILDYLMNDDGLNRNLMQLINDSNLSFKSIIEK EQVTTADKDIQSIVADLAGSPAIKKGILQSLKIVDELVSVMGYPPQTIVVEMARENQTTGKGKNNSRPRYKSLEKAIKEFGSQILKEHPTDNQELRNNRLYLYYLQNGKDMYTGQDLDI HNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGDDVPPLEIVRKRKVFWEKLYQGNLMSKRKFDYLITKAERGGLTEADKARFIHRQLVETRQITKNVANILHQRFNYEKDDHGNI MKQVRIVTLKSALVSQFRKQFQLYKVRDVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGDYHQFDWFKANKATAKKQFYTNIMLFFAQKDRIIDENGEILWDKKYLDTVKKVMSYR QMNIVKKTEIQKGEFFSKATIKPKGNSSKLISRKTNWDPMKYGGLDSPNMAYAVVIEYAKGKNKLVFEKKIIRVTIMERKAFEKDEKAFLEEQGYRQPKVLAKLPKYTLYECEEGRRRML ASANEAQKGNQQVLPNHLVTLLHHAANCEVSDGKSLDYIESNREMFAELLAHVSEFAKRVTLAEANLNKINQLFEQNKEGDIKAIAQSFVDLMAFNAMGAPASFKFFETTIERKRYNNL KELLNSTIIYQSITGLYESRKRLDD	WP_033838504.1 [Listeria innocua]

281	MKKPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKISGDSEKKQIKKNFWGVRLFEKGETAAKRRMSRTARRRIERRRNRISYLQEIFAIQMNEVDDNFFNRLKESFYAESDKKYNRHPFF GTVEEEVAYYKDFPTIYHLRKELIDSQKKADLRLVYLALAHIIKYRGHFLIEGALDTKNTSIDEMFKQFLQIYNQVFANDIEEASLKKTEKNQEVAQILAEKFTRKDKLDKILSLYPGE KTIGVFAQFVNIIVGSTGKFKKHFNLHEKKDINCAEDTYDTDLESLLAIIGDEFAEVFVAAKNAYNAVVLSNIITVTDSTTRAKLSASLIERFENHKEDLKKMKRFVRTYILPEKYDEIF DDTEKHGYAGYISGKTKQADFYKYMKATLEKIEGADYFIAKIEEENFLRKQRTFDNGVIPHQLHLEELEAILHQQAKYYPFLREDYEKIKSLVTFRIPYFVGPLAKGQSEFAWLTRKAD GEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYMVYNELTKVRYIDDQGKTNYFSGQEKQQIFNDLFKQKRKVKKKDLELFLRNINQIESPTIEGLEDSFNASY ATYHDLLKVGMKQEIIDNPLNTEMLEDIVKILTVFEDKRMIKEQLQQFSDVLDGTVLKKLERRHYTGWGRLSAKLLVGIRDKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSIIEK EQVSTTDKDLQSIVAELAGSPAIKKGILQSLKIVDELVSVMGYPPQTIVVEMARENQTTNKGKNNSKPRYKSLEKAIKEFGSQILKEHPTDNQELKNNRLYLYYLQNGKDMYTGQELDI HNLSNYDIDHIVPQSFITDNSVDNLVLTSSAGNREKGDNVPPLEIVQKRKIFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKARFIHRQLVETRQITKNVANILHQRFNYKTDGNKDT METVRIVTLKSALVSQFRKQFQFYKVREVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGEYHQFGWFKANKATAKKQFYTNIMLFFAQKDRIIDENGEILWDKRYLETVKKVLGYR QMNIVKKTEIQKGEFSNVTPNPKGKSNKLIPRKDWDPIKYGGFDSPNVAYAIIIEYEKQKRKVRIEKKLIQINIMEREAFEKDEKTFLEEKGYHQPKVLIKVPKYTLYECENGNRRRML GSANEAHKGNQMLLPNHIMALLYHAEKYEAIDGESLAYIEVHRALFDELLAYISEFARKYTLSNDRLDEINMLYERNKDGDVKSIAESFVSLKKFNAFGVHQDFSFFGTKIERKRDRKL NELLNSTIIYQSITGLYESRKRLDN	WP_003733029.1 [Listeria monocytogenes]
282	MKKPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKISGDSEKKQIKKNFWGVRLFEKGETAAKRRMSRTARRRIERRRNRISYLQEIFAIQMNEVDDNFFNRLKESFYAESDKKYNRHPFF GTVEEEVAYYKDFPTIYHLRKELIDSQKKADLRLVYLALAHIIKYRGHFLIEGALDTKNTSIDEMFKQFLQIYNQVFANDIEEASLKKTEKNQEVAQILAEKFTRKDKLDKILSLYPGE KTIGVFAQFVNIIVGSTGKFKKHFNLHEKKDINCAEDTYDTDLESLLAIIGDEFAEVFVAAKNAYNAVVLSNIITVTDSTTRAKLSASLIERFENHKEDLKKMKRFVRTYILPEKYDEIF DDTEKHGYAGYISGKTKQADFYKYMKATLEKIEEGADYFIAKIEEENFLRKQRTFDNGVIPHQLHLEELEAILHQQAKYYPFLREDYEKIKSLVTFRIPYFVGPLAKGQSEFAWLTRKAD GEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYMVYNELTKVRYIDDQGKTNYFSGQEKQQIFNDLFKQKRKVKKKDLELFLRNINQIESPTIEGLEDSFNASY ATYHDLLKVGMKQEIIDNPLNTEMLEDIVKILTVFEDKRMIKEQLQQFSDVLDGTVLKKLERRHYTGWGRLSAKLLVGIRDKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSIIEK EQVSTATDKDLQSIVADLAGSPAIKKGILQSLKIVDELVSVMGYPPQTIVVEMARENQTTNKGKNNSKPRYKSLEKAIKEFGSQILKEHPTDNQELKNNRLYLYYLQNGKDMYTGQELDI HNLSNYDIDHIVPQSFITDNSVDNLVLTSSAGNREKGDNVPPLEIVQKRKIFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKARFIHRQLVETRQITKNVANILHQRFNYKTDGNKDT METVRIVTLKSALVSQFRKQFQFYKVREVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGEYHQFDWFKANKATAKKQFYTNIMLFFAQKDRIIDENGEILWDKRYLETVKKVLGYR QMNIVKKTEIQKGEFSNVTPNPKGKSNKLIPRKDWDPIKYGGFDSPNVAYAIIIEYEKQKRKVRIEKKLIQINIMEREAFEKDEKTFLEEKGYHQPKVLIKVPKYTLYECENGNRRRML GSANEAHKGNQMLLPNHIMALLYHAEKYEAIDGESLAYIEVHRALFDELLAYISEFARKYTLSNDRLDEINMLYERNKDGDVKSIAESFVSLKKFNAFGVHQDFSFFGTKIERKRDRKL NELLNSTIIYQSITGLYESRKRLDN	WP_031669209.1 [Listeria monocytogenes]
283	MKKRYSYSIGLDIGTNSVGWAVINEDYKVPAKKMTVFGNTDRKTIKKNLLGTVLFDSGETAQARRLKRTNRRRYTRRRYRLCQLQNIFATEMVKVDDTFFQRLSESFFYYQDKAFDKHP IFGNSKEERAYHKTYPTIYHLRKDLADRDQKADLRLIYLALSHIIKFRGHFLIEGKLNSENTDVQKLFIALVTVYNLLFEEEPIAGETCDAKALLITAKTSKSKRLESLISEFPGQKKNG LFGNLIALALGLRPNFKSNFGLSEDAKLQITKDTYEEELDNLLAEIGDHYADLFLAAKNLSDAILLSDILTLSDENTRAPLSASMIKRYEEHQEDLALLKKLVKEQMPEKYWEIFSNAK KNGYAGYIEGKVSQEDFYRYIKPILSRLKGGDEFLAKIDRDDFLRKQRTFDNGSIPHQIHLKELHAILRRQEKYYPFLAEQKEKIEQLLCFRIPYYVGPLAKGGNSSFAWLKRRSDEPI TPWNFKDVVDEESASQAFIEGMTNNDTYLPEEKVLPKHSPLYEMFTVYNELTKVKYIAENMTKPLYLSAEQKEAIIDHLFKQTRKVTVKDLKEKYFSQIEGLENVDVTGVEGAFNASLG TYNDLLKIIKDKAFLDDEANAEILEEIVLILTLFQDEKLIEKRLAKYANLFEKSVLKKLKRKHYRGWGRLSRQLIDGMKDKASGKTILDFLKADDFANRNFIQLINDSSLDFEKLIDDA QKKAIKRESLTEAVANLAGSPAIKKGILQSLKVVDELVKVMGHNPDNIVIEMSRENQTTAQGLKNARQRLKKIKEVIKDTSRELKSDRLFLKILKEHPVENTQLQNDKLYLYYLQNGRDMYTGQDL DINNLSQYDIDHIIPQSFIKDNSIDNLVLTTQKANRGKSDNVPSIEVVRDMKDRVWRRQLANGAISRQKFDHLTKAERGGLADSDKARFIRRQLVETRQITKHVAQILDSRFNSKSNQN KKIARNVKIITLKSKIVSDFRKDFGLYKLREVNNYHHAHDAYLNAVVGTALLKKYPKLEAEFVYGDYKHFDLVKLISKSDPSLGKATAKVFFYSNIMNFFKEELSIADGTILMKRPVIET NTETGEVVWDKVKDFKTIRKVLSYPQVNIVKKTEIQSGAFSKESVLSKGNSDKLIERKKGWDPKKYGGFDSPNTAYSIFVVAKVAKKAQKLKTVKEIVGITIMERSAFEKDPIAFLEK KGYQDIQEKLLIKLPKYSLFELENGRRRLLASANEFQKGNELALSGKYMKFLYLASRYDKLSSKIESEQQKKLFVEQHLHYFDEILDIVVKHATCYIKAENNLKKIISLYKKKEAYSIN EQALNMLNLFIFTSLGAPSTFVFFDETIDRKRYTTSSDVLNGILIQQSITGLYETRIDLSQLGGD	WP_007896501.1 [Streptococcus pseudoporcinus]
284	MKKSYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLDESFLTDDDKNFDSHPIF GNKAEEDAYHQKFPTIYHLRKHLADSTEKADLRLVYLALAHMIKFRGHFLIEGELNAENTDVQKLFADFVGVYDRTFDDSHLSELITVDASSTILTEKISKSRRLEKLINNYPKEKKNTLF GNLIALSLGLQPNFKTNFKLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTLAKNLYDAILLSGILTADDSSTKAPLSASMIKRYAEHHEDLEKLKEFIKANKPELYHDIFKDETKN GYAGYIENGVKQDEFYKYLKNTLSKIAGSDYFLDKIERKDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGDYYPFLKENQDRIEKILTFRIPYYVGPLARKNSRFAWAEYHSDEAVMPW NFDQVIDKESSSAQAFIEHMTNNDLYLPNEKVLPKHSPLYEKYTVYNELTKIKYVTEIGEAKFFDANLKQEIFDGLFKHERKVTVKKKLRTFLDKNFDEFRIVDIQGLDKETETFNASYAT YQDLLKVIKDKVFMDNPENAEILENIVLTLTLFEDREMIKQRLAKYADVFDKKVIDQLARRHYTGWGRLSAKLLNGIRDKQSCKTIMDYLIDDAQSNRNLMQLITDDNLTFKDDIVKAQ YVDNSDDLHQVVQSLAGSPAIKKGILQSLKIVDELVKVMGKEPEQIVVEMARENQTTAKGRRNSQQRVKRLKEAIKSLDRDLNHKILKEHPTDNQALQNERLFLYYLQNGRDMYTGESL DINRLSDYDIDHVIPQAFIKDNSIDNRVLTSSKANRGKSDDVPSEDVVNRMRPFWNKLLSSGLISQRKYNNLITKKELTPDDKAGFIKRQLVETRQITKHVARMLDERFNKEFDDNNKRI RRVKIVTLKSNLVSSFRKEFELYKVREINDYHHAHDAYLNAVVVKALLVKYPKLEPEFVYGEYPKYNSYRERKTATQKMFFYSNIMNMFKSKVKLADDQIVERPMIEVNDETGEIAWDK TKHITTVKKVLSYPQVNIVKKVEEQTIGQNGGLFDDNPKSPLEVIPSKLVPLKKALNPEKYGGYQKPTTAYPILLIVDTKQLIPISVMDKKRFEQNPVKFLKDKGYQQIEKNNFVKLPK YTLVDIGNGIKRLWASSKEVHKGNQLVVSKKSQDLLYHAHHLDNDYSNEYVVKNHYQQFDILFNEITSFSKKCKLGKEHIQKIEEAYSKERDSASIEELADGFIKLLGFTQLGATSPFSF LGIKLNQKQYTGKKDYILLPCMEATLIHQSITGLYETRIDLNKLGGD	WP_002277050.1 [Streptococcus mutans]
285	MKNDYTIGLDIGTNSVGYSVVTDDYKVISKKMNVFGNTEKKSIKKNFWGVRLFESGQTAEARMKRTSRRRIARRKNRICYLQEIFQPEMNHLDNNFFYRLNESFLVADDAKYDKHPIF GTLDEEIHFHEQFPTIYHLRKYLADGDEKADLRLVYLAIAHIIKFRGNFLIEGELNTENNSVIELSKVFVQLYNQTLSELEGFQFIDESIDFSEVLTQQLSKSERADNVLKLFPDEKGI GIFAQFIKLIVGNQGNFKKVFQLEEDQKLQLSTDDYEENIENLLAIIGDEYGDIFVAAQNLYQAILLAGILTSTEKTRAKLSASMIQRYEEHAKDLKLLKRFVKEHIPDKYAEIFNDAT KNGYAGYIDGKTKEEEFYKYLKTTLVQKSGYQYFIEKIEQENFLRKQRIYDNGVIPHQVHAEELRAILRKQEKYYSFLKENHEKIEQILTFRIPYYVGPLAKHNEQSRFAWNIRKSDEP IRPWNNDVVDENASAVATIEYFKDIYLYNENVLPHKSLLTKYEVHELITKKVLKQSPKKKRKVFENPLFMEYRHENGLIERNLVDTRYSIISSPSVKGIEEQFNANFGTY LDLKKFDELHPYLDDEKYQDTLEEVIKVLTVFEDRSMIQNQLEQLPLNLSTKTIKALSRRKYTGWGRLSARLIDGIHDKNSGKTILDYLIEDESDSYIVNRNFMQLINDDHLSFKKIIE DSQPYKEQQSAEELVSELSGSPAIKKGILQSLKIVDELVAIMGYKPKNIVVEMARENQTTGRGKQNSKPRLKGIENGLKEFSDSVLKGSSIDNKQLQNDRLYLYYLQNGKDMYTGHELD IDHLSTYDIDHIIPQSFLKDDSIDNRVLTRSDKNRGKSDNVPSEEVVKRMKDFWRKLLNAKLISERKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRFNSEMDDNFVKC TLKSAIVSEFRKTFNLYKVREINDLHHAHDAYLNAVVALSLLRVYPQLKPEFVYGEYGKNSIHDQNKATIKKQFYSNITRYFASKDYIINDDGEILWNKQETIAQVIKTILGMHQVNVVK KVEIQKGGFSKESIQPKGESQKLIRRKQQWNTKKYGGFDSPVVAYAILSLSFDKGKRKARSFKIVGITIQDRRSFEGNPILYLSKKDYHNPKVEAILPKYSLFEFENGRRRMVASASETQ KGNQLIIPGHLMELLYHSKKIINGKNSDSVSYIQNNKEKFREIFEYIVDFSSKYISADANLNKIEKIFENNFHKASEQEIAKSFINLLTFTAMGAPADFEFFGEKIPRKRVYSISEIID AVFIHQSITGLYETRVRLTEV	WP_007209003.1 [Enterococcus italicus]

286	MKNPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKVAGNSDKKQIKKNFWGVRLFDDGQTAVDRRMNRTARRRIERRRNRISYLQEIFAVEMANIDANFFCRLNDSFYVDSEKRNSRHPFF ATIEEEVAYHKNYRTIYHLREELVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTKNTSVDEVYKQFIQTYNQVFMSNIEEGALAKVEENTEVASILAGKFTRREKFERILRLYPGE KSTGMFAQFISLIVGNKGNFQKVFNLVEKTDIECAKDSYEEDLEALLAIIGDEYAELFVAAKNTYNAVVLSSIITVTATETNAKLSASMIERFDAHEKELGELKAFIKLHLPKQYQEIF NNAEIDGYAGYIDGKTKQVDFYKYLKTTLENVEGADYFITKIEEENFLRKQRTFDNGVIPHQLHLEELEAILHQQAKYYPFLREGYDKIKSLVTFRIPYFVGPLANGQSEFAWLTRKDD GEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYMVYNELTKIRYIDDQGKTNYFSGREKQQIFNDLFKQKRKVKKKDLELFLRNININHIESPTIEGLEDSFNASY ATYHDLLKVGLKQEIIDNPLNTEILEDIVKILTVFEDKRMIKEQLEQFSDVLDGVVLKKLERRHYTGWGRLSAKLLVGIRDKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSIIEK EQVSTTDKDLQSIVADLAGSPAIKKGILQSLKIVDELVSIMGYPPQTIVVEMARENQTTGKGKNNSKPRYKSLEKAIKDFGSQILKEHPTDNQELKNNRLYLYYLQNGKDIYTGQELDI HNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGDVPPLEIVRKRKVFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKARFIHRQLVETRQITKNVANILHQRFNNETDNHGNT MEQVRIVTLKSALVSQFRKQFQLYKVREVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGEYHQFDWFKANKATAKKQFYTNIMLFFAQKERIIDENGEILWDKKYLETIKKVLDYR QINIVKKTEIQKGEFFSKATIKPKGNSSKLIPRKENWDPMKYGGLDSPNMAYAVIIEHAKGKKKIVIEKKLIQINIMERKMFEKDEEAFLEEKGYHQPKVLTKLPKYTLYECEKGRRRML SSANEAQKGNQLVLSNHLVSLLYHAKNCEASDGKSLKYIEAHRETFSELLAQVSEFATKYTLADANLSKINNLFEQNKEGDIKAIAQSFVDLMAFNAMGAPASFKFFEATIDRKRYTNL KELLSSTIIYQSITGLYESRKRLDD	WP_003727705.1 [Listeria monocytogenes]
287	MKNPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKVAGNSDKKQIKKNFWGVRLFDDGQTAVDRRMNRTARRRIERRRNRISYLQEIFAVEMANIDANFFCRLNDSFYVDSEKRNSRHPFF ATIEEEVAYHKNYRTIYHLREELVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTKNTSVDEVYKQFIQTYNQVFMSNIEEGALAKVEENTEVASILAGKFTRREKFERILRLYPGE KSTGMFAQFISLIVGNKGNFQKVFNLVEKTDIECAKDSYEEDLEALLAIIGDEYAELFVAAKNTYNAVVLSSIITVTATETNAKLSASMIERFDAHEKELGELKAFIKLHLPKQYQEIF NNAEIDGYAGYIDGKTKQVDFYKYLKTTLENVEGADYFITKIEEENFLRKQRTFDNGVIPHQLHLEELEAILHQQAKYYPFLREGYDKIKSLVTFRIPYFVGPLANGQSEFAWLTRKDD GEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYMVYNELTKIRYIDDQGKTNYFSGREKQQIFNDLFKQKRKVKKKDLELFLRNININHIESPTIEGLEDSFNASY ATYHDLLKVGLKQEIIDNPLNTEILEDIVKILTVFEDKRMIKEQLEQFSDVLDGVVLKKLERRHYTGWGRLSAKLLVGIRDKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSIIEK EQVSTTDKDLQSIVADLAGSPAIKKGILQSLKIVDELVSIMGYPPQTIVVEMARENQTTGKGKNNSKPRYKSLEKAIKDFGSQILKEHPTDNQELKNNRLYLYYLQNGKDIYTGQELDI HNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGDVPPLEIVRKRKVFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKARFIHRQLVETRQITKNVANILHQRFNNETDNHGNT MEQVRIVTLKSALVSQFRKQFQLYKVREVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGEYHQFDWFKANKATAKKQFYTNIMLFFAQKERIIDENGEILWDKKYLETIKKVLDYR QINIVKKTEIQKGEFFSKATIKPKGNSSKLIPRKENWDPMKYGGLDSPNMAYAVIIEHAKGKKKIVIEKKLIQINIMERKMFEKDEEAFLEEKGYHQPKVLTKLPKYTLYECEKGRRRML SSANEAQKGNQLVLSNHLVSLLYHAKNCEASDGKSLKYIEAHRETFSELLAQVSEFATKYTLADANLSKINNLFEQNKEGDIKAIAQSFVDLMAFNAMGAPASFKFFEATIDRKRYTNL KELLSSTIIYQSITGLYESRKRLDD	WP_003730785.1 [Listeria monocytogenes]
288	MKNPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKVAGNSDKKQIKKNFWGVRLFDDGQTAVDRRMNRTARRRIERRRNRISYLQEIFAVEMANIDANFFCRLNDSFYVDSEKRNSRHPFF ATIEEEVAYHKNYRTIYHLREELVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTKNTSVDGVYEQFIQTYNQVFMSNIEEGTLAKVEENIEVANILAGKFTRREKFERILQLYPGE KSTGMFAQFISLIVGSKGNFQKVFDLIEKTDIECAKDSYEEDLETLLAIIGDEYAELFVAAKNTYNAVVLSSIITVNDTETNAKLSASMIERFDAHEKDLVELKAFIKLNLPKQYEEIF SNAAIDGYAGYIDGKTKQVDFYKYLKTILENIEGSDYFIAKIEEENFLRKQRTFDNGAIPHQLHLEELEAIIHQQAKYYPFLKEDYDKIKSLVTFRIPYFVGPLANGQSEFAWLTRKAD GEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYMVYNELTKVRYIDDQGKTNYFSGQEKQQIFNDLFKQKRKVKKKDLELFLRNINQIESPTIEGLEDSFNASY ATYHDLLKVGMKQEIIDNPLNTEMLEDIVKILTVFEDKRMIKEQLQQFSDVLDGVVLKKLERRHYTGWGRLSAKLLVGIRDKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSIIEK EQVSTTDKDLQSIVAELAGSPAIKKGILQSLKIVDELVSIMGYPPQTIVVEMARENQTTGKGKNNSKPRYKSLEKAIKEFGSQILKEHPTDNQELKNNRLYLYYLQNGKDMYTGQELDI HNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGDVPPLEIVRKRKVFWEKLFQGNLMSKRKFDYLITKAERGGLTEADKATFIHRQLVETRQITKNVANILHQRFNNETDNHGNT MEQVRIVTLKSALVSQFRKQFQLYKVREVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGEYHQFDWFKANKATAKKQFYTNIMLFFAQKERIIDENGEILWDKKYLETIKKVLDYR QMNIVKKTEIQKGEFFSKATIKPKGNSSKLIPRKENWDPMKYGGLDSPNMAYAVIIEHAKGKKKIVIEKKLIQINIMERKMFEKDEEAFLEEKGYRHPKVLTKLPKYTLYECEKGRRRML ASANEAQKGNQLVLSNHLVSLLYHAKNCEASDGKSLKYIEAHRETFSELLAQVSEFATRYTLADANLSKINNLFEQNKEGDIKAIAQSFVDLMAFNAMGAPASFKFFEATIDRKRYTNL KELLSSTIIYQSITGLYESRKRLDD	WP_031665337.1 [Listeria monocytogenes]
289	MKNPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKVAGNSDKKQIKKNFWGVRLFDEGETAADRRMNRTARRRIERRRNRISYLQEIFALEMANIDANFFCRLNDSFYVDSEKRNSRHPFF ATIEEEVAYHKNYRTIYHLREELVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTKNTSVDGVYKQFIQTYNQVFISNIEEGTLAKMEENTTVADILAGKFTRKKLERILQLYPGE KSTGMFAQFISLIVGSKGNFQKVFDLVEKTDIECAKDSYEEDLEALLAIIGDEYAELFVAAKNTYNAVVLSSIITVTDTETNAKLSASMIERFDAHEKDLSELKAFIKLHLPKQYEEIF SNVAIDGYAGYIDGKTKQRDFYKYLKTLLENIEGADYFIAKIEEENFLRKQRTFDNGAIPHQLHLEELEAILHQQAKYYPFLKEAYDKIKSLVTFRIPYFVGPLANGQSDFAWLTRKAD GEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLYYQKYMVYNELTKVRYIDDQGKTNYFSGQEKQQIFNDVFFKQKRKVSKKDLEQFLRNMSHIESPTIEGLEDSFNSSY ATYHDLLKVGIKQEVLENPLNTEMLEDIVKILTVFEDKRMIKEQLQQFSDVLDGAVLKKLERRHYTGWGRLSAKLLVGIRDKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSIIEK EQVSTTDKDLQSIVADLAGSPAIKKGILQSLKIVDELVSIMGYPPQTIVVEMARENQTTVKGKNNSRPRYKSLEKAIKEFGSQILKEHPTDNQELKNNRLYLYYLQNGKDMYTGQELDI HNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGDDVPPLEIVRKRKVFWEKLFQGNLMSKRKFDYLITKAERGGLTEADKATFIHRQLVETRQITKNVANILHQRFNNETDNHGNN MEQVRIVMLKSALVSQFRKQFQLYKVREVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGEYHQFDWFKANKATAKKQFYTNIMLFFAQKERIIDENGEILWDKKYLETIKKVLDYR QMNIVKKTEIQKGEFFSKATIKPKGNSSKLIPRKENWDPMKYGGLDSPNMAYAVIIEHAKGKKKIVIEKKLIQINIMERKMFEKDEKSFLEKQGYRQPKVLTKLPKYTLYECENGRRRML ASANEAQKGNQQVLKGQLITLLHHAKNCEASDGKSLDYIESNREMFGELLAHVSEFAKRYTLADANLSKINQLFEQNKDNDIKVIAQSFVNLMAFNAMGAPASFKFFEATIERKRYTNL KELLSATIIYQSITGLYEARKRLDG	WP_003739838.1 [Listeria monocytogenes]
290	MKNPYTIGLDIGTNSVGWAVLTNQYDLVKRKMKVAGNSDKKQIKKNFWGVRLFDDGQTAVDRRMNRTARRRIERRRNRISYLQEIFAVEMANIDANFFCRLNDSFYVDSEKRNSRHPFF ATIEEEVAYHDNYRTIYHLREKLVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTKNTSVDEVYKQFIETYNQVFMSNIEEGALAKVEENIEVANILAGKFTRRKFERILQLYPGE KSTGMFAQFISLIVGSKGNFQKVFDLIEKTDIECAKDSYEEDLETLLAIIGDEYAELFVAAKNTYNAVVLSSIITVTDTETNAKLSASMIERFDAHEKDLVELKAFIKLNLPKQYEEIF SNAAIDGYAGYIDGKTKQVDFYKYLKTILENIEGSDYFIAKIEEENFLRKQRTFDNGAIPHQLHLEELEAIIHQQAKYYPFLKEDYDKIKSLVTFRIPYFVGPLANGQSEFAWLTRKAD GEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYMVYNELTKIRYIDDQGKTNYFSGREKQQIFNDLFKQKRKVKKKDLELFLRNININHIESPTIEGLEDSFNASY ATYHDLLKVGMKQEIIDNPLNTEMLEDIVKILTVFEDKPMIKEQLQQFSDVLDGGVLKKLERRHYTGWGRLSAKLLVGIREKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSIIEK EQVSTTDKDLQSIVAELAGSPAIKKGILQSLKIVDELVSIMGYPPQTIVVEMARENQTTGKGKNNSKPRYKSLEKAIKEFGSQILKEHPTDNQELKNNRLYLYYLQNGKDMYTGQELDI HNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGGDVPPLEIVRKRKVFWEKLYQGNLMSKRKFDYLITKAERGGLTEADKARFIHRQLVETRQITKNVANILYQRFNKETDNHGNT MEQVRIVTLKSALVSQFRKQFQLYKVREVNGYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGEYHQFDWFKANKATAKKQFYTNIMLFFAQKERIIDENGEILWDKKYLETIKKVLDYR QMNIVKKTEIQKGEFFSKATIKPKGNSSKLIPRKENWDPMKYGGLDSPNMAYAVIIEHAKGKKKIVIEKKLIQINIMERKMFEKDEEAFLEEKGYRHPKVLTKLPKYTLYECEKGRRRML ASANEAQKGNQLVLSNHLVSLLYHAKNCEASDGKSLKYIEAHRETFSELLAQVSEFATRYTLADANLSKINNLFEQNKEGDIKAIAQSFVDLMAFNAMGAPASFKFFEATIDRKRYTNL KELLSSTIIYQSITGLYESRKRLDD	WP_003723650.1 [Listeria monocytogenes]

291	MKNPYTIGLDIGTNSVGWAVLTNQYDLVKRKMKVAGNSDKKQIKKNFWGVRLFDDGQTAVDRRMNRTARRRIERRRNRISYLQEIFAVEMANIDANFFCRLNDSFYVDSEKRNSRHPFF ATIEEEVAYHKNYRTIYHLREELVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTKNTSVDGVYEQFILTYNQVFMSNIEEGTLAKVEENIEVANILAGKFTRREKFERILQLYPGE KSTGMFAQFISLIVGSKGNFQKVFDLIEKTDIECAKDSYEEDLETLLAIIGDEYAELFVAAKNTYNAVVLSSIITVTDTETNAKLSASMIERFDAHEKDLVELKAFIKINLPKQYEEIF SNAAIDGYAGYIDGKTKQVDFYKYLKTTLENVEGADYFITKIEEENFLRKQRTFDNGVIPHQLHLEELEAILHQQAKYYPFLREDYEKIKSLVTFRIPYFVGPLAKGQSEFAWLTRKAD GEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKVMYYNELTKIRYIDDQGKTNYFSGQEKQQIFNDLFKQKRKVKKKDLELFLRNININHVESPTIEGLEDSFNASY ATYHDLMKVGIKQEILDNPLNTEMLEDIVKILTVFEDKRMIKEQLQQFSDVLDGTVLKKLERRHYTGWGRLSAKLLVGIRDKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSIIEK EQVSTADKDLQSIVADLAGSPAIKKGILQSLKVVEELVSVMGYPPQTIVVEMARENQTTNKGKNNSKPRYKSLEKAIKEFGSQILKEHPTDNQELKNNRLYLYYLQNGKDMYTGQELDI HNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGDNVPPLEIVQKRKIFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKARFIHRQLVETRQITKNVANILEQRFNNETDNHGNT MEPVRIVTLKSALVSQFRKQFQLYKVREVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGEYHQFDWFKANKATAKKQFYTNIMLFFAQKERIIDENGEILWDKKYLETIKKVLNYR QMNIVKKTEIQKGEFSNQNPKPRGDSSKLIPKKTNLNPIKYGGFEGSNMAYAIIIEHEKRKKKVTIEKKLIQINIMERKAFEKDEKVFLEGKGYHQPKVLITKLPKYALYECENGRRRML GSANEVHKGNQMLLPNHLMTLLYHAEKREAIDGESLAYIEAHKAVFGELLAHISEFARKYTLANDKLDEINMLYERNKDGDVKSIAESFVSLKKFNAFGVHKDFNFFGTTIKRKRDRKL KELLNSTIIYQSITGLYESRKRLDS	WP_023548323.1 [Listeria monocytogenes]
292	MKNPYTIGLDIGTNSVGWAVLTNQYDLVKRKMKVAGNSDKKQIKKNFWGVRLFDDGQTAVDRRMNRTARRRIERRRNRISYLQEIFAVEMANIDANFFCRLNDSFYVDSEKRNSRHPFF ATIEEEVAYHKNYRTIYHLREELVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTKNTSVDGVYEQFIQTYNQVFMSNIEEGTLAKVEENIEVANILAGKFTRREKFERILQLYPGE KSTGMFAQFISLIVGSKGNFQKVFDLIEKTDIECAKDSYEEDLEALLAIIGDEYAELFVAAKNTYNAVVLSSIITVTATETNAKLSASMIERFDAHEKDLGELKAFIKLHLPKQYQEIF NNAAIDGYAGYIDGKTKQVDFYKYLKTTILENIEGADYFIAKIEEENFLRKQRTFDNGAIPHQLHLEELEAIIHQQAKYYPFLREDYEKIKSLVTFRIPYFVGPLAKGQSEFAWLTRKAD GEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKVMYYNELTKVRYIDDQGKTNYFSGQEKQQIFNDLFKQKRKVKKKDLELFLRNININHVESPTIEGLEDSFNASY ATYHDLLKVGMKQEILDNPLNTEMLEDIVKILTVFEDKPMIKEQLQQFSDVLDGGVLKKLERRHYTGWGRLSAKLLVGIRDKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSIIEK EQVSTTDKDLQSIVADLAGSPAIKKGILQSLKVVEELVSVMGYPPQTIVVEMARENQTTGKGKNNSKPRYKSLEKAIKEFGSKILKEHPTDNQELKNNRLYLYYLQNGKDMYTGQELDI HNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGDVPPLEIVRKRKVFWEKLYQGNLMSKRKFDYLTKAERGGLTDADKARFIHRQLVETRQITKNVANILEQRFNNETDNHGNT MEQVRIVTLKSALVSQFRKQFQLYKVREVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGEYHQFDWFRANKATAKKQFYTNIMLFFGQKERIIDENGEILWDKKYLETIKKVLDYR QMNIVKKTEIQKGEFSKATIKPKGNSSKLIPKRKNWDPMKYGGLDSPNMAYAVIIEHAKGKKKLIFEKKIIRITIMERKMFEKDEEAFLEEKGYRHPKVITKLPKYTLYECEKGRRRML ASANEAQKGNQLVLSNHLVSLLYHAKNCEASDGKSLKYIEAHRETFSELLAQVSEFATRTILADANLSKINNLFEQNKEGDIQAIAQSFVDLMAFNAMGAPASFKFFEATIDRKRYTNL KELLSSTIIYQSITGLYESRKRLDD	WP_014601172.1 [Listeria monocytogenes]
293	MKNPYTIGLDIGTNSVGWAVLTNQYDLVKRKMKVAGNSDKKQIKKNFWGVRLFDDGQTAVDRRMNRTARRRIERRRNRISYLQEIFAVEMANIDANFFCRLNDSFYVDSEKRNSRHPFF ATIEEEVAYHKNYRTIYHLREELVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTKNTSVDGVYEQFIQTYNQVFMSNIEEGTLAKVEENIEVANILARKFTRREKFERILQLYPGE KSTGMFAQFISLIVGSKGNFQKVFDLIEKTDIECAKDSYEEDLETLLAIIGDEYAELFVAAKNTYNAVVLSSIITVTDTETNAKLSASMIERFDAHEKDLVELKAFIKINLPKQYEEIF SNAAIDGYAGYIDGKTKQVDFYKYLKTIILENIEGADYFIAKIEEENFLRKQRTFDNGAIPHQLHLEELEAIIHQQAKYYPFLREDYEKIKSLVTFRIPYFVGPLAKGQSEFAWLTRKAD GEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKVMYYNELTKIRYIDDQGKTNYFSGQEKQQIFNDLFKQKRKVKKKDLELFLRNININHVESPTIEGLEDSFNASY ATYHDLMKVGIKQEILDNPLNTEMLEDIVKILTVFEDKRMIKEQLQQFSDVLDGTVLKKLERRHYTGWGRLSAKLLVGIRDKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSIIEK EQVSTADKDLQSIVADLAGSPAIKKGILQSLKVVEELVSVMGYPPQTIVVEMARENQTTNKGKNNSKPRYKSLEKAIKEFGSQILKEHPTDNQELKNNRLYLYYLQNGKDMYTGQELDI HNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGDNVPPLEIVQKRKIFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKARFIHRQLVETRQITKNVANILEQRFNNETDNHGNT MEPVRIVTLKSALVSQFRKQFQLYKVREVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGEYHQFDWFKANKATAKKQFYTNIMLFFAQKERIIDENGEILWDKKYLETIKKVLNYR QMNIVKKTEIQKGEFSNQNPKPRGDSSKLIPKKTNLNPIKYGGFEGSNMAYAIIIEHEKRKKKVTIEKKLIQINIMERKAFEKDEKVFLEGKGYHQPKVLITKLPKYALYECENGRRRML GSANEVHKGNQMLLPNHLMTLLYHAEKREAIDGESLAYIEAHKAVFGELLAHISEFARKYTLANDKLDEINMLYERNKDGDVKSIAESFVSLKKFNAFGVHKDFNFFGTTIKRKRDRKL KELLNSTIIYQSITGLYESRKRLDS	WP_039320898.1 [Listeria monocytogenes]
294	MKNPYTIGLDIGTNSVGWAVIKDDLTLVRKKIKISGNTDKKEVKKNLWGSFLFEQGDTAQDTRVKRIARRRYERRRFRIRELQKIFDKSMGEVDSNFFHRLDESFLVEEDKEYSKYPIFS NEKEDKNYYDKYPTIYHLRKDLADSNQKADLRLIYLALAHMIKYRGHFLIEGDLKMDGTSISESFQEFIDSYNEVCALEDENLEIIYNDELILTQIENISFKQDISRSKKLDQAIAILFQGV KRQSLFGIFLTLIVGNKANFQKIFNLEDDIKLDLKEEDYDENLEELLSNIDEGYRDVFIQAKNVYNAIELSKIIKTDGKETKAPLSAQMVELYNQHREDLKKYKDIYIKAYLPEKYGETF KDATKNGYAGYIDGKTSQEDFYKFVKAQLKGEENGEYFLEAIENENFLRKQRSFYNGVIPYQIHLQELTAVLDQQEKHYSFLKENRDKIISLLTFRIPYYVGPLAKGESRFAWLERSNS EEKIKPWNFDKIVDIDKSAELFIENLTSRDTYLPDEPVLPKRSLIYQKFTIFNELTKISIVDERGILQNFSSREKIAIFNDLFKNKSKVTKNQLVKYIENKEQIIAPEIKGIEDSFNSN YSTYIDLSKIPDMKKLLEKDEDEIIEEIIKILTIFEDRKMRKRQLMKFKDKLSEKAINQLSKKHYTGWGQLSEKLINGIRDEQSNKTILDYLIDEDNGCPKNMNRNFMQLINDDTLSFKE KIRKAQDINQVNDIKEIVKDLPGSPAIKKGIYQSIRIVDEIIRKMKDRPKNIVIEMARENQTTQEGKNKSKARLKKIQEGLENLDSVHVEKQALDEEMLKSPKYYLYCLQNGKDIYTGK DLDIGQLQTYDIDHIIPRSFITDNSFDNLVLTSSTVNRGKLDNVPSPDIVRQQKGKFWKQLLRAGLMSQRKFNNLTKGKLTDRDRQQFINRQLVETRQITKHVANILLSHHLINEKKEVGEI NIVLLKSALTSQFRKKFDFYKVREVNDYHHAHDAYLNGVIALKLLELYPYMAKDLIYGKYSYHRKIEGDKATQAKYKMSNIIERFSQDLLANPDGEIAWEKDKDLNTIRKVLSSKQINI IKKAEEGKGRLFKETINSRPSKKTEKRIPIKNNLDPNIYGGYIEEKMAYYIAINYLENGKTKKAIVGISIKDKKDFEGQTTEYLGKIGFNKASIINSFELENGSRRMIVGASK ENDSKGELQKGNQMYLPQNLLEFVYHLKHYNEDETSHKFIVEHKAYFDELLNYIVEFANKYILELENSIEKIKDLYHGKGPDVEEKELVESFINLLAITKCGPAADITFLGEKISRKRYR STNCLWGSEVIFQSPTGLYETRIRLE	WP_034440723.1 [Clostridiales bacterium S5-A11]
295	MKVLGNTDRQTVKKNMIGTLLFDSGETAEARRLKRTARRRYTRRINRIKYLQSIFDDEMSKIDSAFFQRIKDSFLVPDDKNDRHPIFGNIKDEVDYHKNYPTIYHLRKKLADSDEKAD LRLIYLALAHIIKFRGHFLIEGDLDSQNTDVNALFLKLVDTYNLMFEDDKIDIQTIDATVILTEKMSKSRRLENLIAKIPNQKKNTLFGNLISLSLGLTPNFKANFELSEDAKLQISKE SFEEDLDNLLAQIGDQYADLFIAAKNLSDAILLSDILTVKGVNTKAPLSASMVQRFNEHQDDLKLLKKLVKVQLPEKYKEIFIDQKNGYAGYINGKTSQEDFYKYIKPILSKLKGAES LISKLEREDFLRKQRTFDNGSIPHQIHLNELKSIIRRQEKYYPFLKDKQVRIEKIFTFRIPYFVGPLANGNSRFAWVKRRSNESITPWNFEEVVDKGASAKVFIERMTNFDTYLPEEKV LPKHSLLYEMFTVYNELTKVKYQAEGMRKPEFISGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLNIIKNQKILDDEQNDSLEDIVLTLTLF EDEKMIAKRLSKYESIFDPSILKKLKKRHYTGWGRLSQKLINGIRDKQTGKTILDFLIEDGQANRNFMQLINDPSLDFASIIKEAQEKTIKSEKLEETIANLAGSPAIKKGILQSVKIV DEVVKVMGYEPSNIVIEMARENQSTQRGINNSRERLRKLEEVHKNIGSKILKEHEISNAQLQSDRVYLYLLQDGKDMYTGKDLDIFDRLSQYDIDHIIPQSFIKDNSIDNIVLTSQESNR GKSDNVPSEVIVKNMKSYWQHQLKSGAISQRKFDNLTKAERGGLSEYDKAGFIKRQLVETRQITKHVAQILNNRFPVKIITLKSKMVSDFRKEFGFYKIREVNDYHH AHDAYINAVVGTALLKKYPKLEAEFVYGDYKHYDLASLVVKSDTSLGKATAKMFFYSNIMNFFKKEVRLADGTVITRPQIETNTETGEIVWDKVKDIKTIRKVLSIPQINVVKKTEVQT GGFSKESILSKRDSDKLIPRKNWDPKKYGGFGSPIIAYSVLVVAKVTKGKSQKTKSVKELVGITIMEQNEFEKDRITFLEKKGYQDIQESLIIKLPKFSLFELENGRKRLLASAKELQ KGNELSLPNKYIQFLYLASRYTSFSGKEEDREKHRHFVESHLHYFDEIKDIIADFSRRYILADANLEKILTLYNEKNQFSIEEQATNMLNLFTFTGLGAPATLKFFNVDIDRKRYTSST EIINSTLIRQSITGLYETRIDLSKIGGD	WP_003107102.1 [Streptococcus parauberis]

296	MMKKEYSIGLDIGTNSVGWAVITDDYKVPAKKMKVLGNTDKQSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTEEMNKVDENFFQRLDDSFLVEEDKQGSKYPI FGTLKEEKEYHKKFPTIYHLREELANSKEKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEEYDKTIEESHLSEITVDALSILTEKVSKSSRLENLIAHYPTEKKNTL FGNLIALSLGLQPNFKTNFQLSEDAKLQFSKDTYEEDLEELLGEIGDEYADLFASAKNLSDAILLSGILAVDDNTTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPDQYNAIFKDKNK KGYAGYIESGVEQDEFYKYLKGILLKINGSGDFLDKIDCEDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARGKGSRFAWAEYKADEKITP WNFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSPLYEFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNSSLG TYHDLRKILDKSPLDDKANEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGTANRNFMQLINDDALSFKEEIARAQ IIGDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSIKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDINNL SQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGAQRRIRN VKIITLKSNLVSNFRKEFELYKVRDINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGDYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKILRKVLSYPQ VNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWNPKKYGGFDSPVVAYSILVFADVEKGKSKKLKRKVQDMVGITIMEKKRFEKNPVDFLEQRGYRNVRLEKIIKLPKYSLFELE NKRRRLLASARELQKGNELVIPQRFTTLLYHSYQIEKNYEPEHREYVEKHKDEFKELLEYISVFSRKYVLADNNLTKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENI DRKRYTSVTECLNATLIHQSITGLYETRIDLSKLGED	WP_037593752.1 [Streptococcus anginosus]
297	MNKAYTIGLDIGTNSVGWAVVTDDYRLMAKKMPVHSKRMEKKIKKNFWGARLFDEGQTAEERRNKRATRRRLIRRRKYRILELQKIFSEEILKKDSHFFARLDESFLIPEDKQYARFPIF PTLLEEKAYYQNYPTIYHLRQKLADSTEKADIRLVYLALAHMIKYRGHFLFEGELDTENTSVEETFKEFIDIYNEQFEEGIIFVKDIPLLILTDKLSKSKKVEKILQVYPKEKTTGCLAQ FLKLIVGNQGNFKQAFHLDEEVKIQLSKETYEEDLEKLLRKSNEEMIDVFLQVKKVYDAILLSGIMTKDTIKAKLSAGMIERYQNHKKDLEELKQFVRAHLHEKVTVFFKDSSKDGY AGYIDGKTTQADYFYKFLKKELTGVDGSEPMLAKIDQENFLLKQRTPTNGSIPHQVHLTEFKAIIDQQKQYYPFLEKSKEKMIQLLTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP NAKDVIDYGASATKFIQRMINYDTYLPTEKVLPKYSMLYQKYTIFNELTKVAYKDDRGIKHQFSSEEKLRIFQELFKKQRRVTKKKLQHFLSANYNIEDAEILGVDKVFNSSYATYHDF LEIAKPTYEKIIELLEQPEMEEMFEDIVKIITIFEDREMVRTQLKKYQRIIGEEIFKKLVKKKYTGWGRLSKRLINGIRDQKTNKTILDYLIDDGISANRNFMQLINDDHLSFKEE IAKELITSQSLLIEVEELIAIIGYKPKNIVIEMARENQTTTGGKNRSKERMKRLKSIEEALKNFGSQILKERPVDNQSLQKDRLYLYYLQNGKDMYTGESL LDIDRLSEYDIDHIIPRSFIVDHSLDNKVLVSSKENRLKKDDVPDSKVVKRMKAYWEKLLRANLISERKFSYLTKLELTDDDKARFIQRQLVETRQITKHVAAILHQYFNQTQELEKEK DIRIITLKSSLVSQFRQVFGIHKVRDINHHHHHAHDAYINAVVAALLKKYPRLAPEFVYGSFAKFHLVKENKATAKKEFYSNIIKFFEKEEQFCDENGEIFWDKRKHIQQIKKVISSHQ VNIVKKTEVQTGSFYKETVNTKEKDFKLIPRKTNKSYWNPKKYGGFGSPVVAYAVFVTYEKGKNTHKKAKAIEGTITMEQAIFEKDPISFLIEKGYSNVNQFIKLPKYTLFELANGQRRMLAS HQELQKANSFILPEKLVTLLYHANHYDEIAYKDSIDYVNEHFSNFQDILDKVIIFAEKVTSAPQKLNQIIATYEKNQEADRKIMAHSFVNLMQFNALGAPADFKFFDTTITRKRYTSLT EIWQSTIIYQSVTGLYETRRRMADLGWGVQ	WP_001040076.1 [Streptococcus agalactiae]
298	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKIRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKYYHEKFPTIYHLRKELADKKEKADLRLVYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNDLLSQDVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEGSEYFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG DLKKILGKDFLDNTDNELILEDIVQTITLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDDGLSFKSIISKAQ HSDNLKEVIGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTAKGLSRSRQRLTTLRESLANLKSNIIEEKKEPKYVKDQVENHHLSDDRLFLYYLQNGKDMYTDDEL DIDNLSQYDIDHIIPQAFIKDDSLDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFVFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRTDKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENKGA NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_017649527.1 [Streptococcus agalactiae]
299	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFPTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAIITDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEGSEYFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETIFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTITLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRVKLLEDGIKELGSQILKEQPVENTQLQNEKLYLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENKGA NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040078.1 [Streptococcus agalactiae]
300	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFPTIYHLRKELADKQEKADLRLIYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVDVEAIITDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGEIYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTITLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRVKLLEDGIKELGSQILKEQPVENTQLQNEKLYLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGLYRKKLSKIVREDKEEKYSEATRKMFFYSNLMNMFKRVVRLADGSIVVRPVIET GRYMGKTAWDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSKFEKNPSAF LESKGYLNIIRDDKLIKNILMFTFSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLSKLGED	WP_001040078.1 [Streptococcus agalactiae]

301	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNITFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040080.1 [Streptococcus agalactiae]
302	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTIFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040081.1 [Streptococcus agalactiae]
303	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIIQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040083.1 [Streptococcus agalactiae]
304	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDDLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040085.1 [Streptococcus agalactiae]
305	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040087.1 [Streptococcus agalactiae]

306	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGGD	WP_001040088.1 [Streptococcus agalactiae]
307	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRDDKLMILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040089.1 [Streptococcus agalactiae]
308	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040090.1 [Streptococcus agalactiae]
309	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLRQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040091.1 [Streptococcus agalactiae]
310	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_017771984.1 [Streptococcus agalactiae]

311	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVASILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVIKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_050198062.1 [Streptococcus agalactiae]
312	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVASILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_050201642.1[St reptococcus agalactiae]
313	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_050881965.1 [Streptococcus agalactiae]
314	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKDIIRRQSEYYPFLKENQDRIEKILTFRIPYYVGPLARGNSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKQLLDFLAKEFEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHSNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVLADIKKGKAQKLKTVKELIGITIMERERFEKNPSAFLESKGYLN IRTDKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQYMKFLYLASRYNESKGKPEEIEKKQEFVNQHISYFDDILQLINDFSKRVILADANLEKINKLYSDNKDNTPVDELAK NIINLFTFTSLGAPAAFKFFDKSVDRKRYTSTKEVLDSTLIHQSITGLYETRIDLGKLGED	WP_050886065.1[St reptococcus agalactiae]
315	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATMQEEKDYHEKFPTIYHLRKELADKKEKADLRLVYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTSFENNHLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSAYMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSEYFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDKIEKILTFRIPYYVGPLARGNSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGEIYFFDSNVKQEIFDGVFKEHRKVSKKQLLDFLAKEFEEFRIVDVTGLDKENKAFNASLG TYHDLKILDKDFLDNPDNESILEDIVQTITLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDRESQKTILDYLISDGRANRNFMQLINDGLSFKSIISKAQ SGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARKAFWKKLLDAKLISQRKYDNLITKAERGGLTPDDKAGFIQRQLVETRQITKHVARILDERFNNKVDDNN DIDNLSQYDIDHIVPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSIEIVKARKAFWKKLLDAKLISQRKYDNLITKAERGGLTPDDKAGFIQRQLVETRQITKHVARILDERFNNKVDDNN KPIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADETVVVKDDIEVNNETGEIA WDKKKHFATVRKVLSYPQVNIVKKTEVQTGGFSKESILAHSNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVLADIKKGKAQKLKTVKELIGITIMERERFEKNPSAFLESKGYLN IRTDKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQYMKFLYLASRYNESKGKPEEIEKKQEFVNQHISYFDDILQLINDFSKRVILADANLEKINKLYSDNKDNTPVDELAK NIINLFTFTSLGAPAAFKFFDKSVDRKRYTSTKEVLDSTLIHQSITGLYETRIDLGKLGED	WP_001040092.1 [Streptococcus agalactiae]

316	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATMQEEKDYHEKFPTIYHLRKELADKKEKADLRLVYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTSFENNHLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSEYFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGNSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNVKQEIFDGVFKEHRKVSKKQLLDFLAKEFEEFRIVDVTGLDKENKAFNASLG TYHDLKKILDKDFLDNPDNESILEDIVQTITLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDRESQKTILDYLISDGRANRNFMQLINDDGLSFKSIISKAQ SGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLEDGVKNLASDILNGDILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIVPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSIDIVKARKAFWKKLLDAKLISQRKYDNLTKAERGGLTPDDKAGFIQRQLVETRQITKHVARILDERFNNKVDDNN KPIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADETVVVKDDIEVNNEIGEIA WDKKKHFATVRKVLSYPQVNIVKKTEVQTGGFSKESILPKRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVKELIGITIMERSRFEKNPSAFLESKGYLN IRTDKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQYMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYSDNKDNTPVDELAK NIINLFTFTSLGAPAAFKFFDKSVDRKRYTSTKEVLDSTLIHQSITGLYETRIDLGKLGED	WP_015058523.1 [Streptococcus agalactiae]
317	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATMQEEKYYHEKFPTIYHLRKELADKKEKADLRLIYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAKYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEGSEYLLEKIKNEDFLRKQRTFDNGSIPHQVHLTELRAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRFLAEGFKDFQFLNRKQKETIFNSLFKEKRKVTEKDIISFLNKVDGYEGIAIKGIEKQFNASLSTYH DLKKILGKDFLDNTDNELILEDIVQTITLFEDREMIRKRLDIYKDFFTESQLKKLYRRHYTGWGRLSAKLINGIRNKENQKTILDYLIDDGSANRNFMQLIKDAGLSFKPIIDKARTGS HSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTAKGLSRSRQRLITTLRESLANLKSNIIEEKKEKPKYVKDQVENHHLSDDRLFLYYLQNGKDMYTGDEL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSVEIVKDCKVFWKKLLDAKLMSQRKYDNLITKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILPKRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRDDKLMILPKYSLFELENGRRRLLASAGELAIPTQFMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENIPVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040094.1[St reptococcus agalactiae]
318	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATMQEEKYYHEKFPTIYHLRKELADKKEKADLRLIYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNDLLSQNVDVEAILTDKISKSAKKDRILARYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEGSEYLLEKIKNEDFLRKQRTFDNGSIPHQVHLTELRAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRFLAEGFKDFQFLNRKQKETIFNSLFKEKRKVTEKDIISFLNKVDGYEGIAIKGIEKQFNASLSTYH DLKKILGKDFLDNTDNELILEDIVQTITLFEDREMIRKRLDIYKDFFTESQLKKLYRRHYTGWERLSAKLINGIRNKENQKTILDYLIDDGSANRNFMQLIKDAGLSFKPIIDKARTGS HLDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTAKGLSRSRQRLITTLRESLANLKSNIIEEKKEKPKYVKDQVENHHLSDDRLFLYYLQNGKDMYTDDEL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLITKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEVQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRTDKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQYMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYSDNKDNTPVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040095.1 [Streptococcus agalactiae]
319	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATMQEEKYYHEKFPTIYHLRKELADKKEKADLRLIYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNDLLSQNVDVEAILTDKISKSAKKDRILARYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEGSEYLLEKIKNEDFLRKQRTFDNGSIPHQVHLTELRAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRFLAEGFKDFQFLNRKQKETIFNSLFKEKRKVTEKDIISFLNKVDGYEGIAIKGIEKQFNASLSTYH DLKKILGKDFLDNTDNELILEDIVQTITLFEDREMIRKRLDIYKDFFTESQLKKLYRRHYTGWGRLSAKLINGIRNKENQKTILDYLIDDGSANRNFMQLIKDAGLSFKPIIDKARTGS HLDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTAKGLSRSRQRLITTLRESLANLKSNIIEEKKEKPKYVKDQVENHHLSDDRLFLYYLQNGKDMYTDDEL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLITKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEVQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRDDKLMILPKYSLFELENGRRRLLASAGELAIPTQFMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENIPVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040096.1[St reptococcus agalactiae]
320	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATMQEEKYYHEKFPTIYHLRKELADKKEKADLRLIYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNDLLSQNVDVEAILTDKISKSAKKDRILARYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEGSEYLLEKIKNEDFLRKQRTFDNGSIPHQVHLTELRAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRFLAEGFKDFQFLNRKQKETIFNSLFKEKRKVTEKDIISFLNKVDGYEGIAIKGIEKQFNASLSTYH DLKKILGKDFLDNTDNELILEDIVQTITLFEDREMIRKRLDIYKDFFTESQLKKLYRRHYTGWGRLSAKLINGIRNKENQKTILDYLIDDGSANRNFMQLIKDAGLSFKPIIDKARTGS HSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTAKGLSRSRQRLITTLRESLANLKSNIIEEKKEKPKYVKDQVENHHLSDDRLFLYYLQNGKDMYTDDEL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSVEIVKDCKVFWKKLLDAKLMSQRKYDNLITKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRDDKLMILPKYSLFELENGRRRLLASAGELAIPTQFMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENIPVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040097.1[St reptococcus agalactiae]

321	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATMQEEKYYHEKFPTIYHLRKELADKKEKADLRLIYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNDLLSQNVDVEAILTDKISKSAKKDRILARYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLIYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEGSEYLLEKIKNEDFLRKQRTFDNGSIPHQVHLTELRAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRFLAEGFKDFQFLNRKQKETIFNSLFKEKRKVTEKDIISFLNKVDGYEGIAIKGIEKQFNASLSTYH DLKKILGKDFLDNTDNEILLEDIVQTLTLFEDREMIRKRLDIYKDFFTESQLKKLYRRHYTGWGRLSAKLINGIRNKENQKTILDYLIDDGSANRNFMQLIKDAGLSFKPIIDKARTGS HSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTAKGLSRSRQRLITTLRESLANLKSNIILEEKKEPKYVKDQVENHHLSDDRLFLYYLQNGKDMYTDDEL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSVEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRDDKIMILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENIPVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040098.1[St reptococcus agalactiae]
322	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATMQEEKYYHEKFPTIYHLRKELADKKEKADLRLIYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNDLLSQNVDVEAILTDKISKSAKKDRILARYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLIYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEGSEYLLEKIKNEDFLRKQRTFDNGSIPHQVHLTELRAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRFLAEGFKDFQFLNRKQKETIFNSLFKEKRKVTEKDIISFLNKVDGYEGIAIKGIEKQFNASLSTYH DLKKILGKDFLDNTDNEILLEDIVQTLTLFEDREMIRKRLDIYKDFFTESQLKKLYRRHYTGWGRLSAKLINGIRNKENQKTILDYLIDDGSANRNFMQLIKDAGLSFKPIIDKARTGS HSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTAKGLSRSRQRLITTLRESLANLKSNIILEEKKEPKYVKDQVENHHLSDDRLFLYYLQNGKDMYTDDEL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSVEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPKVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRDDKIMILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENIPVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHKSITGLYETRIDLGKLGED	WP_001040099.1[St reptococcus agalactiae]
323	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATMQEEKYYHEKFPTIYHLRKELADKKEKADLRLIYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNDLLSQNVDVEAILTDKISKSAKKDRILARYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLIYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEGSEYLLEKIKNEDFLRKQRTFDNGSIPHQVHLTELRAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRFLAEGFKDFQFLNRKQKETIFNSLFKEKRKVTEKDIISFLNKVDGYEGIAIKGIEKQFNASLSTYH DLKKILGKDFLDNTDNEILLEDIVQTLTLFEDREMIRKRLDIYKDFFTESQLKKLYRRHYTGWGRLSAKLINGIRNKENQKTILDYLIDDGSANRNFMQLIKDAGLSFKPIIDKARTGS HSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTAKGLSRSRQRLITTLRESLANLKSNIILEEKKEPKYVKDQVENHHLSDDRLFLYYLQNGKDMYTDDEL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSVEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRDDKIMILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENIPVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHKSITGLYETRIDLGKLGED	WP_001040100.1 [Streptococcus agalactiae]
324	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATMQEEKYYHEKFPTIYHLRKELADKKEKADLRLIYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNDLLSQNVDVEAILTDKISKSAKKDRILARYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLIYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEGSEYLLEKIKNEDFLRKQRTFDNGSIPHQVHLTELRAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRFLAEGFKDFQFLNRKQKETIFNSLFKEKRKVTEKDIISFLNKVDGYEGIAIKGIEKQFNASLSTYH DLKKILGKDFLDNTDNEILLEDIVQTLTLFEDREMIRKRLDIYKDFFTESQLKKLYRRHYTGWGRLSAKLINGIRNKENQKTILDYLIDDGSANRNFMQLIKDAGLSFKPIIDKARTGS HSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTAKGLSRSRQRLITTLRESLANLKSNIILEEKKEPKYVKDQVENHHLSDDRLFLYYLQNGKDMYTDDEL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSVEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRDDKIMILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENIPVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_017643650.1 [Streptococcus agalactiae]
325	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATMQEEKYYHEKFPTIYHLRKELADKKEKADLRLIYLALAHIIKFRGHFLIEDDSFDVRNTDIQKQYQAFLEIFDTTFENNDLLSQNVDVEAILTDKISKSAKKDRILARYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLIYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEGSEYLLEKIKNEDFLRKQRTFDNGSIPHQVHLTELRAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRFLAEGFKDFQFLNRKQKETIFNSLFKEKRKVTEKDIISFLNKVDGYEGIAIKGIEKQFNASLSTYH DLKKILGKDFLDNTDNEILLEDIVQTLTLFEDREMIRKRLDIYKDFFTESQLKKLYRRHYTGWGRLSAKLINGIRNKENQKTILDYLIDDGSANRNFMQLIKDAGLSFKPIIDKARTGS HSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTAKGLSRSRQRLITTLRESLANLKSNIILEEKKEPKYVKDQVENHHLSDDRLFLYYLQNGKDMYTDDEL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSVEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRDDKIMILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENIPVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHKSITGLYETRIDLGKLGED	WP_047209694.1 [Streptococcus agalactiae]

326	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYXIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040104.1 [Streptococcus agalactiae]
327	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKMRVLGNTDKEYIKKNLIGALLFDGGNTADRRLKRTSRRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040105.1 [Streptococcus agalactiae]
328	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTASDRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEDDKRGSKYPIF ATMQEEKDYHEKFPTIYHLRKELADKKEKADLRLFYLALAHIIKFRGHFLIEDDSFDVRNTDIQRQYQAFLEIFDTTFENNHLLSQNIDIEGILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSVAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFTDSSK DGYAGYIEGKTNQGAFYKYLSKLLTKQEGSEYFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEELVDKEASAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLKKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLIHDDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGKAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRDDKLMILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_017647151.1 [Streptococcus agalactiae]
329	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTASDRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEDDKRGSKYPIF ATMQEEKDYHEKFPTIYHLRKELADKKEKADLRLFYLALAHIIKFRGHFLIEDDSFDVRNTDIQRQYQAFLEIFDTTFENNHLLSQNIDVEGILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSVAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFTDSSK DGYAGYIEGKTNQGAFYKYLSKLLTKQEGSEYFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEELVDKEASAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLKKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLIHDDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGKAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRDDKLMILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_017648376.1 [Streptococcus agalactiae]
330	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKMRVLGNTDKEYIKKNLIGALLFDGRNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_047207273.1 [Streptococcus agalactiae]

331	MNKPYSIGLDIGTNSVGYSVVTDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTASDRRLKRTARRYTCRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEDDKRGSKYPIF ATMQEEKDYHEKFPTIYHLRKELADKKEKADLRLVYLALAHIIKFRGHFLIEDDSFDVRNTDIQRQYQAFLEIFDTTFENNHLLSQNIDVEGILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSVAKKLIYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFTDSSK DGYAGYIEGKTNQGAFYKYLSKLLTKQEGSEYFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEELVDKEASAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLKKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLISDGRANRNFMQLIHDDGLSFKPIIDKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQVVVEMARENQTTNQGRRNTRQRYKLLEEGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIVPQAFIKDDSIDNRVLVSSAKNRGKSDNVPSIDIVKARKAFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERFRFEKNPSAFLESKGYLN IRDDKIMILPKYSLFELENGRRRLLASAGELQKGNELALPTQYMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040106.1 [Streptococcus agalactiae]
332	MNKPYSIGLDIGTNSVGYSVVTDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTASDRRLKRTARRYTRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEDDKRGSKYPIF ATMQEEKDYHEKFPTIYHLRKELADKKEKADLRLVYLALAHIIKFRGHFLIEDDSFDVRNTDIQRQYQAFLEIFDTTFENNHLLSQNIDVEGILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSVAKKLIYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFTDSSK DGYAGYIEGKTNQGAFYKYLSKLLTKQEGSEYFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEELVDKEASAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLKKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLISDGRANRNFMQLIHDDGLSFKPIIDKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQVVVEMARENQTTNQGRRNTRQRYKLLEEGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIVPQAFIKDDSIDNRVLVSSAKNRGKSDNVPSIDIVKARKAFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERFRFEKNPSAFLESKGYLN IRDDKIMILPKYSLFELENGRRRLLASAGELQKGNELALPTQYMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040107.1 [Streptococcus agalactiae]
333	MNKPYSIGLDIGTNSVGYSVVTDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTASDRRLKRTARRYTRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEDDKRGSKYPIF ATMQEEKDYHEKFPTIYHLRKELADKKEKADLRLVYLALAHIIKFRGHFLIEDDSFDVRNTDIQRQYQAFLEIFDTTFENNHLLSQNIDVEGILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSVAKKLIYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFTDSSK DGYAGYIEGKTNQGAFYKYLSKLLTKQEGSEYFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEELVDKEASAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLKKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLISDGRANRNFMQLIHDDGLSFKPIIDKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQVVVEMARENQTTNQGRRNTRQRYKLLEEGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGETL DIDNLSQYDIDHIVPQAFIKDDSIDNRVLVSSAKNRGKSDNVPSIDIVKARKAFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHHAHDAYLNAVVAKAILTKYPQLEREFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERFRFEKNPSAFLESKGYLN IRDDKIMILPKYSLFELENGRRRLLASAGELQKGNELALPTQYMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040108.1 [Streptococcus agalactiae]
334	MNKPYSIGLDIGTNSVGYSVVTDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTASDRRLKRTARRYTRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEDDKRGSKYPIF ATMQEEKDYHEKFPTIYHLRKELADKKEKADLRLVYLALAHIIKFRGHFLIEDDSFDVRNTDIQRQYQAFLEIFDTTFENNHLLSQNIDVEGILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSVAKKLIYDSVLLSGILTVTALSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFTDSSK DGYAGYIEGKTNQGAFYKYLSKLLTKQEGSEYFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEELVDKEASAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLKKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLISDGRANRNFMQLIHDDGLSFKPIIDKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQVVVEMARENQTTNQGRRNTRQRYKLLEEGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIVPQAFIKDDSIDNRVLVSSAKNRGKSDNVPSIDIVKARKAFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERFRFEKNPSAFLESKGYLN IRDDKIMILPKYSLFELENGRRRLLASAGELQKGNELALPTQYMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_017771611.1 [Streptococcus agalactiae]
335	MNKPYSIGLDIGTNSVGYSVVTDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTASDRRLKRTARRYTRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEDDKRGSKYPIF ATMQEEKDYHEKFPTIYHLRKELADKKEKADLRLVYLALAHIIKFRGHFLIEDDSFDVRNTDIQRQYQAFLEIFDTTFENNHLLSQNIDVEGILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSVAKKLIYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFTDSSK DGYAGYIEGKTNQGAFYKYLSKLLTKQEGSEYFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEELVDKEASAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLKKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLISDGRANRNFMQLIHDDGLSFKPIIDKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQVVVEMARENQTTNQGRRNTRQRYKLLEEGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIVPQAFIKDDSIDNRVLVSSAKNRGKSDNVPSIDIVKARKAFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERFRFEKNPSAFLESKGYLN IRDDKIMILPKYSLFELENGRRRLLASAGELQKGNELALPTQYMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040109.1 [Streptococcus agalactiae]

336	MNKPYSIGLDIGTNSVGYSVVTDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTASDRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEDDKRGSKYPIF ATMQEEKDYHEKFPTIYHLRKELADKKEKADLRLVYLALAHIIKFRGHFLIEDDSFDVRNTDIQRQYQAFLEIFDTTFENNHLLSQNIDVEGILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSVAKKLIVDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFTDSSK DGYAGYIEGKTNQGAFYKYLSKLLTKQEGSEYFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEELVDKEASAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLKKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLISDGRANRNFMQLIHDDGLSFKPIIDKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQVVVEMARENQTTNQGRRNTRQRYKLLEEGVKNLASDILNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIVPQAFIKDDSIDNRVLVSSAKNRGKSDNVPSEDIVKKMKSYWRQLLNAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERFRFEKNPSAFLESKGYLN IRDDKIMILPKYSLFELENGRRRLLASAGELQKGNELAIPTQYMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_001040110.1 [Streptococcus agalactiae]
337	MNKPYSIGLDIGTNSVGYSVVTDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTASDRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEDDKRGSKYPIF ATMQEEKDYHEKFPTIYHLRKELADKKEKANLRLVYLALAHIIKFRGHFLIEDDSFDVRNTDIQRQYQAFLEIFDTTFENNHLLSQNIDVEGILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSVAKKLIVDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFTDSSK DGYAGYIEGKTTQEAFYKYLSKLLTKQEGSEYFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEELVDKEASAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLKKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLISDGRANRNFMQLIHDDGLSFKPIIDKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQVVVEMARENQTTNQGRRNTRQRYKLLEEGVKNLASDILNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIVPQAFIKDDSIDNRVLVSSAKNRGKSDNVPSEDIVKKMKSYWRQLLNAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQVNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERFRFEKNPSAFLESKGYLN IRDDKIMILPKYSLFELENGRRRLLASAGELQKGNELAIPTQYMKFLYLASRYNELKGKPEEIEQKQEFVVQHVSYFDDILQIINDFSNRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	WP_050204027.1 [Streptococcus agalactiae]
338	MTNNKPYSIGLDIGTNSVGWAVITDDYKVPSKKMKVLGNTDKHFIKKNLIGALLFDEGTTAEDRRLKRTARRRYTRRKNRLRYLQEIFTEEMSKVDISFFHRLDDSFLVPEDKRGSKYPI FATLEEEKEYHKNFPTIYHLRKHLADSKEKADFRLIYLALAHIIKYRGHFLYEESFDIKNNDIQKIFNEFISIYDNTFEGSSLNGQNAQVEAIFTDKISKSAKRERVLKLFPDEKSTGL FSEFLKLIVGNQADFKKHFDLEEKAPLQFSKDTYDEDLENLLVQIGDDFADLFLVAKKLYDAILLSGILTVTDPSTKAPLSASMIDRYENHQKDLAALKQFIKTNLPEKYDEVFSDQSK DGYAGYIDGKTTQEAFYKYIKNLLSKLEGADYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMNAIIRRQGEHYPFLQENKEKIEKILTFRIPYYVGPLARGNRDFAWLTRNSDQAIRP WNFEEIVDKARSAEDFINKMTNYDLYLPEEKVLPKHSLLYEYFTAVYNELTKVRFIAEGLRDYQFFDSGQKKQIVNQLFKEKRKVTEKDIIHYLHNVDGYDGIELKGIEKQFNASLSTYH DLLKIIKDKEFMDDSKNEAILENIVHTLTIFEDREMIKQRLAHYASIFDEKVIKALTRRHYTGWGKLSAKLINGICDKKTGKTILDYLIDDGYNNRNFMQLINDDGLSFKEIIQKAQVV GKTNDVKQVVELPGSPAIKKGILQSLKIVDELVKVMGHAPESIVIEMARENQTTAGKKNSQQRYKRIEDALKNLAPGLDSTISKENPTDNIQLQNDRLFLYYLQNGKDMYTGKSLDI NQLSSCDIDHIIPQAFIKDDSLDNRVLTSSKDNRGKSDNVPSLEIVQKRKAFWQQLLDSKLISERKFNNLTKAERGGLDERDKVGFIKRQLVETRQITKHVARILDARFNTEVDSKDK DRSVKIITLKSNLVSNFRKEFRLYKVREINDYHHAHDPYLNAVVAKAILKKYPKLEPEFVYGDYQKYDLKRYISRTKDPKEVEKATEKYFFYSNLLNFFKEEVHYADGTIVKRENIEYS KDTGEIAWNKEKDFATIKKVLSLPQVNIVKKTEEQTVGQNGGLFDNNIVSKKKVVDASKLTPIKSGLSPEKYGGYARPTIAYSVLVIADIEKGKAKKLKRIKEMVGITVQDKKKFEANP IAYLEECGYKNINPNLIIKLPKYSLFELENGRRRLLASAGELQKGNELIVPYHFTALLYHAQRINKISEPIHKQYVETHQSEFKELLTAIISLSKKYIQKPNVESLLQQAFDQSDKDIY QLSEFSILLKLISFGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLSKLGED	WP_045618028.1 [Streptococcus mitis]
339	MNNNNYSIGLDIGTNSVGWAVITDDYKVPSKKMRVLGNTDKRFIKKNLIGALLFDEGTTAEDRRLKRTARRYTRKNRLRYLQEIFAEEMSKVDSSFFHRLDDSFLVPEDKSGSKYPI FATLAEEKEYHKKFPTIYHLRKHLADSKEKADLRLIYLALAHIIKYRGHFLYEEAFDIKNNDIQKIFNEFINIYDNTFEGSSLSGQNAQVEAIFTGKISKSVKRERVLKFLPDEKSTGL FSEFLKLIVGNQADFKKHFDLEEKASLQFSKDTYDEDLENLLGQIGDDFADLFLVAKKLYDAILLSGILTVTDPSTKAPLSASMIERYENHQEDLAALKQFIKTNLPEKYAEVFSDQSK DGYAGYIDGKTTQEAFYKYIKNLLSKFEGADYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMNAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKKADEKITPW WNFEEIVDKASSAESFINKMTNYDLYLPEEKVLPKHSLLYETFAVYNELTKVRFIAEGLRKYQFFESGQKKQIVNQLFKEKRKVTEKDIIHYLHNVDGYDGIELKGIEKQFNASLSTYH DLLKIIKDKEFMDNHKNQEILENIVHTLTIFEDREMIKQRLAQYDSIFDEKVIKALTRRHYTGWGKLSAKLINGICDKKTGKTILDYLIDDGYNNRNFMQLINDDGLSFKEIIQKAQVV GKTDDLTQVVRELSGSPAIKKGILQSIKIVDELVKVMGYAPESIVIEMARENQTTAKGKKNSQQRYKRIEDALKNLAPGLDSTISKENPTDNIQLQNDRLFLYYLQNGKDMYTGEALDI NQLSSYDIDHIIPQAFIKDDSLDNRVLTSSKDNRGKSDNVPSLEVVKKRKAFWQQLLDSKLISERKFNNLTKAERGGLDERDKVGFIKRQLVETRQITKHVARILDARFNTEVSEKNQK IRSVKIITLKSNLVSNFRKEFKLYKVREINDYHHAHDAYLNAVVAKAILKKYPKLEPEFVYGDYQKYDLKRYISKSKDPKEVEKATEKYFFYSNLLNFFKEEVHYADGTIVKRENIEYS KDTGEIAWNKEKDFVTIKKVLSYPQVNIVKKREVQTGGFSKESILPKGNSDKLIPRKTKDILWDTTKYGGFDSPVIAYSILLIADIEKGKAKKLKTVKTLVGITIMEKAAFEKNPIITFL ENKGYHNVRKENILCLPKYSLFELENGRRRLLASAKELQKGNEIVLPVYLTLLYHSKNVHKLDEPEHLEYIQKHRYEFKDLLNLVSEFSQKYVLAEANLEKIKSLYVDNEQADIEILA NSFINLLTFTALGAPAAFKFFGKDVDRKRYTTVSEILNATLIHQSITGLYETRIDLSKLGED	WP_009754323.1 [Streptococcus sp. taxon 056]
340	MNQKYSIGLDIGTNSVGWAVITDDYKVPAKKMKVLGNTDKQSIKKNLIGALLFDSGETAEATRLKRTARRYTRRRNRLRYLQEIFAEEMMQVDESFFQRLDDSFLVDEDKRGERHPIF GNIAAEVKYHDEPFTIYHLRKHLADSTKQADLRLVYLALAHMIKFRGHFLIEGQLKAENTNVQALFKDFVEVYDKTVEESHLSEITVDALSILITEKVSKSRRLENLIAHYPAEKKNTLF GNLIALFLGLQPNFKTNFQLSEDAKLQFSKDTYEEDLEGLLGEVGDEYADLFASAKNLYDAILLSGILTVDDNSTKAPLSASWVKRYEEHQKDLKKFEDFIKVNALDQYNAIFKDKNKK GYAGYIESGVKQDEFYKYLKGILLQINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKADEKITPW NFDDILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSLLYEYFTVYNELTKVRVNEQGEAKFFDANMKQEIFDHVFKENRKVTKDKLLNYLNKEFEEFRIVNLIGLDKENKVFNSSLGT YHDLRKIINKSFLDNKENAQIIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIAKAQI IGDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDRGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDINNLS QYDIDHIIPQAYIKDDSFDNRVLTSSSENRGKSDNVPSIEVVCARKADWMRLRKAGLISQRKFDNLTKAERGGLTENDKAGFIKRQLVETRQITKHVAQVLDARFNAKHDENKKVIRDV KIITLKSNLVSQFRKDFKFYKVREINDYHHAHDAYLNAVIGTALLKKYPKLASEFVYGEFKKYDVRKFIAKSDKEIGKATAKYFFYSNLMNFFKKEVKFADGTVVERPDIETSEDGEIA WNKQTDFKIVRKVLSYPQVNIVKKTEVQTHGLDRGKPRGLFNANPSPKPKPDSSENLVGVKRNLDPKKYGGYAGISNSYAVLVKAIIEKGVKKETMVLEFQGISILDRITFEKDKRAF LLGKGYKDIHKKIIELPKYSLFELENGRRRLLASISTNNKRGEIHKGNELFVPQKFTTLLYHAKRINNPINKDHIEYVKKHRDDFKELLNYVLEFNEKYVGATKNGERLKEAVADFDSK SNEEICTSFLGAVNSKNAGLFELTSLGSASDFEFLGVKIPRYRDYTPSSLLKDSTLIHQSITGLYETRIDLSKLGED	WP_003041502.1 [Streptococcus anginosus]

341	MQKNYTIGLDIGTNSVGWAVMKDDYTLIRKRMKVLGNTDIKKIKKNFWGVRLFDEGETAKETRLKRGTRRRYQRRNRLIYLQDIFQQPMLAIDENFFHRLDDSFFVPDDKSYDRHPIF GSLEEEVAYHNTYPTIYHLRKHLADNPEKADLRLVYTALAHIVKYRGHFLIEGELNTENTSISETFEQFLDTYSDIFKEQLVGDISKVEEILSSKQSRSRKHEQIMALFPNENKLGNFG RFMMLIVGNTSNFKPVFDLDDEYKLKLSDETYEEDLDTILLGMTDDVFLDVFMAAKNVYDAVEMSAIISTDTGNSKAVLSNQMINFYDEHKVDLAQLKQFFKTHLPDKYYECFSDPSKNG YAGYIDGKTNQEDFYKYIEKVMKTTKSDKDYFLDKIDREVFLRKQRSFYNSVIPHQIHLQEMQAILDRQSQYYPFLAENRDKIESLVTFRIPYYVGPLTVSDQSEFAWMERQSDEPIR PWNFDEIVNKERSAEKFIERMTNMDTYLLEEKVLPKRSLLYQTFEVYNELTKVRYTNEQGKTEKINRQQKAEIIETLFKQKNRVREKDIANYLEQYGVDGTDIKGVEDKFNASLSTYN DLAKIDGAKAYLDDPEYADVWEDIIKILTIFEDKAMRKKQLQTYSDTILSPEILKKLERKHYTGWGRFSKKLINGLRDEGSGNKTILDYLKSDEGSSGPTNRNFMQLIRDNTLSFKKKIED AQTIEDTTHIYDTVAELPGSPAIKKGIRQALKIVEEIIDIIGYEPENIVVEMARESQTTKGKDLSKERLERLKLTEAIKEFDGPSDVKVKDLKNENLRNDRLYLYYLQNGRDMYTNEPLD INNLSKYDIDHIIPQSFTTDNSIDNKVLVSRTKNQGNKSDDVPSINIVKKMKPFWRQLHKAGLISDRKFKNLTKAEHGGLTEADRAHFLNRQLVETRQITKHVANLLDSQYNTAEEQRI NIVLLKSSMTSRFRKEFKLYKVREINDYHHGHDAYLNAVVATTIMKVYPNLKPQFVYGQYKKTSMFKEEKATARKHFYSNITKFFKKEKVVNEETGEILWDTERHLSTIKRVLSWKQMN IVKKVEKQKGQLWKETIYPKGDSSKLIPVKEGMDPQKYGGLSQVSEAFAVVITHEKGKKKQLKSDLISIPIVDQKAYEQHPTAYLEEAGYNNPTVLHELFKYQLFELEDGSRRMIASAK EFQKGNQMVLPLELVELLYHANRYDKVKFPDSIEYVHDNLAKFDDLLEVVIDFSNKYINADKNVQKIQKIYKEHGTEDVELTVESFVNLMTFTAMGAPATFKFFGESITRSRYTSITEF RGSTLIFQSITGLYETRYKLEDN	WP_004636532.1 [Dolosigranulum pigrum]
342	MRKPYSIGLDIGTNSVGWAVITDDYKVPSKKMRIQGTTDRTSIKKNLIGALLFDNGETAEATRLKRTTRRRYTRRKYRIKELQKIFSSEMNELDIAFFPRLSESFLVSDDKEFENHPIF GNLKDEITYHNDYPTIYHLRQTLADSDQKADLRLIYLALAHIIKFRGHFLIEGNLDSENTDVHVLFLNLVNIYNNLFEEDIVETASIDAEKILTSKTSKSRRLENLIAEIPNQKRNMLIF GNLVSLALGLTPNFKTNFELLEDAKLQISKDSYEEDLDNLLAQIGDQYADLFIAAKKLSDAILLSDILRVNTEITKAPLSASMVQRYEEHQQDLALLKNLVKKQLPEKYKEIFDNKERN GYAGYIDGKTTQEEFYKYIKPILEKMDGTEKLLSKLEREDFLRKQRTFDNGSIPHQIHLNELKAIIRRQEEFYPFLKENQDKIEKLFTFRIPYYVGPLANGQSSFAWLKRQSNESITPW NFEEVVDQEASARAFIERMTNFDTYLPEEKVLPKHSLLYEMFMVYNELTKVKYQTEGMKRPVFLSSEDKEEIVNLLFKKERKVTVKQLKEEYFSKMKCFHTVTILGVEDRFNASLGTYH DLLKIFKDKAFLDDEANQDILEEIVWTLTLFEDQAMIERRLVKYADVFEKSVLKKLKKRHYTGWGRLSQKLINGIKDKQTGKTILGFLKDDGVANRNFMQLINDSSLDFAKIIKNEQEK TIKNESLEETIANLAGSPAIKKGILQSIKIVDEIVKIMGQNPDNIVIEMARENQSTMQGIKNSRQRIKELEKSIKELGSKILKEYPVENVSNTQLQSDRLYLYYLQNGKDMYTGKELDYDNL SQYDIDHIIPQSFIKDNSIDNTVLTTQASNRGKSDNVPNIETVNKMKSFWYKQLKSGAISQRKFDHLTKAERGALSDFDKAGFIKRQLVETRQITKHVAQILDSRFNSNLTEDSKSNRN VKIITLKSKMVSDFRKDFGFYKLREVNDYHHAQDAYLNAVVGTALLKKYPKLEAEFVYGDYKHYDLAKLMIQPDSSLGKATTRMFFYSNLMNFFKKEIKLADDTITRPQIEVNTETGE IVWDKVKDMQTIRKVMSYPQVNIVMKTEVQTGGFSKESIWPKGDSDKLIARKKSWDPKKYGGFDSPIIAYSVLVVAKIAKGKTQKLKTIKELVGIKIMEQDEFEKDPIAFLEKKGYQDI QTSSIKLPKYSLFELENGRKRLLASAKELQKGNELALPNKYVKFLYLASHYTKFTGKEDREKKKRSYVESHLYYFDEIMQIIVEYSNRYIILADSNLIKIQNLYKEKDNFSIEEQAINM LNLFTFTDLGAPSAFKFFNGDIDRKRYSSTNEIINSTLIYQSPTGLYETRIDLSKLGGK	WP_038409211.1 [Listeria ivanovii]
343	MRKPYTIGLDIGTNSVGWAVLTDQYNLVKRKMKVAGSAEKKQIKKNFWGVRLFDEGEVAAGRRMNRTTRRRLERRRNRIAYLQEIFAAEMAEVDANFFYRLEDSFYIESEKRHSRHPFF ATIEEEVAYHEEYKTIYHLRKKLVNSSDKADLRLVYLALAHIIKYRGNFLIEGMDLDTKNTSVDEVFKQFIQTYNQIFASDIEEGSILTRLEENKEVAEILSEKLTRRKLDKILKLYTGE KSTGMFARFINLIIGSKGDFKKVFDLDEKAEIECAKDTYEEDLEALIAKIGDEYAEIFVAAKSTYNAVVLSNIITVTDTETKAKLSASMIERFDKHAKDLKRLKAFFKMQLPEKFNEVF NDIEKDGYAGYIDGKTTQEKFYKYMKKMLANIDGADYFIDQIEEENFLRKQRTFDNGTIPHQLHLEELEAILHQQAKYYPFLRKDYEKIRSLVTFRIPYFIGPLANGQSDFAWLTRKAD GRIRPWNIEEKVDFGKSAIDFIEKMTNKDTYLPKENVLPKHSLCYQKMVYNELTKIRYIDDQGKTHHFSGQEKQQIFNGLFKQQRKVKKKDLERRFLYTINHIESPTIEGVEDAFNSSF ATYHDLQKGGVTQEILDNPLNADMLEEIVKILTVFEDKRMIKEQLQSFSDVLDGTILKKLERRHYTGWGRLSAKLLTGIRDKHSHLITILDYLMNDDGLNRNLMQLINDSNLSFKSIIEK EQVSTADKGIQSIVAELAGSPAIKKGILQSLKIVDELVGIMGYPPQTIVVEMARENQTTKGKQNNSKPRFISLEKAIKEFGSQILKEHPTDNQCLKNDRLYLYYLQNGKDMYTGKELDI HNLSNYDIDHIIPQSFITDNSIDNRVLVSSTANREKGDNVPLLEVVRKMKAFWEKLYQAKLMSKRKFDYLTKAERGGLTEADKANFIQRQLVETRQITKNVANILQSRFNCKQDENGNE VEQVRIVTLKSTLVSQFRKQFQLYKVREVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGDYHQFDWFKANKATAKKQFYTNIMRFFAKENQIIDKNGEILWDNRYLDTIKKVLSYR QMNIVKKTEIQKGEFSNATVNPKGNSSKLISRKADWNPIKYGGFDGSNMAYSIVIEYEKKKKTVIKKELIQINIMERVAFEKDQKAFLEEKGYYSPKVLTKIPKYTLYECENGRRRML GSANEAQKGNQMVLPNHLMTLLYHAKNCEAANDGESLAYIEMHREMFAELLAYISEFAKRYTLANDRLEKINMFFEQNKKGDIKVIAKSFDKLKVFNAFGAPRDFEFFETTIKRKRYYNI KELLNATIIYQSITGLYETRKRLED	WP_016631044.1 [Enterococcus multispecies]
344	MRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTDLDENFFARLQESFLVPEDKKWHRHPIFAKLEDEVAYHETYPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHF LIEGKLSTENTSVKDQFQQFMVIYNQTFVNGESRLVSAPLPESVLIEEELTEKASRTKKSEKVLQQFPQEKANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYASESYEEDLEGILAK VGDEYSDVFLAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKKFKRFIRENCPDEYDNLFKNEQKDGYAGYIAHAGKVSQLKFYQVVKKIIQDIAGAEYFLEKIAQENF LRKQQRTFDNGVIPHQIHLAELQALIHRQAAYYPFLKENQEKIEKILTFRIPYYVGPLSKGDASTFAWLKRQSEEPIRPWNLQETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYE KFMVFNELTKISYTDDRGIKANFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQFNASFSTYQDLLKCGITRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQLL STFKGQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILDYLVKDDGVSKHYNRNFMQLINDSQLSFKNAIQKAQSSEHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIM GYAPKRIVVEMARENQTTSTGKRNSIQRLKIVEKAMAEIGSNLLKEQPTNEQDLQNDRLFLYYMQNGKDMYTGDELSLHRLSHYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVP SKEVVKDMKAYWEKLYAAGLISQRKFQRLTKGEQGGLTLEDKAHFIQRQLVETRQITKNVAGILDQRYNAKSKEKKVQIITLKASLTSQFRSIFGLDKVREVNDYHHGQDAYLNCVVAI TLLKVYPNLAPEFVYGEYPKFQTFKENKATAKAIIYTNLLRFFTEDEPRFTKDGEILWSNSYLKTIKKELNYHQMNIVKKVEVQKGGFSKESIKPKGPSNKLIPVKNGLDPQKYGGFDS PVVAYTVLFTHEKGKKPLIKQEILGITIMEKTRFEQNPILFLEEKGFLPRPVIMKLPKYTLYECENGRRRLLASAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSRSLAYVEQHQPEF QELIERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFNAMGAPSTFKFFQKDIERARYTSIKEIFDATIIYQSPTGLYETRRKVVD	WP_000066813.1 [Streptococcus sp. M334]
345	MSNKSYSIGLDIGTNSVGWAVITDDYKVPSKKMKVLGNTDKHFIKKNLIGALLFDEGTTAEDRRLKRTARRRYTRRKNRLRYLQEIFSQEISKVDSSFFHRLDDFFLVPEDKRGSKYPI FATLVEEKEYHKKFPTIYHLRKHLADSKEKTDLRLIYLALAHMIKYRGHFLYEESFDIKNNDIQKIFSEFISIYDNTFEGKSLSGQNAQVEAIFTDKISKSTKRERVLKLFPDEKSTGL FSEFLKLIVGNQADFKKHFDLEEKAPLQFSKDTYDEDLENLLGQIGDDFADLFLVAKKLYDAILLSGILTVKDLSTKAPLSASMIERYENHQKDLAALKQFIQNNLQEKYDEVFSDQSK DGYAGYIDGKTTQEAFYKYIKNLLSKFEGADYFLDKIEREDFLKKQRTFDNGSIPHQIHLQEMNAIIRRQGEHYPFLQENKEKIEKILTFRIPYYVGPLARGNGDFAWLTRNSDQAIRP WNFEEIVDQASSAEDFINKMTNYDLYLPEEKVLPKHSLLYETFAVYNELTKVKFVIAEGLTRYQFLDKKQKKDIFDTFFKAENKRKVTEKDIIHYLHNVDGYDGIELKGIEKQFNASLST YHDLLKIIKDKAFMDDSKNEEILENIIHTLTIFEDREMIKQRLAQYDSLFDEKVIKALRTIRHYTGWGKLSAKLINGIRDKKSGKTILDYLIDDGELHNRNFMQLIHDDGLSFKEIIQKAQ VFGKTNDVKQVVELPGSPAIKKGILQSIKIVDELVKVMGHAPESIVIEMARENQTTARGKKNSQQRYKRIEDSLKNLASGLDSNILKENPTDNQLQNDRLFLYYLQNGRDMYTGKPL EINQLSNYDIDHIIPQAFIKDDSLDNRVLTSSKDNRGKSDNVPSLEVVKKMKAFWQQLLDSKLISERKFNNLTKAERERGGLNELDKVGFIKRQLVETRQITKHVAQFLDARFNNKEVTE KDKKNRNVKIITLKSKLVSNFRKEFGLYKVREINDYHHAHDAYLNAVLAKAILKKYPKLESEFVYGDYQKYDLQRVISKSREPKEVEKATQKYFFYSNLINFFKEEVHYADGTIVKREN IEYSKDTGEIAWNKEKDFATVRKVLSLPQVNIVKKTEVQTGGFSKESILPKGNSDKLIPRKTKEILWDTTKYGGFDSPVIAYSILLIADIEKGKAKKLKTVKTLVGITIMEKATFEKNP ITFLENKGYHNVRKKENILICLPKYSLFELESGRRRMLASAKELQKGNEIVLPVYLTTLLYHSKNVHKLDEPEHLEYIQKHRYEFKDLINLVSEFSQKYVLADANLEKIKNLYADNEQADI EILANSFINLLTFTALGAPAAFKFFGKDVDRKRYTVTSEILNATLIHQSITGLYETRIDLSKLGED	WP_031589969.1 [Kandleria vitulina]
346	MSQNNNKIYNIGLDIGDASVGWAVVDEHYNLLKRHGKHMWGSRLFTQANTAVERRSSRSTRRRYNKRRERIRLLREIMEDMVLDVDPTFFIRLANVSFLDQEDKKDYLKENYHSNYNLF IDKDFNDKTYYDKYPTIYHLRKHLCESKEKEDPRLIYLALAHHIVKYRGNFLYEGQKFSMDVSNIEDKMIDVLRQFNEINLFEYVEDRKKIDEVLNVLKEPLSKKHKAEKAFALFDTTKD NKAAYKELCAALAGNKFNVTMKLMKEAELHDEDEKDISFKFSDATFDDAFVEKQPLLGDCVEFIDLLHDIYSWVELMNILGSAHTSEPSISAAMIQRYEDHKNDLKLLKDVIRKYLPKKV	WP_038409211.1 [Listeria ivanovii]

	FEVFRDEKSKKNNYCNYINHPSKTPVDEFYKYIKKLIEKIDDPDVKTILNKIELESFMLKQNSRTNGAVPYQMQLDELNKILENQSVYYSDLKDNEDKIRSILTFRIPYYFGPLNITKD RQFDWIIKKEGKENERILPWNANEIVDVDKTADEFIKRMRNFCTYFPDEPVMAKNSLTVSKYEVLNEINKLRINDHLIKRDMKDFKMLHTLFMDHKSISANAMKKWLVKNQYFSNTDDIK IEGFQKENACSTSLTPWIDFTKIFGKINESNYDFIEKIIYDVTVFEDKKILRRRLKKEYDLDEEKIKKILKLKYSGWSRLSKKLLSGIKTKYKDSTRTPETVLEVMERTNMNLMQVIND EKLGFKKTIDDANSTSVSGKFSYAEVQELAGSPAIKRGIWQALLIVDEIKKIMKHEPAHVYIEFARNEDEKERKDSFVNQMLKLYKDYDFEDETRKEAANKHLKGEDAKSKIRSERLKLY YTQMGKCMYTGKSLDIDRLDTYQVDHIVPQSLLKDDSIDNKVLVLSSENQRKLDDLVIPSASIRNKMYGFWEKLFNNKIISPKKFYSLIKTEFNEKDQERFINRQIVETRQITKHVAQII DNHYENTKVVTVRADLSHQFRERYHIYKNRDINDFHHAHDAYIATILGTYIGHRFESLDAKYIYGEYKRIFRNQKNKGKEMKRNNDGFILNSMRNIYADKDTGEIVWDPNYIDRIKKCF YYKDCFVTKKLENNGTFFNVTVLPNDTNSDKNTLATVPVNKYRSNVNKYGGFSGVNSFIVAIKGKKKGKKVIEVNKLTGTIPLMYKNADEEIKINYILKQAEDLEEVQIGKEILKNQL IEKDGGLYYIVAPTEIINAKQLILNESQTKLVCEIYKAMKYKNYDNLDSEKIIDLYRLLINKMELYYPEYRKQLVKKFEDRYEQLKVISIEEKCNIIKQILATLHCNSSIGKIMYSDFK ISTTIGRLNGRTISLDDISFIAESPTGMYSKKYKL	WP_029073316.1 [Kandleria vitulina]
347	MSQNNNKIYNIGLDIGDASVGWAVVDENYNLIKRHGKHMWGSRLFTQANTAVERRSSRSTRRRYNKRRERIRLLRGIMEDMVLDVDPTFFIRLANVSFLDQEDKKDYLKENYHSNYNLF IDKDFNDKTYYDKYPTIYHLRKHLCESKEKEDPRLIYLALAHHIVKYRGNFLYEGQKFSMDVSNIEDKMIDVLRQFNEINLFEYVEDRKKIDEVLNVLKEPLSKKHKADKAFALFDTTKKD NKAAYKELCAALAGNKFNVTKMLKEAELHDEDEKDISFKFSDATFDDAFVEKQPLLGDCVEFIDLLHDIYSWVELQNILGSAHTSEPSISAAMIQRYEDHKNDLKLLKDVIRKYLPKKY FEVFRDEKSKKNNYCNYINHPSKTPVDEFYKYIKKLIEKIDDPDVKTILNKIELESFMLKQNSRTNGAVPYQMQLDELNKILENQSVYYSDLKDNEDKIRSILTFRIPYYFGPLNITKD RQFDWIIKKEGKENERILPWNANEIVDVDKTADEFIKRMRNFCTYFPDEPVLAKNSLTVSKYEVLNEINKLRINDHLIKRDIKDFKMLHTLFMDHKSISANAMKKWLVKNQYFSNTDDIK IEGFQKENACSTSLTPWIDFTKIFGBINNSNYELIEKIIYDVTVFEDKKILRRRLKKEYDLDEEKIKKILKLKYSGWSRLSKKLLSGIKTKYKDSTRTPETVLEVMERTNMNLMQVIND EKLGFKKTIDDANSTSVSGKFSYAEVQELAGSPAIKKGIWQALLIVDEIKKIMKHEPAHVYIEFARNEDEKERKDSFVNQMLKLYKDYDFEDETRKEAANKHLKGEDAKSKIRSERLKLY YTQMGKCMYTGKSLDIDRLDTYQVDHIVPQSLLKDDSIDNKVLVLSSENQRKLDDLVIPEMIRNKMFGFWNKLYENKIISPKKFYSLIKSEYSDKDKERFINRQIVETRQITKHVAQII SNHYETTKVVTVRADLSHAFRERYHIYKNRDINDFHHAHDAYIATILGTYIGHRFESLDAKYIYGEYQKIFRNNKNKDKEFNNKNDGFILNSMRNLYADKDTGEVVWDPEWISRIKKCF YYKDCFVTKKLENNGSFFNLTVRPNDEHSEKGTTIAKVPVNKLKSNVHKYGGFEGLKYSIVAIKGKKKGKKIIDVNKLVGIPLMYKNVDDETKINYIKESEGLEEVKIKIEILKNQL IEINGGLFYVTSPTEIVNARQLILDFNCTRIIDGIYKAMKYKNYSELSQEEIMNVYDIFVEKLKLYYPTYKNIATNFENMREQFENISDEEKCEVIRQMLVVMHAGPQNGNITFDDFKL SNRLGRLNCKTISLDTTVFIADSPTGMYSKKYKL	WP_039099354.1 [Lactobacillus curvatus]
348	MSRPYNIGLDIGTSSIGWSVVDQSKLVSVRGKYGYGVRLYDEGQTAAERRSFRTTRRRLKRRKWRLGLLRELFEPYITPVDPTFFLRKKQSNLSPKDQRKLYPQTSLFNDRTDRAFYD DYPTIYHLRYKLMTEKRQFDIRELIYLAMHHIVKYRGHFLNEAPVSSFKSSEINLVAHFDRLNTIFADLFSESGFQLETDKLAEVKALLLDNHQSASNRQRQAALLIITPSTNKAVEKQN KAIATELLKALGLKAKFNVLTGIEAEDVKTWTLTFNAENFDEEMVKLESSLDDNAHQIESLQELYSGVLLAGIVPVENQQSLSQAMITKYDDHQKHLKMLKAVREALAPEDRQRLKQAY DQVVDGQENTKAYSKEDFYGDITKALKNNPDHFVSEIKKLLIELDQFMPKQRTKDNGAIPHQLHLQELDRIIENQQQYYPWLAELNPNSKRQTVAKYKLDELVAFRVPYYVGPLITAEQ QGQSSDAKFAWMIRKAEGQITPWNFDDKVDRQASANEFIKRMTTTDTYLLLAEDVLPKQSLLYQRFEVLNELNKLRINGLKITDDQPITTELKQAIFTDLFMQKTSVTVKNIQDYLVSEKRYASRP AITGLSDENKFNSRLSTYHDLKTIVGDAVDDVDKQADLEKCIEWSTIFEDGKIYSAKLNEIDWITDQQRVQLAAKRYRGWGRLSAKLLTQIVNANGQRIMDLLWDTTDNFMRIVHSEDF DKLITEAANQMMLAENDVQDVINDLVTSPQNKKALRQLILLVVNDIQKAMKGQAPERILIEFARNEDEVNPRLSVQRKRQVEQVYQNISNELLNNTEIRNELKDLSNSALSNTRLFLYFMQG GRDMYTGDSLNIDRLSTYDIDHILPQSFIKDNSLDNRVLVSQRMNRSKADQVPTDFTSVELGQKMQIQWEQMLRAGLITKKYDNLTLNPDHISKYAMKGFINRQLVETRQVIKLATNL LMEQYGEDNIELITVKSGLTHQMRTEFDFPKNRNLNNHHHAFDAYLTAFVGLYLLKRYPKLKPYFVYGEYQKASQQDKWRNFNFLNGLKKDELVDENTEAVIWNKESGLAYLNKIYQFK KIIVTREVHENSGALFNQTLYAAKDDKASGQGGKQLIPAKQDRPTALYGGYSGKTVAYMCIVRIKNKKGDLYKVCGVETSWLAQLKQLTDEDSKKAFLKQKISPQFTKVKKQKGTIVKV VEDFEVIAPHILINQRFFDNGQELITLGSATYKHNEQGLILLDKTAVKLINGALPLTQSEELAEQVYDEILDQVMHYFPLYDTNQFRAKLSAGKAAFMKLPWKSQWDGNKMVQGQQVILD RVLIGLHANAAVSDLGVLKISTPLGKMQQPSGISLSPDTQIIYQSPTGLFERRVAIRDL	WP_012962174.1 [Streptococcus gallolyticus]
349	MTEKNYSIGLDIGTNSVGWAVITDDYKVPAKKMKVLGNTDKKYIKKNLIGALLFDNGETAEATRLKRTARRRYTRRKNRLRYLQEIFAEEMAKVDESFFYRLDESFLTTDDKDFERHPI FGNKADEIKYHQEFPTIYHLRKHLADSHEKADLRLIYLALAHMIKFRGHFLIEGELNAENTDVQKLFEAFVEVYDRTFDDSNLSEITVDASSILTEKFSKSRRLENLIKHYPTEKKNTL FGNIVALALGLTPNFKTSFKLSEDAKLQFSKDTYEEDLEELIGKIGDEYADLFTSAKNLYDAILLSGILTVADNTTKAPLSASMIKRYNEHQVDLKKLKEFIKNNASDKYDEIFNDKDK NGYAGYIENGVKQDEFYKYLKTTLSKIDGSDVFYLDKIEREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYAFLKENQAKIEKILTFRIPYYVGPLARKNSRFAWAEYHSDEKITP WNFDEIIDKEKSAEKFITRMTLNDLYLPEEKVFLPKHSLVYEYFTVYNELTKVKYVNEQGKSNFFDANMKQEIFEHVFKENRKVTKDKFLNYLNKEPPEYRIQDLIGLDKENKSFNASLIG TYHDLKILDKSFLDDKTNETIIEDIIQTLTLFEDRDMIRQRLQKYSDIFTPQQLKKLERRHYTGWGRLSYKLINGIRNKENGKSILDYLIDDGYANRNFMQLISDDTLPFKQIIKDAQ IIGDIDDVTSVVRELPGSPAIKKGILQSVKIVDELVKVMGHNPDNIVIEMARENQKTHRTKPRLKAIENGLKQIGSTLLKEQPTDNKALQKERIYLYYLQNGRDMYTGEPLEI DELDIDHLSDYDIDHIIPQAFIKDDSIDNRVLTSSAANRGKSDDVPSDLIVHDRKADWIRLYKSGLISKRKFDNLTKAERGGLTENDKAGFIKRQLVETRQITKHVAQILDSRFNTERD ENDKVIRNVKVITLKSNLVSQFRKDFKFYKVREINDYHHAHDAYLNAVVGTALLKKYPKLAPEFVYGEYKKYDIRKFITSSGDKATAKYFFYSNIMNFFKRVIRYSNGKVVVRPVIECS KDTGEIAWNKQTDFEKVRKVLSYPQVNIVKKVETQTGGFSKESILPKRKTKKFRWDTPKYGGFDSPNIAYSVFVIADVEKGKAKKLKTVKELVGISIMERSSFEENPVVFL EKKGYQNVQEDNLIKLPKYSLFEFEGGRRRLLASASELQKGNEVVLSRHLVELLYHAHRVNSFNNSEHLKYVSEHKKEFGEVLSCVENFAKSYVDVEKNLGKIRAVADKIDTFSIEDIS ISFVNLLTLTALGAPADFNFLGEKIPRKRYTSTKECLNATLIHQSITGLYETRIDLSKLGEE	WP_034700478.1 [Enterococcus hirae]
350	MTKDYTIGLDIGTNSVGWAVLTDDYQLMKRKMSVHGNTEKKKIKKNFWGARLFDEGQTAEERRTKRTNRRRLARRKYRLSKLQDLFAEELCKQDDCFFVRLEESFLVPEEKQYKPASIF PTLEEEKEYYQKYPTIYHLRQKLVDSTEKEDLRLVYLALAHLLKYRGHFLFEGDLDTENTSIEESFRVFLEQYGKQSDQPLIVHQPVLTILTDKLSKTKKVEEILKYYPTEKINSFFAQ CLKLIVGNQANFKRIFDLEAEVKLQFSKETYEEDLESLLEKIGDEYLDIFLQAKKVHDAILLSEIISSTVKFTQAKLSSGMVERYERHKADLAKFKQFVKENVPQRATVFFKDTTKNGY AGYIKGKTTQEEFYKFVKKELSGVVGSEPFLEKIDQETFLLKQRTYTNGVIPHQVHLIELKAIIDQQKQHYPFLEEAGPKIIALFKFRIPYYVGPLAKEQEASSFAWIERKTAEKINPW NFSEVVDIEKSAMRFIQRMTKQDTYLPTEKVLPKNSLLYQKYMIFNELTKVSYKDERGVKQYFSGDEKQQIFKQLFQKERGKTIVKKLQNFLYTHYHIENAQIFGIEKAFNASYSTYHD DATIIYQSVTGLYETRIRMGDLWAGEQ	
351	MTKDYTIGLDIGTNSVGWAVLTDDYQLMKRKMSVHGNTEKKKIKKNFWGARLFDEGQTAEFRRTKRTNRRRLARRKYRLSKLQDLFAEELCKQDDCFFVRLEESFLVPEEKQYKPASIF PTLEEEKEYYQKYPTIYHLRQKLVDSTEKEDLRLVYLALAHLLKYRGHFLFEGYSKQSDQPLIVHQPVLTILTDKLSKTKKVEEILKYYPTEKINSFFAQ CLKLIVGNQANFKRIFDLEAEVKLQFSKETYEEDLESLLEKIGDEYLDIFLQAKKVHDAILLSEIISSTVKFTKAKLSSGMVERYERHKADLAKFKQFVKENVPQRATVFFKDTTKNGY AGYIKGKTTQEEFYKFVKKELSGVVGSEPFLEKIDQETFLLKQRTYTNGVIPHQVHLIELKAIIDQQKQHYPFLEEAGPKIIALFKFRIPYYVGPLAKEQEASSFAWIERKTAEKINPW NFSEVVDIEKSAMRFIQRMTKQDTYLPTEKVLPKNSLLYQKYMIFNELTKVSYKDERGVKQYFSGDEKQQIFKQLFQKERGKTIVKKLQNFLYTHYHIENAQIFGIEKAFNASYSTYHD	WP_034867970.1 [Enterococcus faecium]

	FMKLAKTNQKAMQEWLEQPEMEPIFEDIVKILTIFEDRQMIKHQLSKYQEVFGEKLLKEFARKHYTGWGRFSAKLIHGIRDRKTNKTILDYLINDDDVPANRNRNLMQLINDEHLSFKE EIAKATVFSKHKSLVDVIQDLPGSPAIKKGIWQSLKIVEELIAIIGYKPKNIVIEMARENQKTHRTSPRLKALENGLKQIGSTLLKEQPTDNKALQKERLYLYYLQNGRDMYTGEPLEI ENLHQYEVDHIIPRSFIVDNSIDDKVLVASKQNQKRDDVPKKQIVNEQRIFWNQLKEAKLISTKKYAYLTKIELTPEDKARFIQRQLVETRQITKHVANILHQSFNQEEEGTDCDGVQ IITLKATLTSQFRQTFGLYKVRENNPHHHAHDAYLNGFIANVLLKRYPKLAPEFVYGKYKYSLARENKATAKKEFYSNILKFLESDEPFCDENGEIYWEKSHHLPRIKKVLSSHQVNV VKKVEQQKGGFYKETVNSKEKPDKLIERKNNWDVTKYGGFGSPVIAYAIAFVYAKGKTQKKTRAIEGITIMEQAAFEKDPTTFLKEKGFPQVTEFIKLPKYTLFEFDNGRRRFLASHKE SQKGNPFILSDQLVTLLIYHAQHYDKITYQESFDYVNTHLSDFSAILTEVLAFAEKYTLADKNIERIQELYEENKYGETSMIAQSFLQLLQFNAIGAPADFKFFGVTIPRKRYTSLTEIW DATIIYQSVTGLYETRIRMGDLWAGEQ	
352	MTKDYTIGLDIGTNSVGWAVLTDDYQILMKRKMSVHGNTEKKKIKKNFWGARLFDEGQTAEPRRTKRTNRRLARRKYRLSKLQDLFAEELCKQDDCFFVRLEESFLVPEEKQYKPASIF PTLEEEKEYYQKYPTIYHLRQKLVDSTEKEDLRLVYLALAHLLKYRGHFLFEGDLDTENTSIEESFRVFLEQYSKQSDQPLIVHQPVLITLTDKLSKTKKVEEILKYYPTEKINSFFAQ CLKLIVGNQANFKRIFDLEAEVKLQFSKETYEEDLESLLEKIGDEYLDIFLQAKKVHDAILLSEIISSTVKETKAKLSSGMVERYERHKADLAKFKQFVKENVPQRATVFFKDTTKNGY AGYIKGKTTQEEFYKFVKKELSGVVGSEPFLEKIDQETFLLKQRTYTNGVIPHQVHLIELKAIIDQQKQHYPFLEEAGPKIIALFKFRIPYYVGPLAKEQEASSFAWIERKTAEKINPW NFSEVVDIEKSAMRFIQRMTKQDTYLPTEKVLPKNSLLYQKYMIFNELTKVSYKDERGVKQYFSGDEKQQIFKQLFQKERGKTIVKKLQNFLYTHYHIENAQIFGIEKAFNASYSTYHD FMKLAKTNQKAMQEWLEQPEMEPIFEDIVKILTIFEDRQMIKHQLSKYQEVFGEKLLKEFARKHYTGWGRFSAKLIHGIRDRKTNKTILDYLINDDDVPANRNRNLMQLINDEHLSFKE EIAKATVFSKHKSLVDVIQDLPGSPAIKKGIWQSLKIVEELIAIIGYKPKNIVIEMARENQKTHRTSPRLKALENGLKQIGSTLLKEQPTDNKALQKERLYLYYLQNGRDMYTGEPLEI ENLHQYEVDHIIPRSFIVDNSIDNKVLVASKQNQKRDDVPKKQIVNEQRIFWNQLKEAKLISPKKYAYLTKIELTPEDKARFIQRQLVETRQITKHVANILHQSFNQEEEGTDCDGVQ IITLKATLTSQFRQTFGLYKVRENNPHHHAHDAYLNGFIANVLLKRYPKLAPEFVYGKYKYSLARENKATAKKEFYSNILKFLESDEPFCDENGEIYWEKSHHLPRIKKVLSSHQVNV VKKVEQQKGGFYKETVNSKEKPDKLIERKNNWDVTKYGGFGSPVIAYAIAFVYAKGKTQKKTRAIEGITIMEQAAFEKDPTTFLKEKGFPQVTEFIKLPKYTLFEFDNGRRRFLASHKE SQKGNPFILSDQLVTLLIYHAQHYDKITYQESFDYVNTHLSDFSAILTEVLAFAEKYTLADKNIERIQELYEENKYGETSMIAQSFLQLLQFNAIGAPADFKFFGVTIPRKRYTSLTEIW DATIIYQSVTGLYETRIRMGDLWAGEQ	WP_010737004.1 [Enterococcus hirae]
353	MTKDYTIGLDIGTNSVGWAVLTDDYQILMKRKMSVHGNTEKKKIKKNFWGARLFDEGQTAEPRRTKRTNRRLARRKYRLSKLQDLFAEELCKQDDCFFVRLEESFLVPEEKQYKPASIF PTLEEEKEYYQKYPTIYHLRQKLVDSTEKGDLRLVYLALAHLLKYRGHFLFEGDLDTENTSIEESFRVFLEQYGKQSDQPLIVHQPVLITLTDKLSKTKKVEEILKYYPTEKINSFFAQ CLKLIVGNQANFKRIFDLEAEVKLQFSKETYEEDLESLLEKIGDEYLDIFLQAKKVHDAILLSGIMTERYERHKADLAKFKQFVKENVPQRATVFFKDTTKNGY AGYIKGKTTQEEFYKFVKKELSGVVGSEPFLEKIDQETFLLKQRTYTNGVIPHQVHLIELKAIIDQQKQHYPFLEEAGPKIIALFKFRIPYYVGPLAKEQEASSFAWIERKTAEKINPW NFSEVVDIEKSAMRFIQRMTKQDTYLPTEKVLPKNSLFYQKYMIFNELTKVSYKDERGVKQYFSGDEKQQIFKQLFQKERGKTIVKKLQNFLYTHYHIENAQIFGIEKAFNASYSTYHD FMKLAKTNQKAMQEWLEQPEMEPIFEDIVKILTIFEDRQMIKHQLSKYQEVFGEKLLKEFARKHYTGWGRFSAKLIHGIRDRKTNKTILDYLINDDDVPANRNRNLMQLINDEHLSFKE EIAKATVFSKHKSLVDVIQDLPGSPAIKKGIWQSLKIVEELIAIIGYKPKNIVIEMARENQKTHRTSPRLKALENGLKQIGSTLLKEQPTDNKALQKERLYLYYLQNGRDMYTGEPLEI ENLHQYEVDHIIPRSFIVDNSIDNKVLVASKQNQKRDDVPKKQIVNEQRIFWNQLKEAKLISPKKYAYLTKIELTPEDKARFIQRQLVETRQITKHVANILHQSFNQEEEGTDCDGVQ IITLKATLTSQFRQTFGLYKVRENNPHHHAHDAYLNGFIANVLLKRYPKLAPEFVYGKYKYSLARENKATAKKEFYSNILKFLESDEPFCDENGEIYWEKSHHLPRIKKVLSSHQVNV VKKVEQQKGGFYKETVNSKEKPDKLIERKNNWDVTKYGGFGSPVIAYAIAFVYAKGKTQKKTRAIEGITIMEQAAFEKDPTTFLKDKGFPQVTEFIKLPKYTLFEFDNGRRRFLASHKE SQKGNPFILSDQLVTLLIYHAQHYDKITYQESFDYVNTHLSDFSAILTEVLAFAEKYTLADKNIERIQELYEENKYGETSMIAQSFLQLLQFNAIGAPADFKFFGVTIPRKRYTSLTEIW DATIIYQSVTGLYETRIRMGDLWAGEQ	WP_010720994.1 [Enterococcus hirae]
354	MTKKNYSIGLDIGTNSVGWAVITDDYKVPAKKMKVLGNTDKKYIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFANEIAKVDESFFQRLDESFLTDDDKTFDSHPI FGNKAEEDAYHQKFPTIYHLRKHLADSSEKADLRLVYLALAHMIKFRGHFLIEGELNAENTDVQKIFADFVGVYNRTFDDSHLSEITVDVASILTEKISKSRRLENLIKYYPTEKKNTL FGNLIALALGLQPNFKTNFKLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTSAKNLYDAILLSGILTVDDNSTKAPLSASMIKRYVEHHEDLEKLKEFIKANKSELYHDIFKDK NGYAGYIENGVKQDEFYKYLKNILSKIKPDGSDYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGDYYPFLKENQEEIEKILTFRIPYYVGPLARGNSRFAWAEYRSDEKIP TPWNFDKVIDKEKSAEKFITRMTLNDLYLPEEKVLPKHSHVYETYAVYNELTKIKYVNEQGKESFFDSNMKQEIFDHVFKENRKVTKEKLLNYLNKEFPEYRIKDLIGLDKENKSFNAS LGTYHDLKKILDKAFLDDKVNEEVIEDIIKTLTLFEDKDMIHERLQKYSDIFTANQLKKLERRHYTGWGRLSYKLINGIRNKENNKTILDYLIDDGSANRNFMQLINDDTLPFKQIIQK SQVVGDVDDIEAVVHDLPGSPAIKKGILQSVKIVDELVKVMGGNPDNIVIEMARENQTTNRGRSQSQQRLKKLQNSLKELGSNILNEEKPSYIEDKVENSHLQNDQLFLYYIQNGKDMY TGDELDIDHIIPQAFIKDDSIDNRVLTSSAKNRGKSDDVPSLDIVRARKAEWVRLYKSGLISKRKFDNLTKAERGGLTEADKAGFIKRQLVETRQITKHVAQILDARFNTE HDENDKVIILKSNLVSQFRKEFKFYKVRENDYHHAHDAYLNAVVGTALIKKYPKLAPEFVYGEYKKYDVRKLVAKSSDDHSEMGKATAKYFFYSNIMNFFKRVIRYSNGKVIVR PVVEYSKDTGEIAWNKRTDFEKVRKVLSYPQVNIVKKVETQTGGFSKESILPKGDSDKLIPRKTKKVLWEPKKYGGFDSPTVAYSVLVVADVEKGKTKKLKTVKELVGISIMERSFFEK NPVEFLKNKGYQNVQEDKLMKLPKYSLFEFEGGRRRLLASATELQKGNEIMLSAHLVALLYHAHRIGNFNSAEHLKYVSEHKKEFEEVLSCVENFANVYVDVEKNLSKIRAAADSMDNF SIEEISDSFINLLTLTALGAPADFNFLGEKIPRKRVNSTKECLNATLIHQSITGLYETRIDLSKLGEE	
355	MTKKNYSIGLDIGTNSVGWAVITDDYKVPAKKMKVLGNTDKKYIKKNLLGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQDIFTEEMAKVDDSFFQRLDESFLTDNDKNFDSHPI FGNKAEEDAYHQKFPTIYHLRKHLADSTEKADLRLVYLALAHMIKFRGHFLIEGELNAENTDVQKLFTDFVGVYDRTFDDSHLSEITVDAASIILTEKISKSRRLENLINNYPKEKKNTL FGNLIALALGLQPNFKTNFKLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTSSKNLYDAILLSASMIKRYVEHHEDLEKLKEFIKANKSELYHDIFKDKTQ NGYAGYIENGVKQDEFYKYLKGILIKINGSDYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGDYYPFLKENQEEIEKILTFRIPYYVGPLARGDSRFAWAEYRSDEKITP WNFDKVIDKEKSAEKFITRMTLNDLYLPEEKVLPKHSHVYETFAVYNELTKVKYVNEQGKDSFFDSNMKQEIFDHVFKENRKVTKEKLLNYLDKEFPEYRIQDLVGLDKENKSFNASLG TYHDLKKILDKSFLDDKVNEEVIEDIIKTLTLFEDREMIQQRLQKYSDIFTKQQLKKLERRHYTGWGRLSYKLINGIRNKENNKTILDYLIDDGSANRNFMQLINDDALSFKQIIQEAQ VVGDVDDIEVVHDLPGSPAIKKGILQSVKIVDELIKVMGDNPDNIVIEMARENQTTNRGRSQSQQRLKKLQNSLKELGSNILNEEKPSYIEDKVENSHLQNDQLFLYYIQNGKDMYTG DELDIDHLSDYDIDHIIPQAFIKDDSIDNRVLTSSAKNRGKSDDVPSLGIVRARKAEWVRLYKSGLISKRKFDNLTKAERGGLTEADKAGFIKRQLVETRQITKHVAQILDARFNTERD ENDKVIILKSNLVSQFRKEFKFYKVRENDYHHADAYLNAVVGTALIKKYPKLAPEFVYGEYKKYDVRKLVAKSSDDHSEMGKATAKYFFYSNIMNFFKRVIRYSNGKVIVR PVVEYSKDTGEIAWNKRTDFEKVRKVLSYPQVNIVKKVETQTGGFSKESILPKRSDKLIPRKTKKVLWEPKKYGGFDSPTVAYSVLVVADVEKGKTKKLKTVKELVGISIMERSFFEK NPVEFLKNKGYQNVQEDKLMKLPKYSLFEFEGGRRRLLASATELQKGNEIMLSAHLVALLYHAHRIGNFNSAEHLKYVSEHKKEFEEVLSCVENFANVYVDVEKNLSKIRAAADSMDNF SIEEISDSFINLLTLTALGAPADFNFLGEKIPRKRVNSTKECLNATLIHQSITGLYETRIDLSKLGEE	
356	MTKKNYSIGLDIGTNSVGWAVITDDYKVPAKKMKVLGNTDKKYIKKNLLGALLFDSGETAEATRLKRTARRYTRRKNRLRYLQEIFAEEMTKVDESFFQRLDESFTRWDDDNKKLGRY PIFGNKADVVKYHQEFPTIYHLRKHLADSSEKADLRLVYLALAHMIKFRGHFLIEGELNAENTDVQKIFADFVGVYDRTFDDSHLSEITVDAASILTEKISKSRRLENLIKYYPTEKKN TLFGNLIALALGLQPNFKMNFKLSEDAKLQFSKDSYEEDLGELLGKIGDDYADLFTSAKNLYDAILLSGILLVDDNSTKAPLSASMIKRYDEHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFDKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIILDKEFLDDSNEEVIEDILLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSAKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVVTLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEMVLPGYLVELLYHAHRIDNGPEQDMTPAEQSIIVQWCKRKSDFKGKPAPDAPKKAGQAKKKK	

	TGDELDIDHLSDYDIDHIIPQAFIKDDSIDNRVLTSSAKNRGKSDDVPSLDIVRARKAEWVRLYKSGLISKRKFDNLTKAERGGLTEADKAGFIKRQLVETRQITKHVAQILDARFNTE SDENDKVIRDVKVITLKSNLVSQFRKDFEFYKVREINDYHHAHDAYLNAVVGTALLKKYPKLASEFVYGEYKKYDIRKFITNSSDKATAKYFFYSNIMNFFKRVIRYSNGKVIVRPVVE YSKDTEDIAWDKKSNFRTICKVLSYPQVNIVKKVETQTGGFSKESILPKGDSDKLIPRKTKKAYWDTKKYGGFDSPTVAYSVFVVADVEKGKAKKLKTVKELVGISIMERSFFEENPVE FLENKGYHNIREDKLIKLPKYSLFEFEGGKRRLLASASELQKGNEMVIPGHLVKLLYHAQRINSFNSTKYLDYVSAHKKEFEKVLSCVEDFANLYVDVEKNLSKIRAVADSMDNFSIEE LSNSFINLLTLTALGAPADFNFLGEKIPRKRYTSTKECLNATLIHQSITGLYETRIDLSKLGEE	WP_009854540.1 [Streptococcus gallolyticus]
357	MTKKNYSIGLDIGTNSVGWAVITDDYKVPAKKMKVLGNTDKKYIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFAEEMTKVDESFFYRLDESFLTTDEKDFERHPI FGNKAEEDAYHQKFPTIYHLRNYLADSSEKADLRLVYLALAHMIKYRGHFLIEGKLNAENTDVQKLFTDFVGVYDRTFDDSHLSEITVDVASTLTEKISKSRRLENLIKYYPTEKKNTL FGNLIALALGLQPNFKMNFKLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTSAKNLVDAILLSGILTVDDNSTKAPLSASMIKRYVEHHEDLEKLKEFIKANKSELYHDIFKDKNK NGYAGYIENGVKQDEFYKYLKNTLSKIDGSDYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGDYYPFLKEKQDRIEKILTFRIPYYVGPLVRKDSRFAWAEYRSDEKITP WNFDKVIDKEKSAEKFITRMTLNDLYLPEEKVLPKHSLLYEIFTVYNELTKVKYVNEQGKESFFDSNMKQEIFDHVFKENRKVTKEKLLNYLNKEFPEYRIKDLIGLDKENKSFNASLG TYHDLKKILDKAFLDDKVNEEVIEDIIKTLTLFEDKDMIHERLQKYSDIFTANQLKKLFERRHYTGWGRLSYKLINGIRNKENNKTILDYLIDDGSANRNFMQLINDDTLPFKQIIQKQS VVGDVDDIEAVVHDLPGSPAIKKGILQSVKIVDELVKVMGDNPDNIVIEMARENQTTNQRGKSNSRMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYTGEELDIHHLS DELDIDHLSDYDIDHIIPQAFIKDDSIDNRVLTSSAKNRGKSDDVPSLDIVRARKAEWVRLYKSGLISKRKFDNLTKAERGGLTEADKAGFIKRQLVETRQITKHVAQILDARFNTEHD ENDKVIRDVKVITLKSNLVSQFRKDFEFYKVREINDYHHAHDAYLNAVVGTALLKKYPKLASEFVYGEYKKYDIRKFITNSSDKATAKYFFYSNIMNFFKTKVKYADGTVFERPIIETN ADGEIAWNKQIDFEKVRKVLSYPQVNIVKKVETQTGGFSKESILPKGDSDKLIPRKTKKYYWDTKKYGGFDSPTVAYSVFVVADVEKGKAKKLKTVKELVGISIMERSFFEENPVEFLE NKGYHNIREDKLIKLPKYSLFEFEGGRRRLLASASELQKGNEMVLPGYLVELLYHAHRADNFNSTEYLINYVSEHKKEFEKVLSCVEDFANLYVDVEKNLSKIRAVADSMDNFSIEEISN SFINLLTLTALGAPADFNFLGEKIPRKRYTSTKECLTATLIHQSITGLYETRIDLSKLGEE	WP_039695303.1 [Streptococcus gallolyticus]
358	MTKKNYSIGLDIGTNSVGWAVITDDYKVPAKKMKVLGNTDKKYIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRLRYLQEIFAEIAKVDESFFQRLDESFLTDDDKTFDSHPI FGNLNAENTDVQKLFTDFVGVYNRTFDDSHLSEITVDVASILTEKISKSRRLENLIKYYPTEKKNTL FGNLIALALGLQPNFKTNFKLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTSAKNLVDAILLSGILTVDDNSTKAPLSASMIKRYVEHHEDLEKLKEFIKANKSELYHDIFKDKNK NGYAGYIENGVKQDEFYKYLKNILSKIKIDGSDYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGDYYPFLKEKQDRIEKILTFRIPYYVGPLVRKDSRFAWAEYRSDEKI TPWNFDKVIDKEKSAEKFITRMTLNDLYLPEEKVLPKHSHVYETYAVYNELTKIKYVNEQGKESFFDSNMKQEIFDHVFKENRKVTKEKLLNYLNKEFPEYRIKDLIGLDKENKSFNAS LGTYHDLKKILDKAFLDDKVNEEVIEDIIKTLTLFEDKDMIHERLQKYSDIFTANQLKKLFERRHYTGWGRLSYKLINGIRNKENNKTILDYLIDDGSANRNFMQLINDDTLPFKQIIQK SQVVGDVDDIEAVVHDLPGSPAIKKGILQSVKIVDELVKVMGGNPDNVIIEMARENQTTNRGRSQSQQRLKKLQNSLKELGSNILNEEKPSYIEDKVENSHLQNDQLFLYYIQNGKDMY TGDELDIDHLSDYDIDHIIPQAFIKDDSIDNRVLTSSAKNRGKSDDVPSLDIVRARKAEWVRLYKSGLISKRKFDNLTKAERGGLTEADKAGFIKRQLVETRQITKHVAQILDARFNTE HDENDKVIRDVKVITLKSNLVSQFRKDFEFYKVREINDYHHAHDAYLNAVVGTALLKKYPKLASEFVYGEYKKYDIRKFITNSSDKATAKYFFYSNIMNFFKTKVKYADGTVFERPIIE TNADGEIAWNKQIDFEKVRKVLSYPQVNIVKKVETQTGGFSKESILPKGDSDKLIPRKTKKVYWDTKKYGGFDSPTVAYSVFVVADVEKGKAKKLKTVKELVGISIMERSFFEENPVEF LENKGYHNIREDKLIKLPKYSLFEFEGGRRRLLASASELQKGNEMVLPGYLVELLYHAHRADNFNSTEYLINYVSEHKKEFEKVLSCVEDFANLYVDVEKNLSKIRAVADSMDNFSIEEI SNSFINLLTLTALGAPADFNFLGEKIPRKRYTSTKECLNATLIHQSITGLYETRIDLSKLGEE	WP_048800889.1 [Streptococcus constellatus]
359	MTQKYSIGLDIGTNSVGWAIVTDDYKVPAKKMKILGNTNKQYIKKNLIGALLFDSGETAKATRLKRTARRYTRRKNRLRYLQEIFIEEMNKVDENFFQRLDDSFLVTEDKRGSKYPIF GTLKEEKEYYKEFETIYHLRKRLADSTGKVDLRLVYLALAHMIKFRGHFLIEGQLKAENTDVQTLFNDFVEVYDKTIEESHLAEITVDALSILTEKVSKSRRLENLVKCYPTEKKNTLF GNLIALSLGLQPNFKTNFQLSEDAKLQFSKDTYEEDLEELLGKIGDDYADLFTSAKNLYDTILLSGILAVDDNSTKALLSASMIKRYEEHQKDLKKLKDFIKVNAPAQYDDIFKDETKN GYAGYIENGVKQDEFYKYLKNTLSKIDGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLVRKGSRFAWAEYKADEKITPW NFDDIILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSLLYEIFTVYNELTKVKYVNEQGEAKFFDANMKQEIFDHVFKENPKVTKDKLINYLDKEFDEFRIVDLTGLDKENKAFNASLGT YHDLRKIILDKSFLDDKANEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENNKTILEYLVDDGYANRNFMQLINDDTLPFKQIIKDAQA IDDVDDIELIVHDLPGSPAIKKGILQSIKIVDELVKVMGYNPDNIVIEMARENQTTTKGRRNSQQRLKKLLQDSITNLDNPVSIKNVENQQLQNDRLFLYYIQNGKDMYTGEELDIHHLS DYDIDHIIPQAFIKDDSIDNRVLTSSAKNRGKSDNVPNLEVVCDRKAGWIRLREAGLISQRKFDNLTKAERGGLTENDKAGFIHRQLVETRQITKHVAQILDARFNPKFRDDNKKVIRDV KIITLKSNLVSQFRRDFKLYKVREINDYHHAHDAYLNAVVGTALLKKYPKLTSEFVYGEYKKYDVRKFIAKSDNDEIGKATAKYFFYSNLMNFFKTEVTKFADGTVVERPDIIETSEDGEI AWNKQTDFKIVRKVLSYPQVNIVKKVEKQTGRFSKESILPKGDSDKLIARKTKENYWDTKKYGGFDSPTVAYSVLVVADIKKGKAKKLKTVKELVGISIMERPFFEKNPIMFLESKGYR NIQKDKLIKLPKYSLFELENGRRRLLASAVELQKGNEMVLPQYLNNLLYHAHRIDNSDNSEHLKYITEHKEEFGKLLSYIENFAKSYVDVDKNLEKIQLAVEKIDSFSVKEISNSFIHL LTLTALGAPADFNFLGEKIPRKRYTSTKECLNATLIHQSITGLYETQTDLSKLGED	WP_054279288.1 [Streptococcus phocae]
360	MVEKKSYSIGLDIGTNSVGWAVITDDYKVPAKKMKVLGNTSRQSIKKNMIGALLFDEGGPAASTRVKRTTRRRYTRRKNRLCYLRDIFSEMMHTIDKHFFLRLEDSFLHKSDKRYEAHP IFGTLVQEEKAYHDNYPTIYHLRKALADNTEKADLRLIYLALAHIIKFRGHFLIEGALSANNTDVQQLVHALVDAYNIMFEEDGLDIEAIDVKAILTEKISKTRRLENLISNIPGQKKNG LFGNLIALSLGLIPNFKSNFNLPEDAKLQLAKDTYDEEINNLLTQIGDEYADLFLSAKNLSDAILLSDILTVNGDGTQAPLSASLIKRYEEHRQDLALLKQMFKEQLPDLYRDVFTDEN KDGYAGYISGKTSQEAFYKYIKPILETLDGAEDFLTKINREDFLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNQEKIEKILTFRIPYYIGPLARGNSRFAWLTRTSDQKIT PWNFDEMVDQEASAQAFIERMTNFDEYLPQEKVLPKHSLLYEYFTVYNELTKVKYVTEGMTKPEFLSAGQKEQKIVELLFKKYRKVTVKQLKEDFFSKIECFDTVDISGVEDKFNASLGTI YHDLLKIIKDKAFLDNSENENIIEDIIITLTLFEDREMIANRLAVYEDLFDQNVLKQLKRRHYTGWGRLSKQLINGMRDKHTGKTILDFLKADGFINRNFMQLINDDNLSFKEEIKKAQ EGGLKDSINDQIRDLAGSPAIKKGILQTINIVDEIVKIMGKAPQHIVVEMARDVQKTDIGVKQSRERMKRVQEVLKKLGSQLLKEHPVENFQLQNERLYLYYLQNGKDMYTGEELSISN LSHYDIDHIIPRSFIKDDSIDNKVLTRSEHNRGKTDNVPSIEVVKRMKPYWQKLLDTKVISQRKFDNLTKAERGGLQESDKANFIQRQLVETRQITKHVAQILDSRFNTERDEKDRPIR RVKVITLKSKFVSDFRQDFGFYKLREINDYHHAHDAYLNAVVGTALLKMYPKLASEFVYGDYQKYDLKRMVGKSGKASGHATAKYFFYSNIMNFFKSEVKLANGNIIKRSPIEVNEETG EIVWDKTKDFGTVRKVLSAPQVNIVKKTEIQTGGFSNETILSKGKSSKLIPRKNKWRDTKYGGFNTPTVAYSVLVVAKVEKGKAKKLKPVKELVGITIMERTKFEEANPIAFLESKGYH DIQEHIMITLPKYSLFELENGRRRLLASATELQKGNEMVLPQSVTLGLYEIRIDLSQFGGD	WP_054279288.1 [Streptococcus phocae]
361	MWGVSLFEAGKTAAERRGYRSTRRRLNHRKFRLRLLEDMFEKEILSKDPSFFIRLKEAFLSPKDEQKQFKGNILFNDKDYTDADYYEQYKTIYHLRYDLISQHRQFDIREVYLAIHHLI KYRGHFIYEDQTFTTDGNQLQHHIKAIITMINSTLNRIIIPETDIDVFEKILLDRMMNRSSKVKFLFELITGEDKQDKPLLKELFNLIVGLKAKPASIFEQENLATIVETMNMSTEQVQ IDLITADVIADEEYDILITAQKIYSAIIDLSEMDGYEYFAEAKKESYRKHQEELIVLVKKMLKSNAITNDERAKFEYFYTDYTGAKSSNVSAAYGKYSKEERLFKHIEL LLAKENVLTTVEHALLEKNITFASLILPLQRSSRNAVIPYQVHEKELVAILENQATYYPFLLEQKDNIHKLITFRIPYYVGPLADQKDSEFAWMVRKQAGKITPFNFEEMVDIDASSEAF IKRMTNKCTYLIHEDVIPKHSFSYAKFEVLNELNKIRLDGKPIDIPILKKRIFEGLFLEKTKVTQTSLKKWLAREHEHMTVSVVQGTQKETEFATSIQAFHRFVKIFDPRETVSNPANEEMF EKIIYWSTVFEDKKIMRKLSEYPQLTEQQVQLAQVREFRGWGRLSQKLINRIKTPVSGDEDHKLSINEIIWQINENFMQIIRNKDYLFKKIIEEQFENETDKISFIQLVESLRKRKFEFIKKNRNVNDYH KGIWQAIKIVKELEKVLQQPAENIFIEFARSDEESKRSTPRDKFIESKAYAQLKKETDTFNLEHLKELKQRSKQLSSQRLFLYFIQNGKCMYSGEHLDIERLDSYEVDHILPQSYIKDNS IENLALVKKVENQRKKDSLLLNSSIINQNYSRWEQLKNAGLIGEKKFRNLTRTKITLDRDKEGFIARQLVETRQITKHVAQILDSRFNTEFDKNDKPIRTVKIITLKSTLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDNIIIKLPKYTLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	WP_029090905.1 [Brochothrix thermosphacta]

	HLDTAIYGGYTAVNYASYALIQFKKGRKLKREIIGIPLAVQTRIDNSETSLQAYIAEQIKSEVEIINGRILKYQLISNNGNRLYIAGPSERHNARQLIVSDEAAKVIWLISTKQADEAM FLKYYRLEHLEAVFEELIRKQAADYQIFEKLIKKIEVNKVYFYSCTYNEKVKVIEELLKITQANATNGDLKLLKMSNREGRLGSVSVALDFKIINQSVTGLYQSIEDYNN	
362	PDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAK NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ IHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNA KLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF PKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRPAATKKAGQAKKKK GGD	
363	MDYKDDDDKDYKDDDDKMAPKKKRKVGIHGVPAADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASG VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKKR YDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVK QLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILD FLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY YSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLK KSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEF SKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRPAATKKAGQAKKKK	
364	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQVHLGELHAILRRQGEFYPFLKENAEKIQQILTFKIPYYVGPLARGNSRFAWASYNSNEKMTP NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLIKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGE IVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENI IHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGPPKKKRKVYPYDVPDYA	
365	MKKPYSIGLDIGTNSVGWSVVTDDYKVPAKKMKVLGNTDKEYIKKNLIGALLFDSGETAEATRMKRTARRRYTRRRNRILYLQDIFSPEILNQVDESFLHRLDDSFLVAEDKRGERHVIF GNIADEVKYHKEFPTIYHLRKHLADSSEKADLRLVYLALAHIIKYRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNMVSLVFGLMPNFKSNFELDEDAKLQFSRDSYDEDLENLLGIIGDEYADVFVAAKKVVDSILLSGILTTNNHSTKAPLSASMIDRYDEHNSDKKLLRDFIRTNIGKEVFKEVFYDTSK NGYAGYIDGKTNQEDFYKYLKNLLQKVDGGDYFIEKIEREDFLRKQRTFDNGSIPHQVHLDEMKAILRRQGEFYPFLKENAEKIQQILTFKIPYYVGPLARGNSRFAWASYNSNEKMTP WNFDNVIDKTSSAQAFIERMTNNDLYLPDQKVLPKHSLLYQKFAVYNELTKIKYVTETGEARLFDVFLKKEIFDGLFKKERKVTKKKILNFLDKNFDEFRITDIQGLDNETGNFNASYG TYHDLLKIIGDKEFMDSSDNVDVLEDIVLSLTLFEDREMIKQRLIKYEDIFSKKVIANLTRRHYTGWGRLSAKLINGIKDKHSKRKTILDYLIDDGHSNRNFMQLINDDNLSFKDEIANS QVIGDGDDLHQVVQELAGSPAIKKGILQSLKIVDELVKVMGYNPEQIVVEMARENQTTARGRNNSQQRLGSLTKAIQDFGSDILKRYPVENNQLQNDQLYLYYLQNGKDMYTGDTDLDIH NLSQYDIDHIIPQSFIKDNSLDNRVLTNSKSNRGKSDNVPSNEVVKRMKGFWLKQLDAKLISQRKFDNLTKAERGGLSAEDKAGFIKRQLVETRQITKHVARILDERFNRDFDKNDKRI RNVKIVTLKSNLVSNFRKEFGFYKVREINNFHHAHDAYLNAVVAKALLIRYPKLEPEFVYGEYPKYNSYRRERKSATEKMFFYSNIMNMFKTTIKLADGRVVEKPVIETANEETGEIAWDK TKHFANVKKVLSYPQVSIVKKVEEQTGGFSKESILPKGGSDKLIARKTKNNYLSTQKYGGFDSPTVAYSIMFVADIEKGKSKRLKTVKEMIGITIMERSRFESNSVTFLEEKGYRNIRE NTIIKFPKYSLFELENGRRRLLASAIELQKGNEMFLPQQFVNLLYHAQHANKEDSVIYLEKHRHELSELFHHIIGVSEKTILKPKVEMTLNEAEFEKHFEFDEVSELAQSFISLLKFTAF GAPGGFKFLDADIKQSNLRYQTVTEVLSSTLIHQSVTGLYETRIDLSKLGGE	
366	MRNMKNPYTIGLDIGTNSVGWAVLTNQYDLVKRKMKVAGNSDKKQIKKNFWGVRLFDDGQTAVDRRMNRTARRRIERRRNRISYLQEIFAVEMANIDANFFCRLNDSFYVDSEKRNSRH PFFATIEEEVAYHKNYRTIYHLRKELVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTKNTSVDGVYEQFIQTYNQVFMSNIEEGTLAKVEENILEVANILAGKFTRREKFERIILQLY PGEKSTGMFAQFISLIVGSKGNFQKVFDLIEKTDIECAKDSYEEDLEALLAIIGDEYAELFVAAKNTYNAVVLSSIITVTATETNAKLSASMIERFDAHEKDLGELKAFIKLHLPKQYQ EIFNNAAIDGYAGYIDGKTKQVDFYKYLKTILENIEGADYFIAKIEEENFLRKQRTFDNGAIPHQLHLEELEAIIHQQAKYYPFLREDYEKIKSLVTFRIPYFVGPLAKGQSEFAWLTR KADGEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYMVYNELTKVRYIDDQGKTNYFSGQEKQQIFNDLFKQKRKVKKKDLELFLRNINHIESPTIEGLEDSFN ASYATYHDLLKVGMKQEIILDNPLNTEMLEDIVKILTVFEDKPMIKEQLQQFSDVLDGGVLKKLERRHYTGWGRLSAKLLVGIRDKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSI IEKEQVSTTDKDIQSIVADLAGSPAIKKGILQSLKIVDELVSIMGYPPQTIVVEMARENQTTQKGKKNSKPRYLKSLEKAIKEFGSKILKEHPTDNQELKNTQLYLYYLQNGKDMYTGQE LDIHNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGDDVPPLEIVRKRKVFWEKLYQGNLMSKRKFDYLTKAERGGLTDADKARFIHRQLVETRQITKNVANILHQRFNNETDNH GNTMEQVRIVTLKSALVSQFRKQFQLYKVREVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGEYHQFDWFKANKATAKKQFYTNIMLFFGQKERIIDENGEILWDKKYLETIKKVL DYRQMNIVKKTEIQKGEFSKATIPKGNSSKKLIPEKRENWDPMKYGGLDSPNMAYAVIIEHAKGKKKLIFEKKIIRITIMERKMFEKDEEAFLEEKGYRHPKVLTKLPKYTLYECEKGRR RMLASANEAQKGNQLVLSNHLVSLLYHAKNCEASDGKSLKYIEAHRETFSELLAQVSEFATRYTLADANLSKINNLFEQNKEGDIQAIAQSFVDLMAFNAMGAPASFKFFEATIDRKRY TNLKELLSSTIIYQSITGLYESRKRLDD	

367	MKNMKNPYTIGLDIGTNSVGWAVLTNQYDLVKRKMKVAGNSDKKQIKKNFWGVRLFDDGQTAVDRRMNRTARRRIERRRNRISYLQEIFAVEMANIDANFFCRLNDSFYVDSEKRNSRH PFFATIEEEVAYHKNYRTIYHLREELVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTKNTSVDGVYEQFIQTYNQVFMSNIEEGTLAKVEENIEVANILARKFTRREKFERILQLY PGEKSTGMFAQFISLIVGSKGNFQKVFDLIEKTDIECAKDSYEEDLETLLAIIGDEYAELFVAAKNTYNAVVLSSIITVTDTETNAKLSASMIERFDAHEKDLVELKAFIKLNLPKQYE EIFSNAAIDGYAGYIDGKTKQVDFYKYLKTILENIEGADYFIAKIEEENFLRKQRTFDNGAIPHQLHLEELEAIIHQQAKYYPFLREDYEKIKSLVTFRIPYFVGPLAKGQSEFAWLTR KADGEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYMVYNELTKIRYIDDQGKTNYFSGQEKQQIFNDLFKQKRKVKKKDLELFLRNINHVESPTIEGLEDSFN ASYATYHDLMKVGIKQEILDNPLNTEMLEDIVKILTVFEDKRMIKEQLKQFSDVLDGTVILKKLERRHYTGWGRLSAKLLVGIRDKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSI IEKEQVSTADKDLQSIVADLAGSPAIKKGILQSLKVVEELVSVMGYPPQTIVVEMARENQTTNKGKNNSKPRYKSLEKAIKEFGSQILKEHPTDNQELKNNRLYLYYLQNGKDMYTGQE LDIHNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGDNVPPLEIVQKRKIFWEKLYQGNLMSKRKFDYLTKAERGGLTPEADKARFIHRQLVETRQITKHVARILDERFNNELDSKG EDIMEPVRIVTLKSALVSQFRKQFQLYKVREVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGEYHQFDWFKANKATAKKQFYTNIMLFFAQKERIIDENGEILWDKKYLETIKKVL NYRQMNIVKKTEIQKGEFSNQNPKPRGDSSKLIPKKTNLNPIKYGGFEGSNMAYAIIIEHEKRKKVTIEKKLIQINIMERKAFEKDEKVFLEGKGYHQPKVLTKLPKYALYECENGRR RMLGSANEVHKGNQMLLPNHLMTLLYHAEKREAIDGESLAYIEAHKAVFGELLAHISEFARKYTLANDKLDEINMLYERNKDGDVKSIAESFVSLKKFNAFGVHKDFNFFGTTIKRKRD RKLKELLNSTIIYQSITGLYETRIDLGKLGED	AKI50529.1 [Listeria monocytogenes]
368	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRIARRYTRRRNRILYLQEIFAEKMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFPTIYHLRKELADKKEKADLRLVYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTTFENNHLLSQNVDVEAILTDKISKSAKKDRILAQYPDQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEELVDKEASAEAFIHCMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEYRKVSKKQLLDFLAKEFEEFRIVDVTGLDKENKAFNASLG TYHDLEKILGKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLDIYKDFFTESQLKKLYRRHYTGWGRLSAKLINGIRNKENQKTILDYLIDDGSANRNFMQLIKDAGLSFKSIISKAQ SGSHSDNLKEVVSELAGSPAIKKGILQSLKIVDELVKVMGYKPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKIMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNDYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRTDKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	CFQ25032.1 [Streptococcus agalactiae]
369	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKASLSDSMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKIMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDERFNNELDSKG RRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIV WDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLN IRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYQDNKENISVDELAN NIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	CFV16040.1 [Streptococcus agalactiae]
370	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSRFAWAEYKADEKITP WNFEDIVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGEAKFFDTNMKQEIFDHVFKEYRKVSKQKLLDFLAKEFDDFKIVDVTGLDKENKAFNASLG TYHDLRKILDKSFLDDVNEKIIEDIIQTLTLFEDREMIRQRLQKYSDIFTTQQLKKLERRHYTGWGRLSYKLINGIRNKENNKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQ IIGYDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLGSQILKEHPVENTQLQNEKLYLYYLQNGKDMYTGETLDINNL SQYDIDHIIPQAFIKDNSLDNRVLTRSDKNRGKSDDVPSIEVVKKMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGAQRRIRN VKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKIIRKVLSYPQ VNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWNPKKYGGFDSPVVAYSILVFADVEKGKSKKLRKVQDMVGITIMEKKRFEKHPVDFLEQRGYRNVRLEKIKLIKLPKYSLFELE NKRRRLLASARELQKGNELVIPQRFTTLLYHSYRIEKDYEPEHREYVEKHKDEFKELLEIYSVFSRKYVLADNNLITKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENI DRKRYTSVTECLNATLIHQSITGLYETRIDLSKLGED	WP_049516684.1 [Streptococcus anginosus]
371	MMKKEYSIGLDIGTNSVGWAVITDDYKVPAKKMKVLGNTDKQSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRRNRLRYLQEIFAEEMMQVDESFFQRLDDSFLVEEDKRGSRYPI FGNIAAEVKYHDDFPTIYHLRKHLVDISQKADLRLVYLALAHMIKFRGHFLIEGQLKAENTNVQALFKDFVEVYDKTVEESHLSEMTVDALSILITEKVSKSRRLENLVECYPTEKKNTL FGNLIALSLGLQPNFKTNFQLSEDAKLQFSKDTYEEDLEGLLGEIGDEYADLFASAKNLYDAILLSGILAVDDNTTKAPLSASMVKRYEEHQKDLKKLKDFIKVNAPAQYDDIFKDETK NGYAGYIENGVKQDEFYKYLKNTLSKIDGSDYFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARGKSRFAWAEYKADEKITP WNFDDIILDKEKSAEKFITRMTNFDLYLPEEKVLPKHSLLYEYFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKEKRKVTKDKLLNYLNKEFEEFRIVNLTGLDKENKAFNASLG TYHDLRKILDKSFLDDVNEKIIEDIIQTLTLFEDREMIRQRLQKYSDIFTTQQLKKLERRHYTGWGRLSYKLINGIRNKENNKTILDYLIDDGYANRNFMQLINDDALSFKEEIARAQ IIGYDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLGSQILKEHPVENTQLQNEKLYLYYIQNGKDMYTGETLDINNL SQYDIDHIIPQAFIKDNSLDNRVLIRSDKNRGKSDDVPSIEVVKKMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGAQRRIRN VKIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFLFYSNILRFFKKEDIQTNEDGEIAWNKEKHIKIIRKVLSYPQ VNIVKKTEEQTGGFSKESILPKGESDKLIPRKTKNSYWNPKKYGGFDSPVVAYSILVFADVEKGKSKKLRKVQDMVGITIMEKKRFEKHPVDFLEQRGYRNVRLEKIKLIKLPKYSLFELE NKRRRLLASARELQKGNELVIPQRFTTLLYHSYRIEKDYEPEHREYVEKHKDEFKELLEIYSVFSRKYVLADNNLITKIEMLFSKNKDAEVSSLAKSFISLLTFTAFGAPAAFNFFGENI DRKRYTSVTECLNATLIHQSITGLYETRIDLSKLGED	WP_049516684.1 [Streptococcus anginosus]

372	IFNDLFKQKRKVKKKDLELFLRNINQIESPTIEGLEDSFNASYATYHDLLKVGMKQEILDNPLNTEMLEDIVKILTVFEDKRMIKEQLQQFSDVLDGVVLKKLERRHYTGWGRLSAKLL VGIRDKQSHLTILEYLMNDDGLNRNLMQLINDSNLSFKSIIEKEQVSTTDKDLQSIVADLAGSPAIKKGILQSLKIVDELVSIMGYPPQTIVVEMARENQTTVKGKNNSRPRYKSLEKA IKEFGSQILKEHPTDNQELKNNRLILYLYQLQNGKDIYTGQELDIHNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGDDVPPLEIVRKRKVFWEKLYQGNLMSKRKFDYLTKAERG GLTEADKARFIHRQLVETRQITKNVANILHQRFNNETDNHGNTMEQVRIVTLKSALVSQFRKQFQLYKVREVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGEYHQFDWFKANKAT AKKQFYTNIMLFFAQKERIIDENGEILWDKKYLETIKKVLDYRQMNIVKKTEIQKGEFSKATIKPKGNSSKLIPRKENWDPMKYGGLDSPNMAYAVIIEHAKGKKIVIEKKLIQINIM ERKMFEKDEEAFLEEKGYRHPKVLITKLPKYTLYECEKGRRRMLASANEAQKGNQLVLSNHLVSLLYHAKNCEASDGKSLKYTEAHRETFSELLAQVSEFATRYTLADANLSKINNLFEQ NKEGDIKXIAQSFVDLMVFNAMGAPASFKFFEATIDRKRYTNLKELLSSTIIYQSITGLYESRKRLDD	EFR83390.1 [Listeria monocytogenes FSL F2-208]
373	MKKPYSIGLDIGTNSVGWAVITDDYKVPAKKMKVLGNTNKESIKKNLIGALLFDAGNTAADRRLKRTARRYTRRRNRILYLQEIFAAEMNKVDESFFHRLDDSFLVPEDKRGSKYPIF GTLEEEKEYHKQFPTIYYLRKILASKEKVDLRLIYLALAHIIKYRGHFLYEDSFDIKNNDIQKIFNEFTILYDNTFEESSLSKGNAQVEEIFTDKISKSAKRDRVLKLFPDEKSTGLF SEFLKLIVGNQADFKKHFDLIEEKAPLQFSKDTYEEDLESLLGQIGDVYADLFVVAKKLYDAILLAGILSVKDPGTKAPLSASMIERYDNHQNDLSALKQFVRRNLPEKYAEVFSDDSKD GYAGYIDGKTTQEGFYKYIKNLISKIEGAEYFLEKIREDFLRKQRTFDNGSIPHQIHLQEMNAILRHQGEYYPFLKENKDKIEQILTFRIPYYVGPLARGNSDFAWLSRNSDEAIRPW NFEEMVDKSSSAEDFIHRMTNYDLYLPEEKVLPKHSLLYETFTVYNELTKVKYIAEGMKDYQFLDSGQKKQIVNQLFKEKRKVTEKDIIHYLHNVDGYDGIELKGIEKHFNSSLSTYHD LLKIIKDKEFMDDPKNEEIFENIVHTLTIFEDRVMIKQRLNQYDSIFDEKVIKALTRRHYTGWGKLSAKLINGIRDKQTSKTILDYLIDDGYSNRNFMQLINDDGLSFKETIQKAQVVG ETNDVKQVVQELPGSPAIKKGILQSLKIVDELVKVMGHAPESVVIEMARENQTTNKGKSKSQQRLKTLSDAISEIGSNILKEHPTDNIQLQNDRLFLYYLQNGKDMYTGEALDINQLSN YDIDHIIPQAFIKDDSLDNRVLTSSKDNRGKSDNVPSLEIVEKMKGFWQQLLDSKLISERKFNNLTKAERGGLDERDKVGFIKRQLVETRQITKHVAQILDDRFNAEVNEKNQKLRSVK IITLKSNLVSNFRKEFGLYKVREINDYHHAHDAYLNAVVAKAILKKYPKLEPEFVYGDYQKYDLKRYISRSRDPKEIEKATEKYFFYSNLLNFFKDKVYYADGTIIQRGNVEYSKDTGE IAWNKKRDFAIVRKVLSYPQVNIVKKTEEQTGGFSKESILPKGNSDKLIPRKTKNVQLDTTKYGGFDSPVIAYSILLVADVEKGKSKKLKTVKSLIGITIMEKVKFEANPVAFLEGKGY QNVVEENIIRLPKYSLFELENGRRRMLASAKELQKGNEMVLPSYLIALLYHAKRIQKKDEPEHLEYIKQHSEFNDLLNFVSEFSQKYVLAESNILEKIKNLYIDNEQTNMEEIANSFIN LLTFTAFGAPAVFKFFGKDIERKRYSTVTEILKATLIHQSITGLYETRIDLSKLGEE	
374	MLGTVLFDSGETAQARRLKRTNRRRYTRRRYRLCQLQNIFATEMVKVDDTFFQRLSESFLYYQDKAFDKHPIFGNSKEERAYHKTYPTIYHLRKDLADRDQKADLRLIYLALSHIIKFR GHFLIEGKLNSENTDVQKLFIALVTYYNLLFEEPIAGETCDAKALIJTAKTSKSKRLESLISEFPGQKKNGLFGNLIALALGLRPNFKSNFGLSEDAKLQITKDTYEEELDNLLAEIGD HYADLFLAAKNLSDAIILSDILTLSDENTRAPLSASMIKRYEEHQEDLALLKKLVKEQMPEKYWEIFSNAKKNGYAGYIEGKVSQEDFYRYIKPILSRLKGGDEFLAKIDRDDFLRKQR TFDNGSIPHQIHLKELHAIIRRQEKYYPFLAEQKEKIEQLLCFRIPYYVGPLAKGGNSRFAWLKRRSDEPITPWNFKDVVDLEESAQAFIEGMTNYDTYLPEEKVLPKHSPLYEMFTVY NELTKVKYIAENMTKPLYLSAEQKEAIIDHLFKQTRKVTVKDLKEKYFSQIEGLENVDVTGVEGAFNASLGTYNDLLKIIKDKAFLDDEANAEILEEIVLILTLFQDEKLIEKRLAKYA NLFEKSVLKKLRKRHYTGWGRLSRQLIDGMKDKASGKTILDFLKSDGFEKLINRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKIVDELVKVMGHNPDNI VIEMSRENQTTAQGLKNARQRLKKIKEVMKKTGSRILEDNSERITNLTLQDNRLFLYLLQLGKDMYTGQDLDIINNLSQYDIDHIIPQSFIKDNSIDNLVLVTTQKANRGKSDNVPSIEVV RDMKDRVWRRQLANGAISRQKFDHLTKAERGGLADSDKARFLRRQLVETRQITKHVAQILDSRFNSKSNQNKKLARNVKIITLKSKIVSDFRKDFQLYKLREVNNYHHAHDAYLNAVVG TALLKKYPKLEAEFVYGDYKHFDLVKLISKSDPSLGKATAKVFFYSNIMNFFKEELSLADGTLMKRPVIETNTETGEVVWDKVKDFKTIRKVLSYPQVNIVKKTEIQSGAFSKESVLSK GNSDKLIERKKGWDPKKYGGFDSPNTAYSIFVVVAKVAKKRKAQKLKTVKEIVGITIMEQAEYEKDNIAFLEKKGYQDIQEKLLIKLPKYSLFELENGRRRLIASANEFQKGNELALSGKY MKFLYLASRYDKLSSKIESEQQKKLFVEQHLHYFDEILDIVVKHATCYIKAENNLKKIISLYKKKEAYSINEQALNMLNLFIFTSLGAPSTFVFFDETIDRKRYTTSSDVLNGILIQQS ITGLYETRIDLSRFGGD	
375	MSNKPYSIGLDIGTNSVGWAVITDDYKVPSKKMKVLGNTDKHFIKKNLIGALLFDEGTTAEDRRLKRTARRYTRRKNRLRYLQEIFSEEISKVDNSFFHRLDDSFLVPEDKRGSKYPI FATLTEEKEYYKQFPTIYHLRKQLADSKEKADLRLIYLTLAHMIKYRGHFLYEESFDIKNNDIQKIFNEFTIKYDNTFEGSSLSSGQNAQVEAIFTDKISKSTKRERVLKLFPDQKSTGL FSEFLKLIVGNQADFKKHFDLEEKAPLQFSKDTYDEDLENLLGQIGDDFADLFLVAKKLYDAILLSGILTVTDPSTKAPLSASMIERYENHQKDLTALKQFIKNNLPEKYDEVFSDQSK EGYAGYIDSKTTQEAFYKYIKNLLSKIDGADYLLDKIEREDFLRKQRTFDNGSIPHQIHLQEMNAILRRQGEHYPFLKENKEKIEKILTFRIPYYVGPLARGNRDFAWLTRNSDQAIRP WNFEEIVDKASSAEAFINKMTNYDLYLPEEKVLPKHSLLYETFAVYNELTKVKFIAEGLRDYQFLDSGQKKQIVNQLFKEKRKVTEKDIIHYLHTVDGYDGIELKGIEKQFNASLSTYH DLLKIIKDKEFMDDPKNEEILENIVLTLTIFEDREMIKQRLKQYSSIFDEKVIKALTRRHYTGWGKLSAKLINGIRDKQTSGKTILDFLKSDGFAKRNFMQLIHDDSLTFKEDIQKAQVSG QGDSLHEHIANLAGSPAIKKGILQSLKIVEELVKVMGHEPESIVIEMARENQTTTRGKKNSQQRYKRIENSLKILASGLNSKILKEHPTDNNQLQNDRLFLYYLQNGKDMYTGEALDI	WP_049531101.1[St reptococcus pseudopneumoniae]
376	MSNKPYSIGLDIGTNSVGWAVITDDYKVPSKKMTVLGNTDKHFIKKNLIGALLFDEGTTAEDRRLKRTARRYTRRKNRLRYLQEIFSGEMSKVDSSFFHRLDDSFLVPEDKRGSKYPI FATLVEEKEYHKQFPTIYHLRKQLADSKEKADLRLIYLVLAHMIKYRGHFLYEEAFDIKNNDIQKIFNEFTISYDNTFEGSSLSSGQNAQVETIFTDKISKSAKRERVLKLFPDEKSTGL FSEFLKLIVGNQADFKKHFDLGEKAPLQFSKDTYDEDLENLLGQIGDDFADLILLSGILTVTDPSTKAPLSASMIERYENHQKDLTTLKQFIKNNLPEKYDEVFSDQSK DGYAGYIDGKTTQEAFYKYIKNLLSKFEGTDYFLEKIREDFLRKQRTFDNGSIPHQIHLQEMNAILRRQGEHYPFLKENKEKIEKILTFRIPYYVGPLARGNRDFAWLTRNSDQAIRP WNFEEIVDKASSAEDFINKMTNYDLYLPEEKVLPKHSLLYETFAVYNELTKVKFIAEGLRDYQFLDSGQKKQIVNQLFKEKRKVTEKDIIHYLHTVDGYDGIELKGIEKQFNASLSTYH DLLKIIKDKEFMDDSKNEAILENIVHTLTIFEDREMIKQRLAQYDSIFDEKVIKALTRRHYTGWGKLSAKLINGIRDKQTSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG QGDSLHEHIANLAGSPAIKKGILQSLKIVEELVKVMGHEPESIVIEMARENQTTTRGKKNSQQRYKRIENSLKILASGLNSKILKEHPTDNNQLQNDRLFLYYLQNGKDMYTGEALDI	WP_049549711.1 [Streptococcus pseudopneumoniae]
377	MSNKPYSIGLDIGTNSVGWVITTDDYKVPSKKMKVLGNTDKHFIKKNLIGALLFDEGTTAEDRRLKRTARRYTRRKNRLRYLQEIFAEEMNKVDSSFFHRLDDSFLVPEDKRGSKYPI FATLAEEKEYHKNFPTIYHLRKQLADSKEKADLRLIYLALAHMIKYRGHFLYEEAFDIKNNDIQKIFNEFTINYDNTFEGSSLSGQNEQVEAIFSDKISKSAKRERVLKLFPDEKSTGL FSEFLKLIVGNQADFKKHFDLEEKAPLQFSKDTYDEDLENLLGQIGDGFADLFLVAKKLYDAILLSGILTVTDPSTKAPLSASMIERYQNHQNDIASLKQFIKNNLPEKYDEVFSDQSK DGYAGYVDGKTTQEAFYKYIKNLLSKFEGADYFLEKIREDFLRKQRTFDNGSIPHQIHLQEMNAILRRQGEHYPFLKENKEKIEKILTFRIPYYVGPLARGNRDFAWLTRNSDQAIRP WNFEEIVDKASSAEDFINKMTNYDLYLPEEKVLPKHSLLYETFAVYNELTKVKFIAEGLRDYQFLDSGQKKQIVNQLFKEKRKVTEKDIIQYLHNVDGYDGIELKGIEKQFNASLSTYH DLLKIIKDKEFMDDSKNEEILENIVHTLTIFEDREMIKQRIAQYDSIFDEKVIKALTRRHYTGWGKLSAKLINGIRDKQTGKTILDYLIDDGYSNRNFMQLINDDGLSFKEIIQKAQVF GKTNDVKQVVQELPGSPAIKKGILQSLKIVEELVKVMGHEPESIVIEMARENQTTTRGKKNSQQRYKRIENSLKILASGLNSKILKEHPTDNNQLQNDRLFLYYLQNGKDMYTGEALDI	WP_049538452.1 [Streptococcus pseudopneumoniae]

	NQLSSCDIDHIIPQAFIKDDSLDNRVLTSSKENRGKSDNVPCLEVVDRMKVFWQQLLDFKLISYRKFNNLTKAERGGLDERDKVGFIRRQLVETRQITKHVAQILDSRFNTEVTEKDKK NRNVKIITLKSNLVSNFRKEFGLYKVREINDYHHAHDAYLNAVVAKALLKKYPKLEPEFVYGDYQKYDLKRYISRSKDPKDIEKATEKYFFYSNLLNFFKEEVHYADGTIVKRENIEYS KDTGEIAWNKEKDFATIKKILSLPQVNIVKKTEEQTVGQNGGLFDNNIVSKKKVVDASKLIPIKSGLSPEKYGGYARPTIAYSVLVIADIEKGKTKKLKRIKEMIGITVQDKKIFESNP IAYLEECGYKNINPNLIIKLPKYSLFEFNGGQRRLLASSIELQKGNELILPYHFTALLYHTQRINKISEPIHKQYVEAHQNEFKELLTTIISLSKRYIQKPNVESLLQQAFEQADKDIY QLSESFISLLKLTSFGAPGAFKFLGVEISQSSVRYKPNSQFLDATLIHQSITGLYETRIDLSKLGED	BAQ51233.1 [Streptococcus pyogenes]
378	MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDE HHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTF RIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDEIVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSV KELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKR VILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	WP_049474547.1 [Streptococcus mutans]
379	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKTDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSAKKLYDSIILSGILTVTDVSTKAPLSASMIQRYNEHQMDLTQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERENDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLKDNQDRIEKILTFRIPYYVGPLASGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYVTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILIDKLKNDEDLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAELINGIRDKESQKTILDYLIDDGRSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLITDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYIL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNRLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNID RKRYTSTTEILKATLIHQSITGLYETRIDLSKLGGD	KLJ37842.1 [Streptococcus agalactiae]
380	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILKDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLKDGVKNLASDLINGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDELFNNELDSKG RRIRKVKIVTVKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVADIKKGKSKKLKTVTELLGITIMERSRFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKFFDKNIDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGGD	KLJ72361.1 [Streptococcus agalactiae]
381	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLEKILKDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDDGLSFKSIISKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLKDGVKNLASDLINGNILKEYPTDNQALQNERLFLYYLQNGRDMYTGEAL DIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVARILDELFNNELDSKG RRIRKVKIVTVKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVADIKKGKSKKLKTVTELLGITIMERSRFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKFFDKNIDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGGD	KLL20707.1 [Streptococcus agalactiae]
382	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQYQDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKERQDRIEKILTFRIPYYIGPLARGKSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLLYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDVIGLDKENKVFNASYGT YHDLEKILKDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDDGLSFKSIISKAQ DDGLSFKSIISKAQAGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLKDGVKNLASDLINGNILKEYPTDNQALQNERLFLY YLQNGRDMYTGEALDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFIQRQLVETRQITKHVAR ILDERFNNELDSKGRRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLESEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVV KDDIEVVNNDTGEIVWDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSKESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKSKKLKTVTELLGITIMERSRFE	KLL20707.1 [Streptococcus agalactiae]

	KNPSAFLESKGYLNIRADKLIILPKYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYIASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLY QDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED	
383	MNKPYSIGLDIGTNSVGYSVVTDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTASDRRLKRTARRRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEDDKRGSKYPIF ATMQEEKDYHEKFPTIYHLRKELADKKEKADLRLVYLALAHIIKFRGHFLIEGNLDSENTDVHVLFLNLVNIYNNLFEEDIVETASIDAEKILTSKTSKSRRLENLIAEIPNQKRNMLF FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSVAKKLVDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFTDSSK DGYAGYIEGKTNQGAFYKYLSKLLTKQEGSEYFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENLDRIEKILTFRIPYYVGPLARGKSDFAWMTRKTDDSIRP WNFEELVDKEASAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEHRKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLG TYHDLKKILDKQDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIKSDGRANRNFMQLIHDDGLSFKPIIDKAQ AGSHSDNLKEVVGELAGSPAIKKGILQSLKIVDELVKVMGYEPEQVVVEMARENQTTNKGRRNTRQRYKLLKESIKEFGSQILKEHPTDNQELKNNRLYLYYLQNGKDMY DIDNLSQYDIDHIVPQAFIKDDSIDNRVLVSSAKNRGKSDNVPSDIVKKARKAFWKKLLDAKLMSQRKYDNLTKAERGGLSTSDDKARFIHRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KRVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGALTNESIYARGSFDKLISRKHRFESSKYGGFGSPTVTYSVLVVAKSKVQDGKVKKIKTGKELIGITLLDKLVFEKNPLKFIEDKGYGN VQIDKCIKLPKYSLFEFENGTRRMLASVMANNNSRGDLQKANEMFLPAKLVTLLYHAHKIESSKELEHEAYIILDHYNDLIQLLSYIERFASLYVDVEKNISKVKELFSNIESYSISEIC SSVINLLTLTASGAPADFKFLGITIPRKRYGSPQSILSSTLIHQSITGLYETRIDLSQLGGD	
384	MRKPYSIGLDIGTNSVGWAVITDDYKVPSKKMRIQGTTDRTSIKKNLIGALLFDNGETAEATRLKRTTRRRYTRRKYRIKELQKIFSSEMNELDIAFFPRLSESFLVSDDKEFENHPIF GNLKDEITYHNDYPTIYHLRQTLADSDQKADLRLIYLALAHIIKFRGHFLIEGNLDSENTDVHVLFLNLVNIYNNLFEEDIVETASIDAEKILTSKTSKSRRLENLIAEIPNQKRNMLF GNLVSLALGLTPNFKTNFELLEDAKLQISKDSYDEDLDNLLAQIGDQYADLFIAAKKLSDAILLSDIITVKGASTKAPLSASMVQRYEEHQDDLALLKNLVKKQIPEKYKEIFFDNKEKN GYAGYIDGKTSQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSAKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEAIQKAQVS GQGHSLHEQIANLAGSPAIKKGILQSVKVVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEI VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGALTNESIYARGSFDKLISRKHRFESSKYGGFGSPTVTYSVLVVAKSKVQDGKVKKIKTGKELIGITLLDKLVFEKNPLKFIEDKGYGN VQIDKCIKLPKYSLFEFENGTRRMLASVMANNNSRGDLQKANEMFLPAKLVTLLYHAHKIESSKELEHEAYIILDHYNDLIQLLSYIERFASLYVDVEKNISKVKELFSNIESYSISEIC SSVINLLTLTASGAPADFKFLGITIPRKRYGSPQSILSSTLIHQSITGLYETRIDLSQLGGD	
385	LKLYPGEKSTGMFAQFISLIVGSKGNFQKPFDLIEKSDIECAKDSYEEDLESLIALIGDEYAELFVAAKNAYSAVVLSSITTVAETETNAKLSASMIERFDTHEEDLGELKAFIKLHLDP KHYEEIFSNTEKHGYAGYIDGKTKQADFYKYMKTTLENIEGADYFIAKIEKENFLRKQRTFDNGAIPHQLHLEELEAILHQQAKYYPFLKENYDKIKSLVTFRIPYFVGPLANGQSEFA WITRKADGEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYLVYNELTKVRYINDQGKTSYFSGQEKEQIFNDLFKQKRKVKKKDLELFLRNMSHVESPTIEGLE DSFNSSYSTYHDLLKVGIKQEILDNPVNTEMLENIVKILTVFEDKRMIKEQLQQFSDVLDGVVLKKLERRHYTGWGRLSAKLIMGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEAIQKAQVS FKSIIEKEQVTTADKDIQSIVADLAGSPAIKKGILQSLKIVDELVSVMGYPPQTIVVEMARENQTTGKGKNNSRPRYKSLEKAIKEFGSQILKEHPTDNQELKNNRLYLYYLQNGKDMY TSQQLDIDHLSVQFIVPQSFITDNSIDNLVLTSSAGNREKGNDVPPLEIVQKRKVFWEKLLQAKLISQRKFDYLTKAERGGLTEADKARFIHRQLVETRQITKNVANILHQRFNYG KDDHGNTMKQVRIVTLKSALVSQFRKQFQLYKVRGVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGDYHQFDWFKANKATAKKQFYTNIMLFFAQKDRIIDENGEIIWDKKYLDTV KKVMSYRQMNIVKKTEIQKGEFSKATIKPKGNSSKLIPRKTNWDPMKYGGLDSPNMAYAVVIEYAKGKNKLVFEKKIIRVTIMERKAFEKDEKAFLEEQGYRQPKVLAKLPKYTLYECE EGRRRMLASANEAQKGNQQVLPNHLVTLLHHAANCEVSDGKSLDYIESNREMFAELLAHVSEFAKRYTLAEANLNKINQLFEQNKEGDIKAIAQSFVDLMAFNAMGAPASFKFFETTIE KRKYNNLKELLNSTIIYQSITGLYESRKRLDD	
386	MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRIRYLQEIFSSEMSKVDDSFFHRLEESFLVEEDKKHERHPIF GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF GNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLAKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSAKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEAIQKAQVS GQGDSLHEQIANLAGSPAIKKGILQSVKVVDELVKVMGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV KVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGEFKKYDVRKFIAKSDKEIGKATAKYFFYSNIMNFFKKEVKLADGTVVERPLIETNEETGEI VWDKGRDFATVRKVLSMPQVNIVKKTEVQTGALTNESIYPKGNSKLIPRKTNWDPMKYGGYAGISNSYAVLVKAIIEKGVKKKETMVLEFQGISILDRITFEKDKRAF KLIGKGYKDIKKIELPKYSLFELKDGSRRMLASILSTNNKRGEIHKGNELFVPQKFTTLLYHAKRINNPINKDHIEYVKKHRDDFKELLNYVLEFNEKYVGATKNGERLKEAVADFDSK SNEEICTSFLGAVNSKNAGLFELTSLGSASDFEFLGVKIPRYRDYTPSSLLIKDSTIGLYETRIDLSKLGED	
387	MDQKYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKQSIKKNLIGALLFDSGETAEATRLKRTARRYTRRRNRLRYLQEIFAEEMNKVDENFFQRLDDSFLVDEDKRGERHPIF GNIAAEVKYHDDFPTIYHLRKHLADISQKADLRLVYLALAHMIKFRGHFLIEGQLKAENTNVQALFKDFVEVYDKTVEESHLSEMTVDALSILITPKVSKSRRLENLIAHYPAEKKNTLF GNLIALSLGLQPNFKTNFQLSEDAKLQFSKDTYEEDLGLLGEIGDEYADLFAAKNLVQAIELSGILTVTDPGTKAPLSASMIKRYDEHHQDLTLLKELVRTQLPDKYKEIFFDQSKN GYAGYIENGVKQDEFYKFLKGILQKINGSGDFLDKIDREDFLRKQRTFDNGSIPHQIHLQEMHAILRRQEDFYPFLKDNQEKIEKILTFRIPYYVGPLARKGSRFAWAEYKAKADEKITPW NFDDIILDKEKSAEKFITRMTLNDLYLPEEKVLPKHSLLYETFTVYNELTKVKYVNEQGETAFFDANMKQEIFDHVFKENRKVTKDALFSFLKTEGCFEKIVDLIGLDKENKVFNSSLGT YHDLIRKIILDKSFLDNEENEQIIEDIIQTITLFEDREMIRQRLQKYSDIFTKAQLKKLERCHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIAKAQV IGETDDLNQVVSDIAGSPAIKKGILQSLKIVDELVKVMGYKPENIVVEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDIDNLS SNEEICTSFLGAVNSKNAGLFELTSLGSASDFEFLGVKIPRYRDYTPSSLIKDSTIGLYETRIDLSKLGED	
388	MKKMLANIDGADYFIDQIEEENFLRKQRTFDNGTIPHQLHLEELEATILHQQAKYYPFLKDYEKIRSLVTFRIPYFIGPLANGQSDFAWLTRKADGEIRPWNIEEKVDFGKSAIDFIEK MTNKDTYLPKENVLPKHSLCYQKYMYYNELTKIRYIDDQGKTHHFSGQEKQQIFNGLFKQQRKVKKKDLERFLYTTINHIESPTIEGVEDAFNSSFATYHDLQKGGVTQEILDNPLNADM LEEIVKILTVFEDKRMIKEQLQSFSDVLDGTILKKLERRHYTGWGRLSAKLITGIRDKHSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSIIEKEQVSTADKGIQSIVAELAGSPAIK	

	KGILQSLKIVDELVGIMGYPPQTIVVEMARENQTTGKGKNNSKPRFISLEKAIKEFGSQILKEHPTDNQCLKNDRLYLYYLQNGKDMYTGKELDIHNLSNYDIDHIIPQSFITDNSIDN RVLVSSTANREKGDNVPLLEVVRKKKAFWEKLYQAKLMSKRKFDYLITKAERGGLTEADKANFIQRQLVETRQITKNVANILYQRFNCKQDENGNEVEQVRIVTLKSTLVSQFRKQFQLY KVREVNDYHHAHDAYLNGVVVANTLLKVYPQLEPEFVYGDYHQFDWFKANKATAKKQFYTNIMRFFAKENQIIDKNGEILWDNRYLDTIKKVLSYRQMNIVKKTEIQKGEFSNATVNPKG NSSKLISRKADWNPIKYGGFDGSNMAYSIVIEYEKRKKTVIKKELIQINIMERVAFEKDQKAFLEEKGYYSPKVLTKIPKYTLYECENGNRRMLGSANEAQKGNQMVLPNHLMTLLYH AKNCEANDGESLAYIEMHREMFAELLAYISEFAKRYTLANDRLEKINMFFEQNKKGDIKVIAKSFDKLKVFNAFGAPRDFEFFETTIKRKRVYNIKELLNATIIYQSITGLYEARKRLE D	ivanovii FSL F6-596]
389	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYYENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEDLEELLGKIGDDYADLFTLAKNLYDAIILSGILTADDSSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGT YHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKPYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRETNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIANIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFELEN GRKRLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDIKAIAQSFVDLMAFNAMGAPASFKFFETTIERKRY RKRYTSTTEILNATLIHQSITGLYETRIDLSKLGGD	ALF27331.1 [Streptococcus intermedius]
390	MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLIGALLFDSGNTAEDRRLKRTARRYTRRRNRILYLQEIFSEEMGKVDDSFFHRLEDSFLVTEDKRGERHPIF GNLEEEVKYHENFPTIYHLRQYLADNPEKTDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRF AEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIEDNYAELFLSAKKLYDSIILSGILTVTDVSTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKD GYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIERKDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKILTFRIPYYIGPLARGKSDFSWLSRKSADKITPW NFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPRHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKAFNASYGT YHDLRKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKNLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQV IGETDNLNQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYTGEELDIDYL SQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQ VKIVTLKSNLVSNFRKEFELYKVRETNDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQV NIVKKVEEQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIANIEKGKSKKLKLVKDLVGITIMERTIFEKNPVAFLERKGYRNVQEENIVKLPKYSLFELEN GRRRLLASARELQKGNEIVLPNHLGTMLYHAKNIHKVDEPKHLDEFKELLDVVSNFSKKNILAESNLEKIEELYAQNNNKDITELASSFINLLTFTAIGAPAAAFKFFDTNIDRKRY RKRYTSTTEILNATLIHQSITGLYETRIDLSRLGGD	EHN60060.1 [Listeria innocua ATCC 33091]
391	MKNMKKPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKIAGDSEKKQIKKNFWGVRLFDEGQTAADRRMARTARRRIERRRNRISYLQGIFAEEMSKTDANFFCRLSDSFYVDNEKRNSRH PFFATIEEEVEYHKNYPTIYHLRELVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTQNTSVDGIYKQFIQTYNQVFASGIEDGSLKKLEDNKDVAKILVEKVTRKEKLERILKLIY PGEKSAGMFAQFISLIVGSKGNFQKPFDLIEKSDIECAKDSYEEDLESLLALIGDEYAELFVAAKNAYSAVVLSSIITVAETETNAKLSASMIERFDTHEEDLGELKAFIKLHLPKHYE EIFSNTEKHGYAGYIDGKTKQADFYKYMKMTLENIEGADYFIAKIEKENFLRKQRTFDNGAIPHQLHLEELEAILHQQAKYYPFLKENYDKIKSLVTFRIPYFVGPLANGQSEFAWLTR KADGEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYLVYNELTKVRYINDQGKTSYFSGQEKEQIFNDLFKQKRKVKVKKDLELFLRNMSHVESPTIEGLEDSFN SSYSTYHDLLKVGIKQEILDNPVNTEMLENIVKILTVFEDKRMIKEQLQQFSDVLDGVVLKKLERRHYTGWGRLSAKLLMGIRDKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSI IEKEQVTTADKDIQSIVADLAGSPAIKKGILQSLKIVDELVSVMGYPPQTIVVEMARENQTTGKGKNNSRPRYKSLEKAIKEFGSQILKEHPTDNQELRNNRLYLYYLQNGKDMYTGQD LDIHNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGDDVPSPELEIVKKRKVFWEKLIQGNLMSKRKFDYLITKAERGGLTEADKARFIHRQLVETRQITKNVANILHQRFNNYEKDDH GNTMKQVRIVTLKSALVSQFRKQFQLYKVRDVNDYHHAHDAYLNGVVANTLLKVYPQLEPEFVYGDYHQFDWFKANKATAKKQFYTNIMLFFAQKDRITDENGEILWDKKYLDTVKKVM SYRQMNIVKKTEIQKGEFSKATIKPKGNSDKLIPRKTKDPMKYGGLDSPNVAYAVVIEVQKGKNKLKTVKELIGITIMERERFEKNPSAFLESKGYLN RMLASANEAQKGNQQVLPNHLVTLLHHVANCEVSDGKSLDYIESNREMFAELLAHVSEFAKRYTLAEANLNKINQLFEQNKEGDIKAIAQSFVDLMAFNAMGAPASFKFFETTIERKRY NNLKELLNSTIIYQSITGLYESRKRLDD	AHN30376.1 [Streptococcus agalactiae 138P]
392	MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRYTRRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVEEDKRGSKYPIF ATMQEEKDYHEKFPTIYHLRKELADKKEKADLRLVYLALAHIIKFRGHFLIEDDRFDVRNTDIQKQYQAFLEIFDTSFENNHLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGI FAEFLKLIVGNQADFKKHFNLEDKTPLQFAKDSYDEDLENLLGQIGDEFADLFSAAKKLYDSVLLSGILTVTDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSK DGYAGYIEGKTNQEAFYKYLSKLLTKQEDSEYFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDKIEKILTFRIPYYVGPLARGNSDFAWMTRKTDDSIRP WNFEDLVDKEKSAEAFIHRMTNNDLYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNVKQEIFDGVFKEHKRVSKKQLLDFLAKEFEEFRIVDVTGLDKENKAFNASLG TYHDLKKILDKDFLDNPDNESILEDIVQTITLFEDREMIKKRLENYKDLFTESQLKKLYRRHYTGWGRLSAKLINGIRDRESQKTILDYLISDGRANRNFMQLINDDGLSFKSIISKAQ SGSHSDQYDIDHIVPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSIDIVKARKAFWKKLLDAKLISQRKYDNLTKAERGGLSPDDKAGFIQRQLVETRQITKHVARILDERFNTRMDDNN KPIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFVYGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADETVVKDDIEVNNETGEIA WDKKKHFATVRKVLSYPQVNIVKKTEVQTGGFSKESILPKRKTKDIYLDPKKYGGFDSPIVAYSVLVLADIKKGKAQKLKTVKELIGITIMERERFEKNPSAFLESKGYLN IRTDKLIILPKYSLFELENGRKRMLASAGELQKGNELALPTQYMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKINKLYSDNKDNTPVDELAK NIINLFTFTSLGAPAAAFKFFDKSVDRKRYTSTKEVLDSTLIHQSITGLYETRIDLGKLGED	AHY15608.1 [Streptococcus iniae]
393	MRKPYSIGLDIGTNSVGWAVITDDYKVPSKKMRIQGTTDRTSIKKNLIGALLFDNGETAEATRLKRTTRRYTRRKYRIKELQKIFSSEMNELDIAFFPRLSESFLVSDDKEFENHPIF GNIKDEITYHNDYPTIYHLRQTLADSDQKADLRLIYLALAHIIKFRGHFLIEGNLDSENTDVQKLFIQLVQTYNQLFEENGISKTSKSRRLENLIAEIPNQKRNMLF GNLVSLALGLTPNFKTNFELLEDAKLQISKDSYEEDLDNLLAQIGDQYADLFIAAKKLSDAILLSDILRVNTEITKAPLSASMVQRYEEHQQDLALLKNLVKKQIPEKYKEIFDNKEKN GYAGYIDGKTSQEEFYKFIKPILEKLDGTEKLISKLERRDFLRKQRTFDNGSIPHQIHLNELHAIIRRQEKFYPFLKENQKKIEKILTFRIPYYVGPLANGQSSFAWLKRQSNESITPW NFEEVVDQEASAAFISERLIERMTNFDTYLPKEKVLPKHSLIYEKFMVYNELTKVKYVTEGMKKPAFLSSEDKEFIVNLLFKKERKVTVKQLKEEFYSFMKCFHTVTILGVEDRFNASLGTYH DLIKIFKDKAFLDDEANQDILEEIVWTLTLFEDQAMIERRLVKYADVFEKSVLKKLKKRHYTGWGRLSQKLINGIKDKQTGKTILGFLKDDGVANRNFMQLINDSSLDFAKIIKNEQEK TIKNESLEETIANLAGSPAIKKGILQSIKIVDELVKIMGQNPDNIVIEMARENQSTMQGIKNSRQRLRKLEEVHKNTGSKILKEYNVSNTQLQSDRLYLYLLQDGKDMYTGKELDYDNL SQYDIDHIIPQSFIKDNSIDNTVLTTQASNRGKSDNVPNIETVNKMKSFWYKQLKSGAISQRKFDHLTKAERGALSDFDKAGFIKRQLVETRQITKHVAQILDSRFNSNLTEDSKSNRN	AHY15608.1 [Streptococcus iniae]

	VKIITLKSKMVSDFRKDFGFYKLREVNDYHHAQDAYLNAVVGTALLKKYPKLEAEFVYGDYKHYDLAKLMIQPDSSLGKATTRMFFYSNLMNFFKKEIKLADDTIFTRPQIEVNTETGE IVWDKVKDMQTIRKVMSYPQVNIVMKTEVQTGGFSKESIWPKGDSDKLIARKKSWDPKKYGGFDSPIIAYSVLVVAKIAKGKTQKLKTIKELVGIKIMEQDEFEKDPIAFLEKKGYQDI QTSSIIKLPKYSLFELENGRKRLLASAKELQKGNELALPNKYVKFLYLASHYTKFTGKEEDREKKRSYVESHLYYFXEVKSSF	
394	MRKPYSIGLDIGTNSVGWAVITDDYKVPSKKMRIQGTTDRTSIKKNLIGALLFDNGETAEATRLKRTTRRRYTRRKYRIKELQKIFSSEMNELDIAFFPRLSESFLVSDDKEFENHPIF GNLKDEITYHNDYPTIYHLRQTLADSDQKADLRLIYLALAHIIKFRGHFLIEGNLDSENTDVHVLFLNLVNIYNNLFEEDIVETASIDAEKILTSKTSKSRRLENLIAEIPNQKRNMLF GNLVSLALGLTPNFKTNFELLEDAKLQISKDSYEEDLDNLLAQIGDQYADLFIAAKKLSDAILLSDIITVSGASTKAPLSASMVQRYEEHQDDLALLKNLVKKQLPEKYKEIFDNKENN GYAGYIDGKTSQEEFYKYIKPILLKLDGTEKLISKLEREDFLRKQRTFDNGSIPHQIHNELKAIIRRQEKFYPFLKENQKKIEKLFTFKIPYYVGPLANGQSSFAWLKRQSNESITPW NFEEVVDQEASARAFIERMTNFDTYLPEEKVLPKHSPLYEYFMVYNELTKVKYQTEGMKRPVFLSSEDKEEIVNLLFKKKERKVTVKQLKEEYFSKMKCFHTVTILGVEDRFNASLGTYH DLLKIFKDKAFLDDEAENQDILEEIVWTLTLFEDQAMIERRLVKYADVFEKSVLKKLKKRHYTGWGRLSQKLINGIKDKQTGKTILGFLKDDGVANRNFMQLINDSSLDFAKIIKNEQEK TIKNESLEETIANLAGSPAIKKGILQSIKIVDEIVKIMGQNPDNIVIEMARENQSTMQGIKNSRQRLRKLEEVVKHKNTGSKILKEYNVSNTQLQSDRLYLYLLQDGKDMYTGKELDYDNL SQYDIDHIIPQSFIKDNSIDNTVLTTQASNRGKSDNVPNIETVNKMKSFWYKQLKSGAISQRKFDHLTKAERGALSDFDKAGFIKRQLVETRQITKHVAQILDSRFNSNLITEDSKSNRN VKIITLKSKMVSDFRKDFGFYKLREVNDYHHAQDAYLNAVVGTALLKKYPKLESEFVYGDYKHYDLAKLMIQPDSSLGKATTRMFFYSNLMNFFKKEIKLADDTIFTRPQIEVNTETGE IVWDKVKDMQTIRKVMSYPQVNIVMKTEVQTGGFSKESIWPKGDSDKLIARKKSWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKDPIDF QTSSIIKLPKYSLFELENGRKRLLASAKELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDK PIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	KGE60856.1 [Streptococcus pyogenes SS1447]
396	MDEEAKIQLSKESYEEELESLILEKSGEEFRDVFLQAKKVYDAILLSDILSTKKQNSKAKLSIGMIERYDSHKKDLEELKQFVKANLPEKTAIFFKDSSKNGYAGYIDGKTTQEDFYKFL KKEINGIAGSERFMEKVDQENFLLKQRTTANGVIPHQVHLTELKAIIERQKPYYPSLEEARDKMIRLLTFRIPYYVGPLAQGEETSSFAWLERKTPEKVTPWNATEVIDYSASAMKFIQ RMINYDTYLPTEKVLPKHSLLYQKYTIFNELTKVAYKDERGIKHQFSSKEKREIFKELFQKQRKVTVKKLKQDFLSAHYNIEDAEILGVDKAFNSSYATYHDFLDILAKPNTERVAELLEQ PEMNAFEDIVKIITIFEDREMIRTQLKKYQSVLGDGFFKKLVKKHYTGWGRLSEKLINGIRDKKTNKTILDYLIDDDDFPYNRNRNFMQLINDDSLSFKEEIAKAQVIGETDNL EALLGSPAIKKGIWQTLKIVDELVKIMGRQPENIVVEMARENQRTNRSKPRLKALEEAIKSFDSPLLKEQPVDNQALQKDRLYLYYLQNGKDMYTGEALDIDRLSEYDIDHIIPRSFIV DNSIDNKVLVSSSKENRLKMDDVPPDQKVVIRMRRYWEKLLRANLISERKFAYLTKLELTPEDKARFIQRQLVETRQITKHVAAILDQYFNQPEESKNKGIRIITLKSSLVSQFRKTFGIN KVREINNHHHAHDAYLNGVVAIALLKKYPKLEPEFVYGNYTKFNIATAKKEFYSNIIRFFEKEEVSYDENGEIFWDKARHIPQIKKVISSHQVNIVKKVEVQTGGFYKETVNPK GKPDKLIQRKAGWDVSKYGGFGSPVVAYAVAFIYEKGKARKKAKAIEGITIMKQSLFEQDPIGFLSNKGYSNVTKFIKLSKYTLYELENGRRRMVASHKEAQKANSFILPEKLVTLLYH AQHYDEIAHKESFDYVNDHLSEFRELILDQVIDFSNRYTIAAKNTEKIAELFEQNQESTVQSLSQSFINLMQLNAMGAPADFKFFDVIIPRKRYPSLITEIWESTIIYQSITGLRETRTRM ATLWDGEQ	
397	MDLIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIKKNLLGALLFDSGNTAADRRLKRTARRRYTRRNRILYLQEIFAEEMSKVDDSFFHRLEDSFLVTEDKRGERHPIFGNLEEEV KYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRFAEFLKLI VGNQADFKKHFELEEKAPLQFSKDTYEEELEVLLAQIGDNYAELFLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKDGYAGYID GKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLAKDNQDRIEKILTFRIPYYVGPLARGKSDFAWLSRKSADKITPWNFDEIVD KESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYVTEQGKTAFFDANMKQEIFDGVFKVKRKVTVKDLKNDYFKKMEVFDSVEISGVEDRFNTSLGTYHDLLKIV KDKFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQVIGETDNL LDKFDLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYSDLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQVIGETDNL NQVVSDIAGSPAIKKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVEHSQLQNDRLFLYYLQNGRDMYTGEELDIDYLSQYDIDH IIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKNVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVAQILDSRFNTNFDQNGKDMYTGETLDINNLS SNLVSNFRKEFELYKVREINDYHHAHDAYLNAVIGKALLGVYPQLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDNGEIIWKKDEHISNIKKVLSYPQVNIVKKVE EQTGGFSKESILPKGNSDKLIPRKTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLENGRKRLLA SARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELIYAQNNGEDLKELASSFINLLTFTAIGAPATFKFFDKNIDRKRYTST TEILNATLIHQSITGLYETRIDLSKLGGD	ESR09100.1 [Streptococcus iniae IUSA1]
398	MEQDEFEKDPIAFLEKKGYQDIQTSSIIKLPKYSLFELENGRKRLLASAKELQKGNELALPNKYVKFLYLASHYTKFTGKEEDREKKRSYVESHLYYFDEIMQIIVEYSNRYIILADSNL IKIQNLYKEKDNFSIEEQAINMLNLFTFTDLGAPSAFKFFNGDIDRKRYISSTNEIINSTLIYQSPTGLYETRIDLSKLGGK	
399	MKKEYSIGLDIGTNSVGWAVITDDYKVPAKKMKVLGNTDKQSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRLRYLQEIFTEEMNKVDENFFQRLDDSFLVEEDKQGSKYPIF GTLKEEKEYHKKFPTIYHLRKELANSKEKADLRLVYLALAHMIKFRGHFLYEGDLKAENTDVQALFKDFVEVEYDKTIEESHLSEITVDALSILITEKVSKSSRLENLIAHYPTEKKNTLF GNLIALSLGLQPNFKTNFQLSEDAKLQFSKDTYEEDLEELLGEIGDEYADLFAASAKNLYDAILLSGIILAVDDNTTKAPLSASMVKRYEEHQKDLKLLKDFIKVNAPDQYNAIFKDKNKK GYAGYIEGVEQDEFYKYLKGLLKKINGSGDFLDKIDCEDFLRKQRTFDNGSIPHQIHLQEMHAILRRQGEHYPFLKENQDKIEKILTFRIPYYVGPLARKGSRFAWAEYKAEKITPWN NFDDIILDKEKSAEKFITRMTLNDLILYLPEEKVLPKHSPLYEYFTVYNELTKVKYVNEQGEAKFFDTNMKQEIFDHVFKENRKVTDKDLLINYLNKEFEEFRIVNLTGLDKENKAFNSSLGT YHDLRKIILDKSFLDDKANEKTIEDIIQTLTLFEDREMIRQRLQKYSDIFTKAQLKKLERRHYTGWGRLSYKLINGIRNKENKKTILDYLIDDGYANRNFMQLINDDALSFKEEIAKAQI IGDVDDIANVVHDLPGSPAIKKGILQSVKIVDELVKVMGHNPANIIIEMARENQTTDKGRRNSQQRLKLLQDSLKNLDNPVNIKNVENQQLQNDRLFLYYIQNGKDMYTGETLDINNLS QYDIDHIIPQAFIKDNSLDNRVLITRSDKNRGKSDDVPSIEVVVHEMKSFWSKLLSVKLITQRKFDNLTKAERGGLTEEDKAGFIKRQLVETRQITKHVAQILDERFNTEFDGAQRRIRNV KIITLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVVGNALLLKYPQLEPEFVYGEYPKYNSYRSRKSATEKFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRESDKLIARKKDWDPKKYGGFDSPTVAYSVLVIADVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEMALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEILEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	
400	MTKKEQPYNIGLDIGTGSVGWAVTNDNYDLLNIKKKNLWGVRLFEGAQTAKETRLNRSTRRYRRRKNRIRWLNELIFSEELANTDPSFLIRLQNSWVSKKDPDRKRDKYNLFIDNPYTD KEYYRE FPTIFHLRKELIINKNKADIRLVYLALAHNILKYRGNFTYEHQKFNISTLLNSNLSKLEIELNQQLIKYDISFPDNCDWNHISDILIIGRGNAIQKSSNILNNFTLDKETKKLLKE VINLILGNVAHLNTIFKTSLTKDEEKLSFSGKDIELDDLSIILDDQFTVLDTANRIYSTITLNEIILNGESYFSMAKVNQYENHAIDLCKLRDMWHTTKNEKAVGLSRQAYDDYINK PKYGTKELYTSLKKFLKVALPTNLAKEAEEKISKGTYLVKPRNSENGVVPYQINKIEMEKIIDNQSQYYPFLKENKEKLLSILSFRIPYYVGPLQSSEKNPFAWMERKSNGHARPWNFD EIVDKEKSSNKFIRRMTVTGYLYGEPVLNENLINNIRITENLKTNPTGSRITVEKTLKQHINELFDKNYKKKITVKKLKLTKWLIAQGYYKNDLRTPKDSLITYL DMKKIFGSSFMENNKNYNQIEELIEWLTIFEDKQIINEKLKSHSNLSKIILTCITTETNTPKSLSQLSNYSVLDLMWTTNNNFISIISNDKYDFKNYIEN HNLNKNEDQNISNLVNDIHVSPALKRGITQSIKIVQEIVKFMGHAPKYIFIEVTRETKKSRTTTSRGKRIQRLQSKLLNKANGFKPQLRKYLVPNEKIQEELKKHKNDISSERIMLYFL	

	QNGKSLYSEESLININKLSDYQVDHILPRTYIPDDSLENKALVLAKENQRKADDLLLNSNVIDKNLERWTYMLNNNMMGLKKFKNLTRRVITDKDKLGFIHRQLVQTSQMVKGVANILNS MYKNQGTTCIQARANLSTAFRKALSDQDDRYHFKHPELVKNRNINDFHHAQDAYLASFLGTYRLRRFPTDEMLLMNGEYNKFYGQVKELYSKKKKLPDSRKNGFIISPLVNGTTQYDRN TGEIIWNVGFRDKILKIFNYHQCNVTRKTEIKTGQFTDQTIYSPKNPKYKKLIAQKKDMDPNIYGGFSGDNKSSITIVKIDNNKIKPVAIPIRLINDLKDKKTLQNWLEENVKHKKSIQ IIKNNVPIGQIIYSKKVGLLSLNSDREIANRQQLLILPPEHSALLRILQIPDEDPDQIILAFYDKNILVEIILQELITKMKKFYPFYKNEQEFLASNIENFNQATTSEKINSLEELITLLHA NSTSAHLIFNNIEKKAFGRKTHGLTLNDTDFIYQSVTGLYETRIHIE	KGE60162.1 [Streptococcus pyogenes MGAS2111]
401	QELDINRLSGYDVDHIVPQSFLKDDSIDNKVLTRSDNRGKSDNVPSEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKVGFIKRQLVETRQITKHVAQILDSRMNTKYD ENDKLIREVRVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLI ETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKDPIDFL EAKGYKEVRKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKP IREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	CUO82355.1 [Roseburia hominis]
402	MSQNIVDYCIGLDLGTGSVGWAVVDMNHRLMKRNGKHLWGSRLFSNAETAATRRSSRSIRRRYNKRRERIRLLRAILQDMVLEKDPTFFIRLEHTSFLDEEDKAKYLGTDYKDNYNLFI DEDFNDYTYYHKYPTIYHLRKALCESTEKADPRLIYLALAHIIVKYRGNFLIEGQQKFNMDASNIEDKLSDVFTQFADFNNIPYEDDEKKNLEIILEKLPLSKKAKVDEVMALISPEKEF KSAYKELVTGIAGNKMNVTKMILCESIKQGDSEIKLKFSDSNYDDQFSEVENDLGEYVEFIDSLHNIYSWVELQTIMGATHTDNASISEAMVSRYNKHHEDLQLLKKCIKDNVPKKYFD MFRNDSEKVKGYINYINRPNKAPVDEFYKFVKKCIEKVDTPEAKQLLHDIELENFLLKQNSRTNGSVPYQMQLDEMIKIIDNQAKYYPVLKEKRRFQLLSILTFRIPYYFGPLNETSEHA WIKRLEGKENQRILPWNYQDTVDVDATAEGFIKRMRSYCTYFPDEEVLPKNSLIVSKYEVYNELNKIRVDDKLLEVDVKNDIYNELFMKNKTVTEKKLKNWLVNNQCCRKDAEIKGFQK ENQFSTSLTPWIDFTNIFGKIDQSNFDLIEKIIYDLTVFEDKKIMKRRLKKKYALPDDKIKQILKLKYKDWSRLSKKLLDGIVADNRFGSSVTVLDVLEMSRLNLMEIINDKELGYAQM IEEASSCPKDGKFTYEEVAKLAGSPALKRGIWQSLQIVEEITKVMKCRPKYIYIEFERSREAKERTESKIKKLENVYKDLDEQTKKEYKSVLEELKGFDNTKKISSDSLFLYFTQLGKC MYSGKKLDIDSLDKYQIDHIVPQSLVKDDSFDNRVLVLPSENQRKLDDLVVPFDIRDKMYKFFWKLLFDHELISPKKFYSLIKTEYTERDEERFINRQLVETRQITKNVTQILEDHYSTT KVAAIRANLSHEFRVKNHIYKNRDINDYHHAHDAYIVALIGGFMRDRYPNMHDSKAVYSEYMKMFRKNKNDQKKWKDGFVINSMNYPYEVDGKLIWNPDLINEIKKCFYYKDCYCTTKL DQKSGQMFNLTVLPNDAHSAKGTTEAVIPVNKNRKDVNKYGGFSGLQYVIAAIEGTKKKGKKLVKVRKLSGIPLYLKQADIKRQIEYVEKEKELSDVKIIKNNIPLNQLIEIDGRQYLLL TSPTECVNAMQLVLNEEQCKLIADIYNAIYKQDFDGLDNMLMIQLYLQLIDKLKTLYPIYMGIVEKFEKLTEDFVSISKEEKANVIKQMLIIMHKGPQNGNIIYDDFNVGKRIGRLNGR TFYLDNIEFISQSPTGIYTKKYKL	Cas9 [Geobacillus thermodenitrifica ns]
403	MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRLRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARILL HLAKRRGFRSNRKSERTNKENSTMLKHIEENQSILSSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVARDDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVG FCTFEPKEKRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHDVRTLLNLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVYGKGAAKS FRPIDFDTFGYALTMFKDDTDIRSYLRNEYEQNGKRMENLADKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQKTVLLPNIPPIANPVVMRA LTQARKVVNAIIKKYGSPVSIHIELARELSQSFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVDHVIPYSRSLDDSYT NKVLVLTKENREKGNRTPAEYLGLGSERWQQFETFVLTNKQFSKKKRDRLLRLHYDENENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVNGRITAHLRSRWNFNKNRE ESNLHHAVDAAIVACTTPSDIARVTAFYQRREQNKELSKKTDPQFPQPWPHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMPKKRSITGAAHQETLRRYIGIDERSGKI QTVVKKLSEIQLDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGELGPIIRTIKIDTTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTIDMMKGI LPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIDSSNGGLSLVSHDNNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRGEKRVGVASSS HSKAGETIRPL	
404	KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHN VNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTY FPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIY QSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIE LAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSD SKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIF KEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDP QTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKL KKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	
405	(His-X-Glu-X23-26-Pro-Cys-X2-4-Cys	Zn2+-coordinating motif

EQUIVALENTS AND SCOPE

[00549]Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents of the embodiments described herein. The scope of the present disclosure is not intended to be limited to the above description, but rather is as set forth in the appended claims.

[00550]Articles such as "a," "an," and "the" may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include "or" between two or more members of a group are considered satisfied if one, more than one, or all of the group members are present, unless indicated to the contrary or otherwise evident from the context. The disclosure of a group that includes "or" between two or more group members provides embodiments in which exactly one member of the group is present, embodiments in which more than one members of the group are present, and embodiments in which all of the group members are present. For purposes of brevity those embodiments have not been individually spelled out herein, but it will be understood that each of these embodiments is provided herein and may be specifically claimed or disclaimed.

[00551]It is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitation, element, clause, or descriptive term, from one or more of the claims or from one or more relevant portion of the description, is introduced into another claim. For example, a claim that is dependent on another claim can be modified to include one or more of the limitations found in any other claim that is dependent on the same base claim. Furthermore, where the claims recite a composition, it is to be understood that methods of making or using the composition according to any of the methods of making or using disclosed herein or according to methods known in the art, if any, are included, unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise.

[00552]Where elements are presented as lists, *e.g.*, in Markush group format, it is to be understood

that every possible subgroup of the elements is also disclosed, and that any element or subgroup of elements can be removed from the group. It is also noted that the term "comprising" is intended to be open and permits the inclusion of additional elements or steps. It should be understood that, in general, where an embodiment, product, or method is referred to as comprising particular elements, features, or steps, embodiments, products, or methods that consist, or consist essentially of, such elements, features, or steps, are provided as well. For purposes of brevity those embodiments have not been individually spelled out herein, but it will be understood that each of these embodiments is provided herein and may be specifically claimed or disclaimed.

[00553]Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value within the stated ranges in some embodiments, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise. For purposes of brevity, the values in each range have not been individually spelled out herein, but it will be understood that each of these values is provided herein and may be specifically claimed or disclaimed. It is also to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values expressed as ranges can assume any subrange within the given range, wherein the endpoints of the subrange are expressed to the same degree of accuracy as the tenth of the unit of the lower limit of the range.

[00554]In addition, it is to be understood that any particular embodiment of the present invention may be explicitly excluded from any one or more of the claims. Where ranges are given, any value within the range may explicitly be excluded from any one or more of the claims. Any embodiment, element, feature, application, or aspect of the compositions and/or methods of the invention, can be excluded from any one or more claims. For purposes of brevity, all of the embodiments in which one or more elements, features, purposes, or aspects is excluded are not set forth explicitly herein.

CLAIMS

What is claimed is:

1. A Cas9 protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of a Cas9 as provided by any one of SEQ ID NO: 2, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations in an amino acid residue selected from the group consisting of amino acid residues 10, 177, 218, 322, 367, 409, 427, 589, 599, 614, 630, 631, 654, 673, 693, 710, 715, 727, 743, 753, 757, 758, 762, 763, 768, 803, 859, 861, 865, 869, 921, 946, 1016, 1021, 1028, 1054, 1077, 1080, 1114, 1134, 1135, 1137, 1139, 1151, 1180, 1188, 1211, 1219, 1221, 1223, 1256, 1264, 1274, 1290, 1318, 1317, 1320, 1323, and 1333 of the amino acid sequence provided in SEQ ID NO: 2.

2. The Cas9 protein of claim 1, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations selected from the group consisting of X10T, X177N, X218R, X322V, X367T, X409I, X427G, X589S, X599R, X614N, X630K, X631A, X654L, X673E, X693L, X710E, X715C, X727I, X743I, X753G, X757K, X758H, X762G, X763I, X768H, X803S, X859S, X861N, X865G, X869S, X921P, X946D, X1016D, X1021T, X1028D, X1054D, X1077D, X1080S, X1114G, X1134L, X1135N, X1137S, X1139A, X1151E, X1180G, X1188R, X1211R, X1219V, X1221H, X1223S, X1256R, X1264Y, X1274R, X1290G, X1318S, X1317T, X1320V, X1323D, and X1333K of the amino acid sequence provided in SEQ ID NO: 2, wherein X represents any amino acid.

3. The Cas9 protein of claim 1 or 2, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations selected from the group consisting of A10T, D177N, K218R,

I322V, A367T, S409I, E427G, A589S, K599R, D614N, E630K, M631A, R654L, K673E, F693L, K710E, G715C, L727I, V743I, R753G, E757K, N758H, E762G, M763I, Q768H, N803S, R859S, D861N, G865G, N869S, L921P, N946D, Y1016D, M1021T, E1028D, N1054D, G1077D, F1080S, R1114G, F1134L, D1135N, P1137S, V1139A, K1151E, D1180G, K1188R, K1211R, E1219V, Q1221H, G1223S, Q1256R, H1264Y, S1274R, V1290G, L1318S, N1317T, A1320V, A1323D, and R1333K of the amino acid sequence provided in SEQ ID NO:2, wherein X is any amino acid.

4. A Cas9 protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of a Cas9 as provided by SEQ ID NO: 2, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations in an amino acid residue selected from the group consisting of amino acid residues 10, 177, 218, 322, 367, 427, 589, 599, 614, 630, 631, 693, 710, 743, 753, 757, 758, 762, 768, 803, 859, 861, 865, 869, 921, 946, 1016, 1021, 1028, 1054, 1077, 1080, 1114, 1134, 1135, 1137, 1151, 1180, 1188, 1211, 1221, 1223, 1274, 1290, 1317, 1320, 1323, and 1333 of the amino acid sequence provided in SEQ ID NO: 2.

5. The Cas9 protein of claim 4, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations selected from the group consisting of X10T, X177N, X218R, X322V, X367T, X427G, X589S, X599R, X614N, X630K, X631A, X693L, X710E, X743I, X753G, X757K, X758H, X762G, X768H, X803S, X859S, X861N, X865G, X869S, X921P, X946D, X1016D, X1021T, X1028D, X1054D, X1077D, X1080S, X1114G, X1134L, X1135N, X1137S, X1151E, X1180G, X1188R, X1211R, X1221H, X1223S, X1274R, X1290G, X1317T, X1320V, X1323D, and X1333K of the amino acid sequence provided in SEQ ID NO: 2, wherein X represents any amino acid.

6. The Cas9 protein of claim 4 or 5, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations selected from the group consisting of A10T, D177N, K218R, I322V, A367T, E427G, A589S, K599R, D614N, E630K, M631A, F693L, K710E, V743I, R753G, E757K, N758H, E762G, Q768H, N803S, R859S, D861N, N869S, L921P, N946D, Y1016D, M1021T, E1028D, N1054D, G1077D, F1080S, R1114G, F1134L, D1135N, P1137S, K1151E, D1180G, K1188R, K1211R, Q1221H, G1223S, S1274R, V1290G, N1317T, A1320V, A1323D, and R1333K of the amino acid sequence provided in SEQ ID NO: 2, wherein X represents any amino acid.

7. The Cas9 protein of any one of claims 1-6, wherein the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones listed in Table 1, or a combination of conservative mutations thereto.

8. The Cas9 protein of any one of claims 1-7, wherein the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones listed in Table 1.

9. The Cas9 protein of any one of claims 1-8, wherein the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones selected from the group consisting of N3.19.4c-3; N3.19.4c-4; P4.2-72-4; P4.2-72-5; P10.6.144.2; P10.5.192.7; P10.5.192.10; P10.6.144.5; P10.6.192.1; P10.6.192.9; P10.6.192.12; P13.2-8; P13.3-3; P13.4-3; P16.2-120-1; P16.2-120-2; P16.2-120-3; P16.2-120-4; P16.2-120-5; P16.2-120-6; P16.1-3; P16.3-2; P16.4-5(es); and P16.6-2, or a combination of conservative mutations thereto.

10. The Cas9 protein of any one of claims 1-9, wherein the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones selected from the group consisting of N3.19.4c-3; N3.19.4c-4; P4.2-72-4; P4.2-72-5; P10.6.144.2; P10.5.192.7; P10.5.192.10; P10.6.144.5; P10.6.192.1; P10.6.192.9; P10.6.192.12; P13.2-8; P13.3-3; P13.4-3; P16.2-120-1; P16.2-120-2; P16.2-120-3; P16.2-120-4; P16.2-120-5; P16.2-120-6; P16.1-3; P16.3-2; P16.4-5(es); and P16.6-2.

11. The Cas9 protein of any one of claims clim 1-10 comprising an amino acid sequence that is at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence of a Cas9 as provided by SEQ ID NOs: 2.

12. The Cas9 protein of any one of claims 1-11, wherein the Cas9 exhibits an increased activity on a target sequence that does not comprise the canonical PAM (5′-NGG-3′) at its 3′ end as compared to *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2.

13. The Cas9 protein of any one of claims 1-12, wherein the Cas9 protein exhibits an activity on a target sequence having a 3′ end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 5-fold, at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold increased as compared to the activity of *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2 on the same target sequence.

14. The Cas9 protein of claim 12 or 13, wherein the 3′ end of the target sequence is directly adjacent to an AAA, GAA, CAA, or TAA sequence.

15. The Cas9 protein of any one of claims 12-14, wherein the activity is measured by a nuclease assay, a deamination assay, or a transcriptional activation assay.

16. The Cas9 protein of any one of claims 1-15, wherein the Cas9 protein comprises a D10A and/or a H840A mutation of the amino acid sequence provided in SEQ ID NO: 2.

17. A Cas9 protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of a Cas9 as provided by SEQ ID NO: 2, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations in an amino acid residue selected from the group consisting of amino acid residues 472, 562, 565, 570, 570, 589, 608, 625, 627, 629, 630, 631, 638, 647, 652, 653, 654, 670, 673, 676, 687, 703, 710, 711, 716, 740, 742, 752, 753, 767, 771, 775, 789, 790, 795, 797, 803, 804, 808, 848, 866, 875, 890, 922, 928, 948, 959, 990, 995, 1014, 1015, 1016, 1021, 1030, 1036, 1055, 1057, 1114, 1127, 1135, 1156, 1177, 1180, 1184, 1207, 1219, 1234, 1246, 1251, 1252, 1286, 1301, 1332, 1335, 1337, 1338, 1348, 1349, 1365, 1367, and 1368 of the amino acid sequence provided in SEQ ID NO: 2.

18. The Cas9 protein of claim 1, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations selected from the group consisting of X472I, X562F, X565D, X570T, X570S, X589V, X608R, X625S, X627K, X629G, X630G, X631I, X631V, X638P, X647A, X647I, X652T, X653K, X654L, X654I, X654H, X670T, X673E, X676G, X687R, X703P, X710E, X711T, X716R, X740A, X742E, X752R, X753G, X767D, X771H, X775R, X789E, X790A, X795L, X797N, X803S, X804A, X808D, X848N, X866R, X875I, X890E, X890N, X922A, X928T, X948E, X959N, X990S, X995S, X1014N, X1015A, X1016C, X1016S, X1021L, X1030R, X1036H, X1036D,

X1055E, X1057S, X1057T, X1114G, X1127A, X1127G, X1135N, X1156E, X1156N, X1177S, X1180E, X1184T, X1207G, X1219V, X1234D, X1246E, X1251G, X1252D, X1286H, X1301S, X1332N, X1332G, X1335Q, X1337N, X1338T, X1348V, X1349R, X1365L, X1367E, X1367T, X1367fs?, and X1368D of the amino acid sequence provided in SEQ ID NO: 2, wherein X represents any amino acid.

19. The Cas9 protein of claim 17 or 18, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations selected from the group consisting of T472I, I562F, V565D, I570T, I570S, A589V, K608R, L625S, E627K, R629G, E630G, M631I, M631V, T638P, V647A, V647I, K652T, R653K, R654L, R654I, R654H, I670T, K673E, G676G, G687R, T703P, K710E, A711T, Q716R, T740A, K742E, G752R, R753G, N767D, Q771H, K775R, K789E, E790A, I795L, K797N, N803S, T804A, N808D, K848N, K866R, V875I, K890E, K890N, V922A, K948E, K959N, N990S, T995S, K1014N, V1015A, Y1016C, Y1016S, M1021L, G1030R, Y1036H, Y1036D, I1055E, I1057S, I1057T, R1114G, D1127A, D1127G, D1135N, K1156E, K1156N, N1177S, D1180E, A1184T, E1207G, E1219V, N1234D, K1246E, D1251G, N1252D, N1286H, P1301S, D1332N, D1332G, R1335Q, T1337N, S1338T, I1348V, H1349R, L1365L, G1367E, G1367T, G1367fs?, and D1368D of the amino acid sequence provided in SEQ ID NO: 2, wherein X is any amino acid.

20. A Cas9 protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of a Cas9 as provided by any one of SEQ ID NO: 2, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations in an amino acid residue selected from the group consisting of amino acid residues 472, 562, 565, 570, 570, 589, 608, 625, 627, 629, 630, 631, 638, 647, 647, 652, 653, 654, 654, 654, 670, 676, 687, 703, 710, 716, 740, 742,

752, 753, 767, 771, 775, 789, 790, 795, 797, 803, 804, 808, 848, 866, 875, 890, 890, 922, 948, 959, 990, 995, 1014, 1015, 1016, 1016, 1021, 1030, 1036, 1036, 1055, 1057, 1057, 1114, 1127, 1135, 1156, 1156, 1177, 1180, 1184, 1234, 1246, 1251, 1252, 1286, 1301, 1332, 1332, 1335, 1338, 1348, 1349, 1367, 1367, 1367, and 1368 of the amino acid sequence provided in SEQ ID NO: 2.

21. The Cas9 protein of claim 20, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations selected from the group consisting of X472I, X562F, X565D, X570T, X570S, X589V, X608R, X625S, X627K, X629G, X630G, X631I, X631V, X638P, X647A, X647I, X652T, X653K, X654L, X654I, X654H, X670T, X676G, X687R, X703P, X710E, X716R, X740A, X742E, X752R, X753G, X767D, X771H, X775R, X789E, X790A, X795L, X797N, X803S, X804A, X808D, X848N, X866R, X875I, X890E, X890N, X922A, X948E, X959N, X990S, X995S, X1014N, X1015A, X1016C, X1016S, X1021L, X1030R, X1036H, X1036D, X1055E, X1057S, X1057T, X1114G, X1127A, X1127G, X1135N, X1156E, X1156N, X1177S, X1180E, X1184T, X1234D, X1246E, X1251G, X1252D, X1286H, X1301S, X1332N, X1332G, X1335Q, X1338T, X1348V, X1349R, X1367E, X1367T, X1367fs?, and X1368D of the amino acid sequence provided in SEQ ID NO: 2, wherein X represents any amino acid.

22. The Cas9 protein of claim 20 or 21, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations selected from the group consisting of T472I, I562F, V565D, I570T, I570S, A589V, K608R, L625S, E627K, R629G, E630G, M631I, M631V, T638P, V647A, V647I, K652T, R653K, R654L, R654I, R654H, I670T, G676G, G687R, T703P, K710E, Q716R, T740A, K742E, G752R, R753G, N767D, Q771H, K775R, K789E, E790A, I795L, K797N, N803S, T804A, N808D, K848N, K866R, V875I, K890E, K890N, V922A, K948E, K959N, N990S, T995S,

K1014N, V1015A, Y1016C, Y1016S, M1021L, G1030R, Y1036H, Y1036D, I1055E, I1057S, I1057T, R1114G, D1127A, D1127G, D1135N, K1156E, K1156N, N1177S, D1180E, A1184T, N1234D, K1246E, D1251G, N1252D, N1286H, P1301S, D1332N, D1332G, R1335Q, S1338T, I1348V, S1349R, G1367E, G1367T, G1367fs?, and D1368D of the amino acid sequence provided in SEQ ID NO: 2, wherein X represents any amino acid.

23. The Cas9 protein of any one of claims 17-22, wherein the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones listed in Table 2, or a combination of conservative mutations thereto.

24. The Cas9 protein of any one of claims 17-23, wherein the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones listed in Table 2.

25. The Cas9 protein of any one of claims 17-24, wherein the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones selected from the group consisting of N4.CAC-1; N4.CAC-5; N4.CAC06; SacB.CAC.4h; N3.CAC-1; N3.CAC-5; N3.CAC-6; N3.CAC-8; P15.1.166-3; P15.1.166-8; P15.2.166-2; P15.3.166-4; P15.3.166-5; P15.3.166-7; P15.4.166-4; P15.4.166-8; P17.1.144-1; P17.1.144-2; P17.1.144-3; P17.1.144-4; P17.1.144-5; P17.1.144-7; P17.1.144-8; P17.2.144-1; P17.2.144-2; P17.2.144-3; P17.2.144-4; P17.2.144-5; P17.2.144-6; P17.2.144-7; P17.2.144-8; P17.1-1; P17.1-5; and P17.1.7-4(fn), or a combination of conservative mutations thereto.

26. The Cas9 protein of any one of claims 17-25, wherein the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones selected from the group consisting of N4.CAC-1; N4.CAC-5; N4.CAC06; SacB.CAC.4h; N3.CAC-1; N3.CAC-5; N3.CAC-6;

N3.CAC-8; P15.1.166-3; P15.1.166-8; P15.2.166-2; P15.3.166-4; P15.3.166-5; P15.3.166-7; P15.4.166-4; P15.4.166-8; P17.1.144-1; P17.1.144-2; P17.1.144-3; P17.1.144-4; P17.1.144-5; P17.1.144-7; P17.1.144-8; P17.2.144-1; P17.2.144-2; P17.2.144-3; P17.2.144-4; P17.2.144-5; P17.2.144-6; P17.2.144-7; P17.2.144-8; P17.1-1; P17.1-5; and P17.1.7-4(fn).

27. The Cas9 protein of any one of claims claim 17-26 comprising an amino acid sequence that is at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence of a Cas9 as provided by SEQ ID NO: 2.

28. The Cas9 protein of any one of claims 17-27, wherein the Cas9 exhibits an increased activity on a target sequence that does not comprise the canonical PAM (5′-NGG-3′) at its 3′ end as compared to *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2.

29. The Cas9 protein of any one of claims 17-28, wherein the Cas9 protein exhibits an activity on a target sequence having a 3′ end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 5-fold, at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold increased as compared to the activity of *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2 on the same target sequence.

30. The Cas9 protein of claim 28 or 29, wherein the 3′ end of the target sequence is directly adjacent to an AAC, GAC, CAC, or TAC sequence.

31. The Cas9 protein of any one of claims 28-30, wherein the activity is measured by a nuclease assay, a deamination assay, or a transcriptional activation assay.

32. The Cas9 protein of any one of claims 17-31, wherein the Cas9 protein comprises a D10A and/or a H840A mutation of the amino acid sequence provided in SEQ ID NO: 2.

33. A Cas9 protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of a Cas9 as provided by SEQ ID NOs: 2, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations in an amino acid residue selected from the group consisting of amino acid residues 575, 596, 631, 649, 654, 664, 710, 740, 743, 748, 750, 753, 765, 790, 797, 853, 922, 955, 961, 985, 1012, 1049, 1057, 1114, 1131, 1135, 1150, 1156, 1162, 1180, 1191, 1218, 1219, 1221, 1227, 1249, 1253, 1256, 1286, 1293, 1308, 1317, 1320, 1321, 1332, 1335, and1339 of the amino acid sequence provided in SEQ ID NO: 2.

34. The Cas9 protein of claim 33, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations selected from the group consisting of X575S, X596Y, X631L, X649R, X654L, X664K, X710E, X740A, X743I, X748I, X750A, X753G, X765X, X790A, X797E, X853E, X922A, X955L, X961E, X985Y, X1012A, X1049G, X1057V, X1114G, X1131C, X1135N, X1150V, X1156E, X1162A, X1180G, X1180A, X1191N, X1218S, X1219V, X1221H, X1227V, X1249S, X1253K, X1256R, X1286K, X1293T, X1308D, X1317K, X1320V, X1321S, X1332G, X1335L, and X1339I of the amino acid sequence provided in SEQ ID NO: 2 wherein X represents any amino acid.

35. The Cas9 protein of claim 33 or 34, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations selected from the group consisting of F575S, D596Y, M631L, K649R, R654L, R664K, K710E, T740A, V743I, V748I, V750A, R753G, R765X, E790A, K797E, D853E, V922A, V955L, K961E, H985Y, D1012A, E1049G, I1057V, R1114G, Y1131C, D1135N, E1150V, K1156E, E1162A, D1180G, D1180A, K1191N, G1218S, E1219V, Q1221H, A1227V, P1249S, E1253K, Q1256R, N1286K, A1293T, N1308D, N1317K, A1320V, P1321S, D1332G, R1335L, and T1339I of the amino acid sequence provided in SEQ ID NO: 2 wherein X is any amino acid.

36. A Cas9 protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of a Cas9 as provided by SEQ ID NO: 2 werein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations in an amino acid residue selected from the group consisting of amino acid residues 575, 596, 631, 649, 664, 710, 740, 743, 748, 750, 753, 765, 790, 797, 853, 922, 961, 985, 1012, 1049, 1057, 1114, 1131, 1135, 1150, 1156, 1162, 1180, 1191, 1218, 1221, 1249, 1253, 1286, 1293, 1308, 1317, 1320, 1321, 1332, 1335, and1339 of the amino acid sequence provided in SEQ ID NO: 2.

37. The Cas9 protein of claim 36, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations selected from the group consisting of X575S , X596Y, X631L, X649R, X664K, X710E, X740A, X743I, X748I, X750A, X753G, X765X, X790A, X797E, X853E, X922A, X961E, X985Y, X1012A, X1049G, X1057V, X1114G, X1131C, X1135N, X1150V, X1156E, X1162A, X1180G, X1180A, X1191N, X1218S, X1221H, X1249S, X1253K, X1286K,

X1293T, X1308D, X1317K, X1320V, X1321S, X1332G, X1335L, and X1339I of the amino acid sequence provided in SEQ ID NO: 2, or in a corresponding mutation, or mutations, in any of the amino acid sequences provided in SEQ ID NO: 2 wherein X represents any amino acid.

38. The Cas9 protein of claim 36 or 37, wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations selected from the group consisting of F575S, D596Y, M631L, K649R, R664K, K710E, T740A, V743I, V748I, V750A, R753G, R765X, E790A, K797E, D853E, V922A, K961E, H985Y, D1012A, E1049G, I1057V, R1114G, Y1131C, D1135N, E1150V, K1156E, E1162A, D1180G, D1180A, K1191N, G1218S, Q1221H, P1249S, E1253K, N1286K, A1293T, N1308D, N1317K, A1320V, P1321S, D1332G, R1335L, and T1339I of the amino acid sequence provided in SEQ ID NO: 2 wherein X represents any amino acid.

39. The Cas9 protein of any one of claims 33-38, wherein the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones listed in Table 3 or a combination of conservative mutations thereto.

40. The Cas9 protein of any one of claims 33-39, wherein the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones listed in Table 3

41. The Cas9 protein of any one of claims 33-40, wherein the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones selected from the group consisting of SacB.N4.19.TAT-4h-1; SacB.N4.19.TAT-4h-3; P12.2.b9-8; P12.3.b9-8; P12.3.b9-8 (ax); P12.3.b10-6; SacB.P12a2.AAT.3hr.maj; SacB.P12a2.AAT.3hr.min; P17.4-1; P17.4-2;

P17.4-3; P17.4-4; P17.4-5; P17.4-6; P17.4-8; P17-4-1-1; P17-4-3-1; and P17-4-6-1, or a combination of conservative mutations thereto.

42. The Cas9 protein of any one of claims 33-41, wherein the Cas9 protein comprises the combination of mutations of any one of the Cas9 clones selected from the group consisting of SacB.N4.19.TAT-4h-1; SacB.N4.19.TAT-4h-3; P12.2.b9-8; P12.3.b9-8; P12.3.b9-8 (ax); P12.3.b10-6; SacB.P12a2.AAT.3hr.maj; SacB.P12a2.AAT.3hr.min; P17.4-1; P17.4-2; P17.4-3; P17.4-4; P17.4-5; P17.4-6; P17.4-8; P17-4-1-1; P17-4-3-1; and P17-4-6-1.

43. The Cas9 protein of any one of claims claim 33-42 comprising an amino acid sequence that is at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the amino acid sequence of a Cas9 as provided by SEQ ID NOs: 2

44. The Cas9 protein of any one of claims 33-43, wherein the Cas9 exhibits an increased activity on a target sequence that does not comprise the canonical PAM (5′-NGG-3′) at its 3′ end as compared to *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2.

45. The Cas9 protein of any one of claims 33-44, wherein the Cas9 protein exhibits an activity on a target sequence having a 3′ end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 5-fold, at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold increased as compared to the activity of *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2 n the same target sequence.

46. The Cas9 protein of claim 44 or 45, wherein the 3′ end of the target sequence is directly adjacent to an AAT, GAT, CAT, or TAT sequence.

47. The Cas9 protein of any one of claims 44-46, wherein the activity is measured by a nuclease assay, a deamination assay, or a transcriptional activation assay.

48. The Cas9 protein of any one of claims 33-47, wherein the Cas9 protein comprises a D10A and/or a H840A mutation of the amino acid sequence provided in SEQ ID NO: 2 or a corresponding mutation, or mutations, in another Cas9 amino sequence.

49. The Cas9 protein of any one of claims 1-48, wherein the Cas9 exhibits an increased activity on a target sequence comprising a PAM sequence selected from the group consisting of AAA, AAC, AAG, AAT, CAA, CAC, CAG, CAT, GAA, GAC, GAG, GAT, TAA, TAC, TAG, TAT, ACA, ACC, ACG, ACT, CCA, CCC, CCG, CCT, GCA, GCC, GCG, GCT, TCA, TCC, TCG, TCT, AGA, AGC, AGT, CGA, CGC, CGT, GGA, GGC, GGT, TGA, TGC, TGT, ATA, ATC, ATG, ATT, CTA, CTC, CTG, CTT, GTA, GTC, GTG, GTT, TTA, TTC, TTG, and TTT at its 3′ end as compared to *Streptococcus pyogenes* Cas9 as provided by SEQ ID NO: 2.

50. The Cas9 protein of any one of claims 1-49, wherein the Cas9 protein exhibits lower off-target activity as compared to an off-target activity of the *Streptococcus pyogenes* Cas9 domain as provided by SEQ ID NO: 2.

51. A fusion protein comprising (i) the Cas9 protein of any one of claims 1-50, and (ii) an effector domain.

52. The fusion protein of claim 51, wherein the effector domain is a domain that comprises nuclease activity, nickase activity, recombinase activity, deaminase activity, methyltransferase activity, methylase activity, acetylase activity, acetyltransferase activity, transcriptional activation activity, or transcriptional repression activity.

53. The fusion protein of claim 51 or 52, wherein the effector domain is a nucleic acid editing domain.

54. The fusion protein of claim 53, wherein the nucleic acid editing domain comprises a deaminase domain.

55. The fusion protein of claim 54, wherein the deaminase domain is a cytidine deaminase domain.

56. The fusion protein of claim 55, wherein the cytidine deaminase domain is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase.

57. The fusion protein of claim 55 or 56, wherein the cytidine deaminase domain is at least 80%, at least 85%, at least 90%, at least 92%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identical to the cytidine deaminase domain of any one of SEQ ID NOs: 27-61.

58. The fusion protein of claim 55 or 56, wherein the cytidine deaminase domain comprises the amino acid sequence of any one of SEQ ID NOs: 27-61.

59. The fusion protein of any one of claims 51-58, wherein the fusion protein further comprises a uracil glycosylase inhibitor (UGI) domain.

60. The fusion protein of claim 59, wherein the UGI domain comprises the amino acid sequence of SEQ ID NO: 115.

61. The fusion protein of any one of claims 51-60, wherein the fusion protein comprises the amino acid sequence of SEQ ID NO: 122.

62. The fusion protein of any one of claims 51-60, wherein the fusion protein comprises the amino acid sequence of SEQ ID NO: 123.

63. The fusion protein of any one of claims 51-62, wherein the fusion protein further comprises a second UGI domain.

64. The fusion protein of claim 63, wherein the fusion protein comprises the amino acid sequence of SEQ ID NO: 123.

65. The fusion protein of claim 63, wherein the fusion protein comprises the amino acid sequence of SEQ ID NO: 124.

66. The fusion protein of claim 54, wherein the deaminase domain is an adenosine deaminase domain.

67. The fusion protein of claim 66 further comprising a second adenosine deaminase domain.

68. The fusion protein of claim 67, wherein the first adenosine deaminase domain and the second adenosine deaminase domain comprises an ecTadA domain, or variant thereof.

69. The fusion protein of claim 68, wherein the first adenosine deaminase domain and the second adenosine deaminase domain comprise the amino acid sequence of any one of SEQ ID NOs: 62-84.

70. The fusion protein of claim 69, wherein the first adenosine deaminase comprises the amino acid sequence of SEQ ID NO: 62-84.

71. The fusion protein of claim 69, wherein the second adenosine deaminase comprises the amino acid sequence of SEQ ID NO: 62-84.

72. The fusion protein of any one of claims 66-71, wherein the fusion protein comprises the amino acid sequence of SEQ ID NO: 127.

73. The fusion protein of any one of claims 66-71, wherein the fusion protein comprises the amino acid sequence of SEQ ID NO: 128.

74. A complex comprising the fusion protein of any one of claims 51-73, and a guide RNA bound to the Cas9 protein.

75. The complex of claim 74, wherein the guide RNA is about 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence.

76. The complex of claim 75, wherein the 3′ end of the target sequence is directly adjacent to an AAA, AAC, AAG, AAT, CAA, CAC, CAG, CAT, GAA, GAC, GAG, GAT, TAA, TAC, TAG, TAT, ACA, ACC, ACG, ACT, CCA, CCC, CCG, CCT, GCA, GCC, GCG, GCT, TCA, TCC, TCG, TCT, AGA, AGC, AGT, CGA, CGC, CGT, GGA, GGC, GGT, TGA, TGC, TGT, ATA, ATC, ATG, ATT, CTA, CTC, CTG, CTT, GTA, GTC, GTG, GTT, TTA, TTC, TTG, or TTT sequence.

77. The complex of claim 75 or 76, wherein the 3′ end of the target sequence is directly adjacent to an AAA, GAA, CAA, or TAA sequence.

78. The complex of claim 75 or 76, wherein the 3′ end of the target sequence is directly adjacent to an AAC, GAC, CAC, or TAC sequence.

79. The complex of claim 75 or 76, wherein the 3′ end of the target sequence is directly adjacent to an AAT, GAT, CAT, or TAT sequence.

80. The complex of any one of claims 74-79, wherein the guide RNA is 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides long.

81. The complex of any one of claims 75-80, wherein the guide RNA comprises a sequence of 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 contiguous nucleotides that is complementary to a target sequence.

82. The complex of any one of claims 75-81, wherein the target sequence is a DNA sequence.

83. The complex of claim 82, wherein the target sequence is a sequence in the genome of a mammal.

84. The complex of claim 83, wherein the target sequence is a sequence in the genome of a human.

85. The complex of any one of claims 75-84, wherein the target sequence comprises a sequence associated with a disease or disorder.

86. The complex of claim 85, wherein the target sequence comprises a point mutation associated with a disease or disorder.

87. The complex of claim 86, wherein the complex edits a point mutation in the target sequence.

88. The complex of claim 87, wherein the point mutation is located between about 10 to about 20 nucleotides upstream of the PAM in the target sequence.

89. The complex of claim 87 or 88, wherein the target sequence comprises a T to C point mutation.

90. The complex of claim 89, wherein the complex deaminates the target C point mutation, and wherein the deamination results in a sequence that is not associated with a disease or disorder.

91. The complex of claim 90, wherein the target C point mutation is present in the DNA strand that is not complementary to the guide RNA.

92. The complex of claim 87 or 88, wherein the target sequence comprises a G to A point mutation.

93. The complex of claim 92, wherein the complex deaminates the target A point mutation, and wherein the deamination results in a sequence that is not associated with a disease or disorder.

94. The complex of claim 93, wherein the target A point mutation is present in the DNA strand that is not complementary to the guide RNA.

95. The complex of any one of claims 74-94, wherein the complex exhibits increased deamination efficiency of a point mutation in a target sequence that does not comprise the canonical PAM (5′-NGG-3′) at its 3′ end as compared to the deamination efficiency of a complex comprising Streptococcus pyogenes Cas9 protein as provided by SEQ ID NO: 2.

96. The complex of claim 95, wherein the complex exhibits increased deamination efficiency of a point mutation in a target sequence having a 3′ end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 2-fold, at least 3-fold, at least 4-fold, at least 5-fold, at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at

least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold increased as compared to the deamination efficiency of complex comprising the *Streptococcus pyogenes* Cas9 protein as provided by SEQ ID NO: 2 on the same target sequence.

97. The complex of any one of claims 90-96, wherein a deamination activity is measured using a deamination assay, PCR, or sequencing.

98. The complex of any one of claims 74-97, wherein the complex produces fewer indels in a target sequence that does not comprise the canonical PAM (5′-NGG-3′) at its 3′ end as compared to the amount of indels produced by a complex comprising *Streptococcus pyogenes* Cas9 protein as provided by SEQ ID NO: 2.

99. The complex of claim 98, wherein the complex produces fewer indels in a target sequence having a 3′ end that is not directly adjacent to the canonical PAM sequence (5′-NGG-3′) that is at least 2-fold, at least 3-fold, at least 4-fold, at least 5-fold, at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold lower as compared to the amount of indels produced by a complex comprising *Streptococcus pyogenes* Cas9 protein as provided by SEQ ID NO: 2 on the same target sequence.

100. The complex of any one of claims 98-99, wherein indels are measured using high-throughput sequencing.

101. The complex of any one of claims 74-100, wherein the complex exhibits a decreased off-target activity as compared to the off-target activity of a complex comprising *Streptococcus pyogenes* Cas9 protein as provided by SEQ ID NO: 2.

102. The complex of claim 101, wherein the off-target activity of the complex is at least 2-fold, at least 3-fold, at least 4-fold, at least 5-fold, at least 10-fold, at least 50-fold, at least 100-fold, at least 500-fold, at least 1,000-fold, at least 5,000-fold, at least 10,000-fold, at least 50,000-fold, at least 100,000-fold, at least 500,000-fold, or at least 1,000,000-fold decreased as compared to the off-target activity of a complex comprising *Streptococcus pyogenes* Cas9 protein as provided by SEQ ID NO: 2.

103. The complex of any one of claims 75-102, wherein the target sequence is in the genome of an organism.

104. The complex of claim 103, wherein the organism is a prokaryote.

105. The complex of claim 104, wherein the prokaryote is a bacterium.

106. The complex of claim 103, wherein the organism is a eukaryote.

107. The complex of claim 103, wherein the organism is a plant or fungus.

108. The complex of claim 103, wherein the organism is a vertebrate.

109. The complex of claim 108, wherein the vertebrate is a mammal.

110. The complex of claim 109, wherein the mammal is a human.

111. The complex of claim 103, wherein the organism is a cell.

112. The complex of claim 111, wherein the cell is a human cell.

113. A method comprising contacting a nucleic acid with the fusion protein of any one of claims 51-73, and with a guide RNA, wherein the guide RNA is about 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence.

114. A method comprising contacting a cell with the fusion protein of any one of claims 51-73, and with a guide RNA, wherein the guide RNA is about 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence.

115. A method comprising contacting a nucleic acid with the complex of any one of claims 74-112.

116. A method comprising contacting a cell with the complex of any one of claims 74-112.

117. The method of any one of claims 113-116, wherein the contacting is performed *in vitro*.

118. The method of any one of claims 114-116, wherein the contacting is performed *in vivo*.

119. A method comprising administering to a subject the fusion protein of any one of claims 51-73, and a guide RNA, wherein the guide RNA is about 15-100 nucleotides long and comprises a sequence of at least 10 contiguous nucleotides that is complementary to a target sequence.

120. A method comprising administering to a subject the complex of any one of claims 74-112.

121. The method of any one of claims 113-120, wherein the target sequence of the nucleic acid is a DNA sequence.

122. The method of any one of claims 113-121, wherein the 3′ end of the target sequence is not immediately adjacent to the canonical PAM sequence (5′-NGG-3′).

123. The method of claim 122, wherein the 3′ end of the target sequence is directly adjacent to a sequence selected from the group consisting of AAA, GAA, CAA, and TAA.

124. The complex of claim 122, wherein the 3′ end of the target sequence is directly adjacent to a sequence selected from the group consisting of AAC, GAC, CAC, and TAC.

125. The complex of claim 122, wherein the 3′ end of the target sequence is directly adjacent to a sequence selected from the group consisting of AAT, GAT, CAT, and TAT.

126. The method of any one of claims 113-125, wherein the target sequence comprises a sequence associated with a disease or disorder.

127. The method of claim 126, wherein the target DNA sequence comprises a point mutation associated with a disease or disorder.

128. The method of claim 127, wherein the activity of the fusion protein, or the activity of the complex, results in a correction of the point mutation.

129. The method of any one of claims 127-128, wherein the target DNA sequence comprises a T to C point mutation associated with a disease or disorder, and wherein the deamination of the mutant C base results in a sequence that is not associated with a disease or disorder.

130. The method of claim 129, wherein the target DNA sequence encodes a protein and wherein the point mutation is in a codon and results in a change in the amino acid encoded by the mutant codon as compared to the wild-type codon.

131. The method of claim 130, wherein the deamination of the mutant C results in a change of the amino acid encoded by the mutant codon.

132. The method of claim 131, wherein the deamination of the mutant C results in the codon encoding the wild-type amino acid.

133. The method of any one of claims 127-128, wherein the target DNA sequence comprises a G to A point mutation associated with a disease or disorder, and wherein the deamination of the mutant A base results in a sequence that is not associated with a disease or disorder.

134. The method of claim 133, wherein the target DNA sequence encodes a protein and wherein the point mutation is in a codon and results in a change in the amino acid encoded by the mutant codon as compared to the wild-type codon.

135. The method of claim 134, wherein the deamination of the mutant A results in a change of the amino acid encoded by the mutant codon.

136. The method of claim 135, wherein the deamination of the mutant A results in the codon encoding the wild-type amino acid.

137. The method of any one of claims 113-136, wherein the contacting is in vivo in a subject.

138. The method of claim 137, wherein the subject has or has been diagnosed with a disease or disorder.

139. The method of claim 137 or 138, wherein the disease or disorder is a proliferative disease, a genetic disease, a neoplastic disease, a metabolic disease, or a lysosomal storage disease.

140. A kit comprising a nucleic acid construct, comprising:

(a) a nucleic acid sequence encoding the fusion protein of any one of claims 51-73; and

(b) a heterologous promoter that drives expression of the sequence of (a).

141. A kit comprising a nucleic acid construct, comprising:

(a) a nucleic acid sequence encoding the complex of any one of claims 74-112; and

(b) a heterologous promoter that drives expression of the sequence of (a).

142. The kit of claim 140 further comprising an expression construct encoding a guide RNA backbone, wherein the construct comprises a cloning site positioned to allow the cloning of a nucleic acid sequence identical or complementary to a target sequence into the guide RNA backbone.

143. A polynucleotide encoding the fusion protein of any one of claims 51-73 or the complex of any one of claims 74-112.

144. A vector comprising a polynucleotide of claim 143.

145. The vector of claim 144, wherein the vector comprises a heterologous promoter driving expression of the polynucleotide encoding the fusion protein or the polynucleotide encoding the complex.

146. A method comprising contacting a cell with the vector of claim 144 or 145.

147. The method of claim 146, wherein the cell vector is transfected into the cell.

148. The method of claim 147, wherein the vector is transfected into the cell using electroporation, heat shock, or a composition comprising a cationic lipid.

149. A cell comprising the fusion protein of any one of claims 51-73, or a nucleic acid molecule encoding the fusion protein of any one of claims 51-73.

150. A cell comprising the complex of any one of claims 74-112, or a nucleic acid molecule encoding the complex of any one of claims 74-112.

151. A cell comprising the vector of claim 144 or 145.

152. An SpCas9 comprising the amino acid sequence of SEQ ID NO: 122, wherein the SpCas9 has a non-canonical PAM specificity.

153. An SpCas9 comprising the amino acid sequence of SEQ ID NO: 123, wherein the SpCas9 has a non-canonical PAM specificity.

154. An SpCas9 comprising the amino acid sequence of SEQ ID NO: 124, wherein the SpCas9 has a non-canonical PAM specificity.

155. A fusion protein comprising an SpCas9 of any of claims 152-154 and a cytidine deaminase.

156. The fusion protein of claim 155, wherein the cytidine deaminase comprises any one of SEQ ID NOs: 27-61.

157. A fusion protein comprising an SpCas9 of any of claims 152-154 and an adenosine deaminase.

158. The fusion protein of claim 155, wherein the adenosine deaminase comprises any one of SEQ ID NOs: 62-84.

159. A complex comprising a fusion protein of any one of claims 155-158 and a guide RNA.

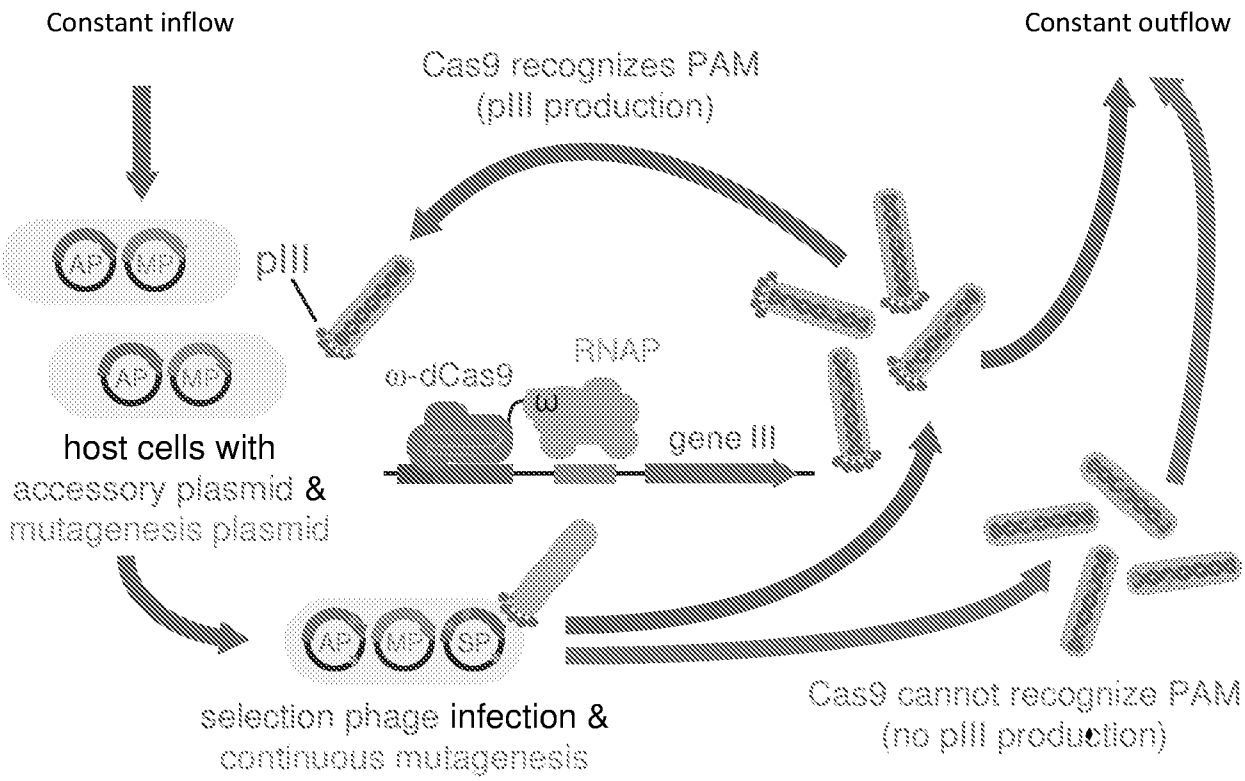

Constant inflow
Constant outflow
Cas9 recognizes PAM
(pIII production)
AP
MP
pIII
RNAP
ω-dCas9
ω
gene III
host cells with
accessory plasmid &
mutagenesis plasmid
AP
MP
SP
selection phage infection &
continuous mutagenesis
Cas9 cannot recognize PAM
(no pIII production)

FIG. 1A

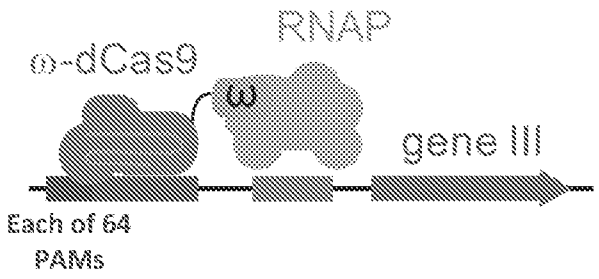

RNAP
ω-dCas9
ω
gene III
Each of 64
PAMs

FIG. 1B

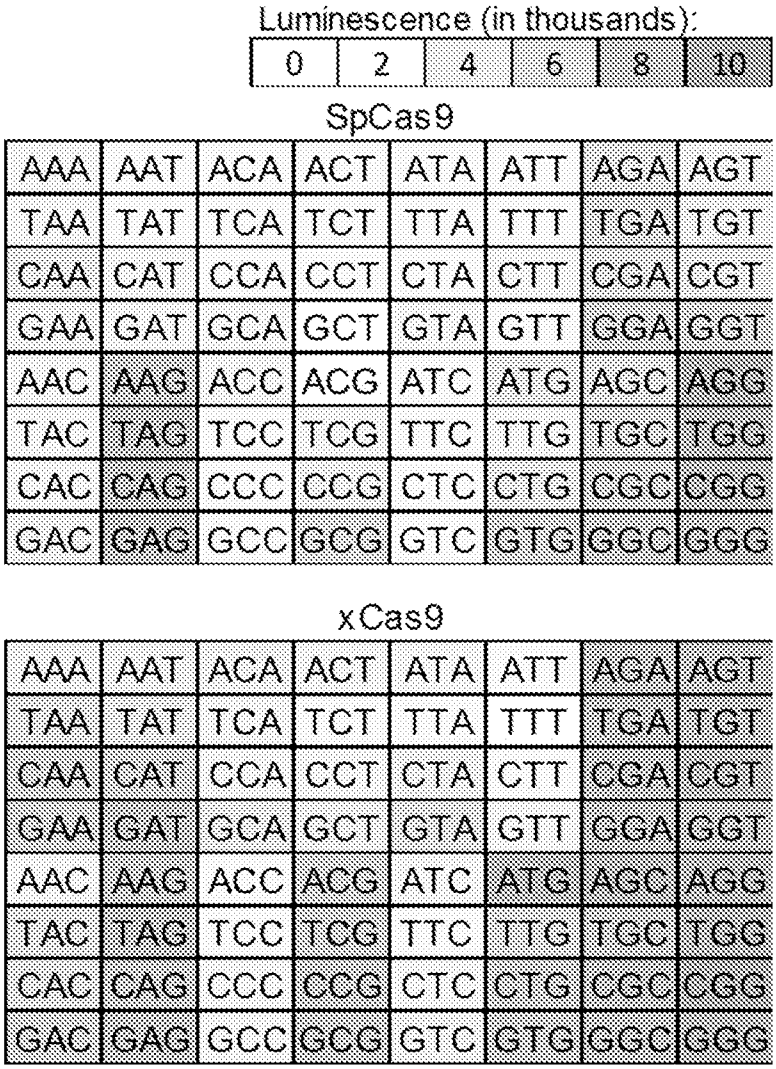

Luminescence (in thousands):
0
2
4
6
8
10
SpCas9
AAA AAT ACA ACT ATA ATT AGA AGT
TAA TAT TCA TCT TTA TTT TGA TGT
CAA CAT CCA CCT CTA CTT CGA CGT
GAA GAT GCA GCT GTA GTT GGA GGT
AAC AAG ACC ACG ATC ATG AGC AGG
TAC TAG TCC TCG TTC TTG TGC TGG
CAC CAG CCC CCG CTC CTG CGC CGG
GAC GAG GCC GCG GTC GTG GGC GGG
xCas9
AAA AAT ACA ACT ATA ATT AGA AGT
TAA TAT TCA TCT TTA TTT TGA TGT
CAA CAT CCA CCT CTA CTT CGA CGT
GAA GAT GCA GCT GTA GTT GGA GGT
AAC AAG ACC ACG ATC ATG AGC AGG
TAC TAG TCC TCG TTC TTG TGC TGG
CAC CAG CCC CCG CTC CTG CGC CGG
GAC GAG GCC GCG GTC GTG GGC GGG

FIG. 1C

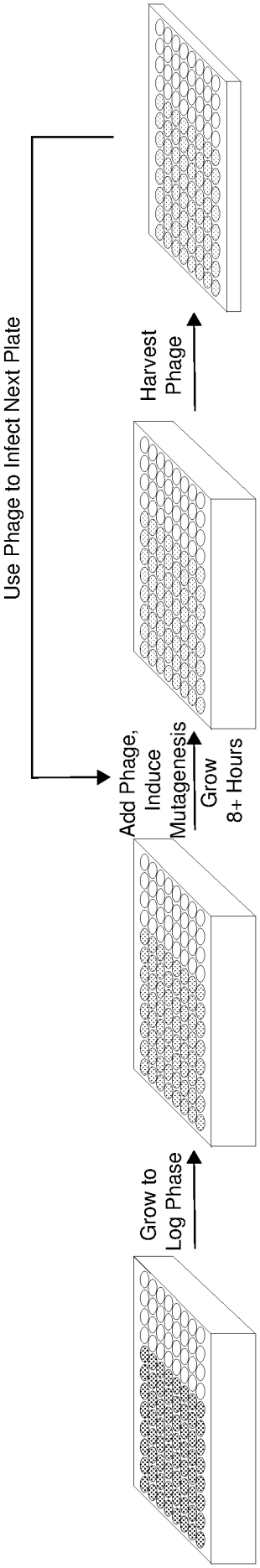

Grow to
Log Phase
Add Phage,
Induce
Mutagenesis
Grow
8+ Hours
Harvest
Phage
Use Phage to Infect Next Plate

FIG. 2A

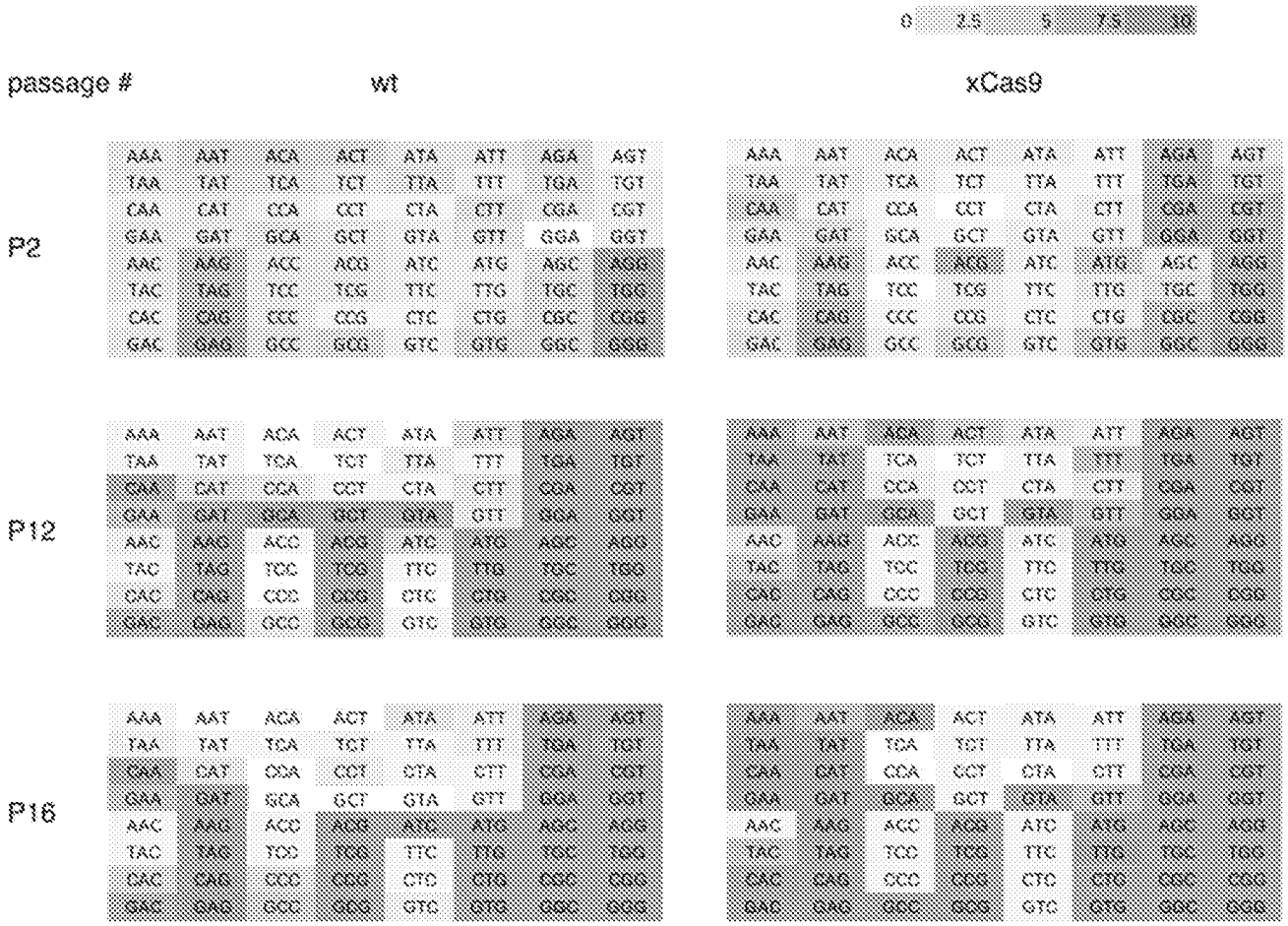


FIG. 2B

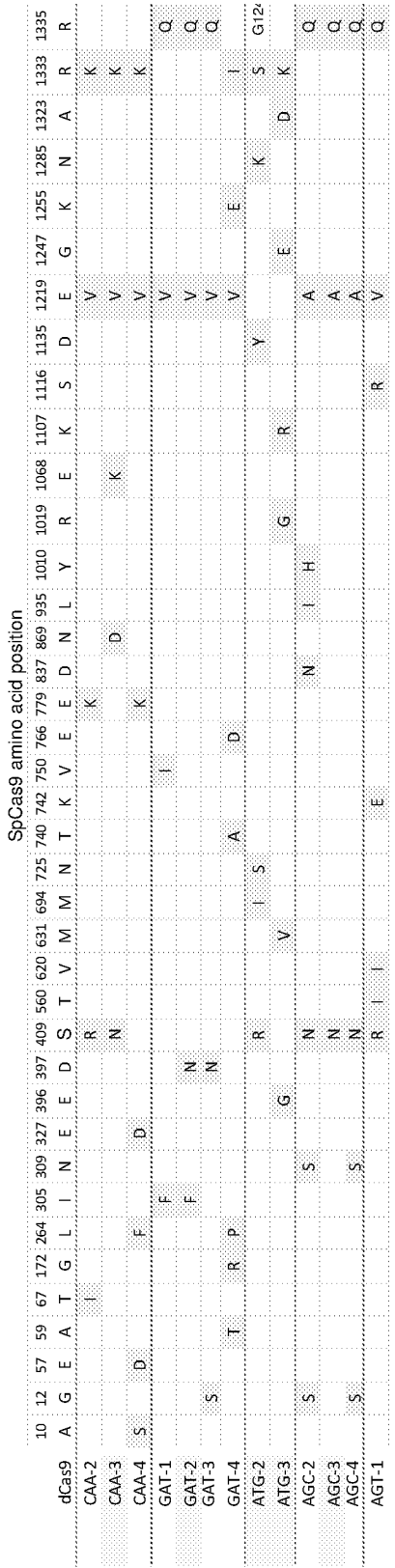

SpCas9 amino acid position
dCas9
CAA-2
CAA-3
CAA-4
GAT-1
GAT-2
GAT-3
GAT-4
ATG-2
ATG-3
AGC-2
AGC-3
AGC-4
AGT-1

FIG. 3A

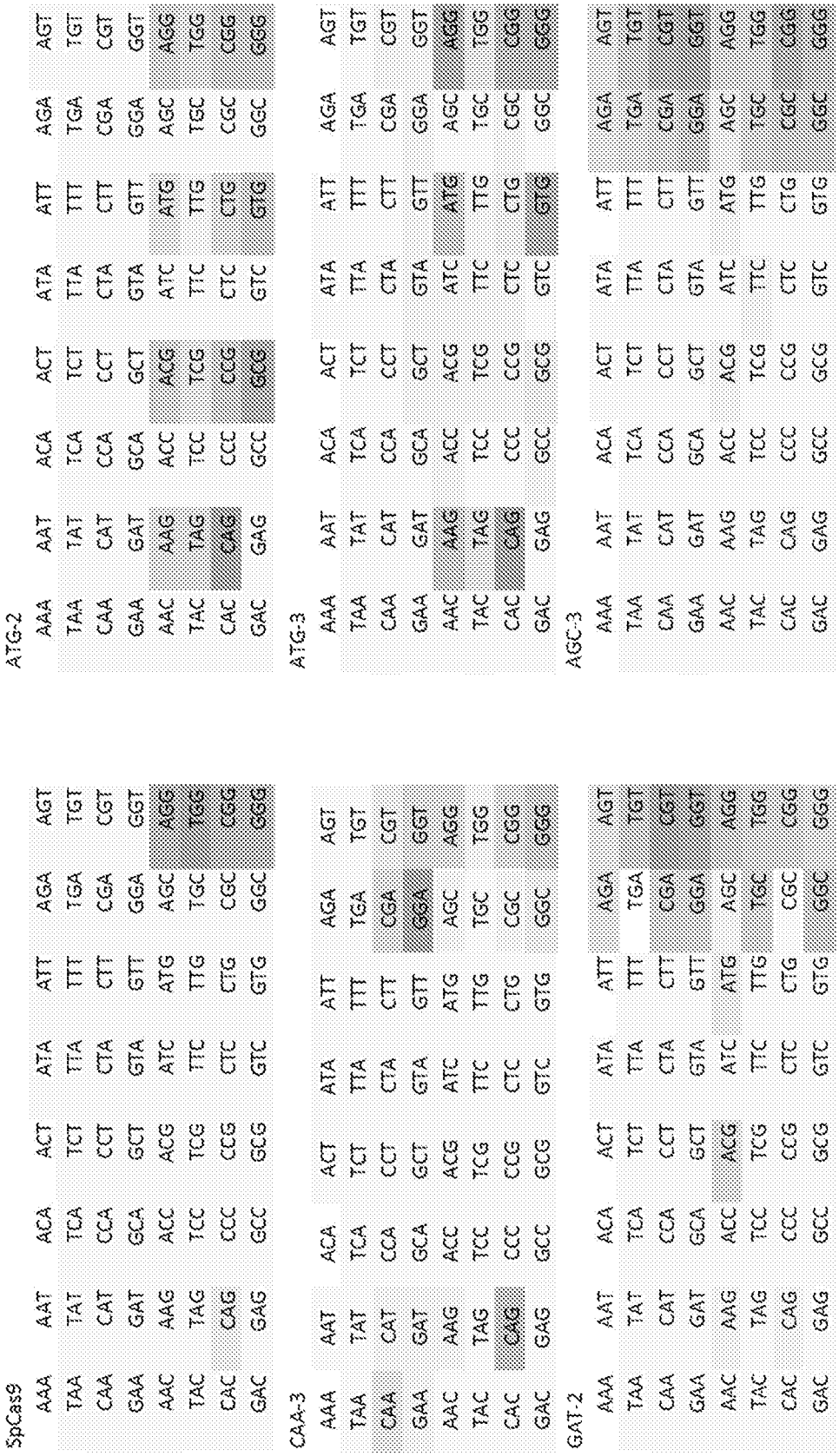

SpCas9
AAA AAT ACA ACT ATA ATT AGA AGT
TAA TAT TCA TCT TTA TTT TGA TGT
CAA CAT CCA CCT CTA CTT CGA CGT
GAA GAT GCA GCT GTA GTT GGA GGT
AAC AAG ACC ACG ATC ATG AGC AGG
TAC TAG TCC TCG TTC TTG TGC TGG
CAC CAG CCC CCG CTC CTG CGC CGG
GAC GAG GCC GCG GTC GTG GGC GGG
ATG-2
AAA AAT ACA ACT ATA ATT AGA AGT
TAA TAT TCA TCT TTA TTT TGA TGT
CAA CAT CCA CCT CTA CTT CGA CGT
GAA GAT GCA GCT GTA GTT GGA GGT
AAC AAG ACC ACG ATC ATG AGC AGG
TAC TAG TCC TCG TTC TTG TGC TGG
CAC CAG CCC CCG CTC CTG CGC CGG
GAC GAG GCC GCG GTC GTG GGC GGG
ATG-3
AAA AAT ACA ACT ATA ATT AGA AGT
TAA TAT TCA TCT TTA TTT TGA TGT
CAA CAT CCA CCT CTA CTT CGA CGT
GAA GAT GCA GCT GTA GTT GGA GGT
AAC AAG ACC ACG ATC ATG AGC AGG
TAC TAG TCC TCG TTC TTG TGC TGG
CAC CAG CCC CCG CTC CTG CGC CGG
GAC GAG GCC GCG GTC GTG GGC GGG
AGC-3
AAA AAT ACA ACT ATA ATT AGA AGT
TAA TAT TCA TCT TTA TTT TGA TGT
CAA CAT CCA CCT CTA CTT CGA CGT
GAA GAT GCA GCT GTA GTT GGA GGT
AAC AAG ACC ACG ATC ATG AGC AGG
TAC TAG TCC TCG TTC TTG TGC TGG
CAC CAG CCC CCG CTC CTG CGC CGG
GAC GAG GCC GCG GTC GTG GGC GGG
CAA-3
AAA AAT ACA ACT ATA ATT AGA AGT
TAA TAT TCA TCT TTA TTT TGA TGT
CAA CAT CCA CCT CTA CTT CGA CGT
GAA GAT GCA GCT GTA GTT GGA GGT
AAC AAG ACC ACG ATC ATG AGC AGG
TAC TAG TCC TCG TTC TTG TGC TGG
CAC CAG CCC CCG CTC CTG CGC CGG
GAC GAG GCC GCG GTC GTG GGC GGG
GAT-2
AAA AAT ACA ACT ATA ATT AGA AGT
TAA TAT TCA TCT TTA TTT TGA TGT
CAA CAT CCA CCT CTA CTT CGA CGT
GAA GAT GCA GCT GTA GTT GGA GGT
AAC AAG ACC ACG ATC ATG AGC AGG
TAC TAG TCC TCG TTC TTG TGC TGG
CAC CAG CCC CCG CTC CTG CGC CGG
GAC GAG GCC GCG GTC GTG GGC GGG

FIG. 3B

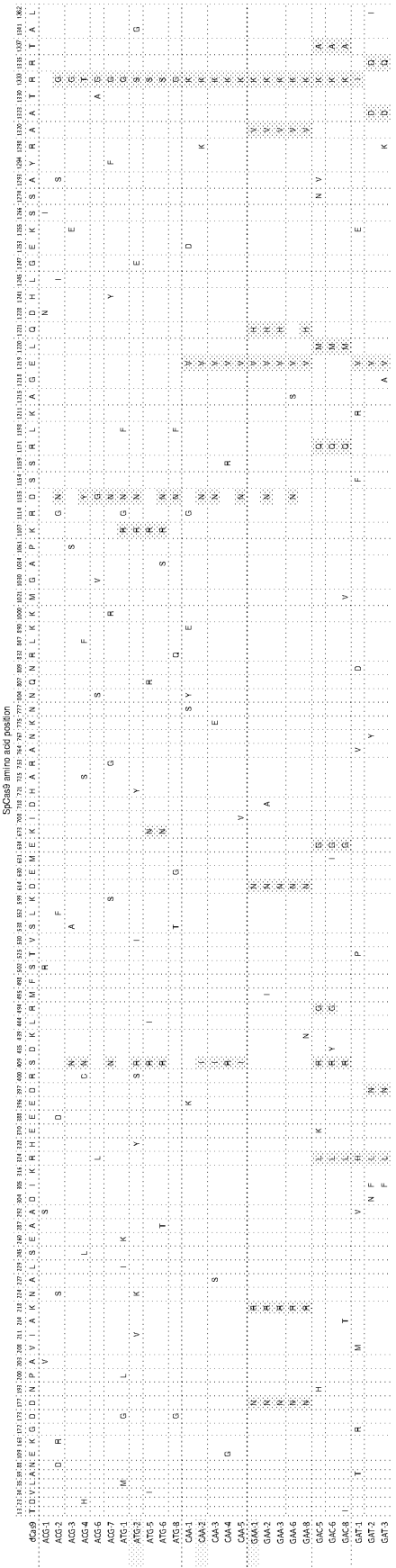

SpCas9 amino acid position
dCas9
AGG-1
AGG-2
AGG-3
AGG-4
AGG-6
AGG-7
ATG-1
ATG-2
ATG-5
ATG-6
ATG-8
CAA-1
CAA-3
CAA-3
CAA-4
CAA-5
GAA-1
GAA-2
GAA-3
GAA-6
GAA-8
GAC-5
GAC-6
GAC-8
GAT-1
GAT-2
GAT-3

FIG. 4A

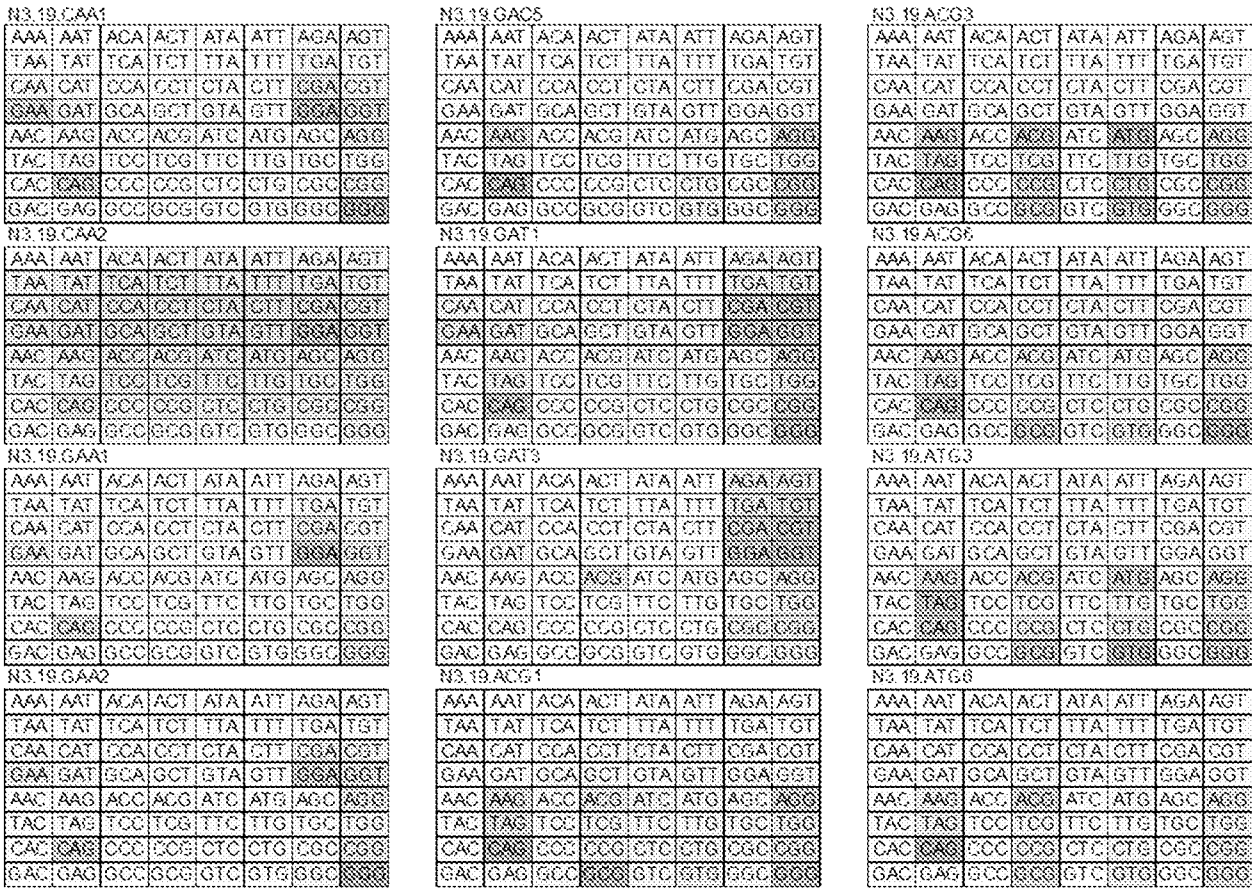

N3.19.CAA1
N3.19.CAA2
N3.19.GAA1
N3.19.GAA2
N3.19.GAC5
N3.19.GAT1
N3.19.GAT3
N3.19.ACG1
N3.19.ACG3
N3.19.ACG6
N3.19.ATG3
N3.19.ATG8

FIG. 4B

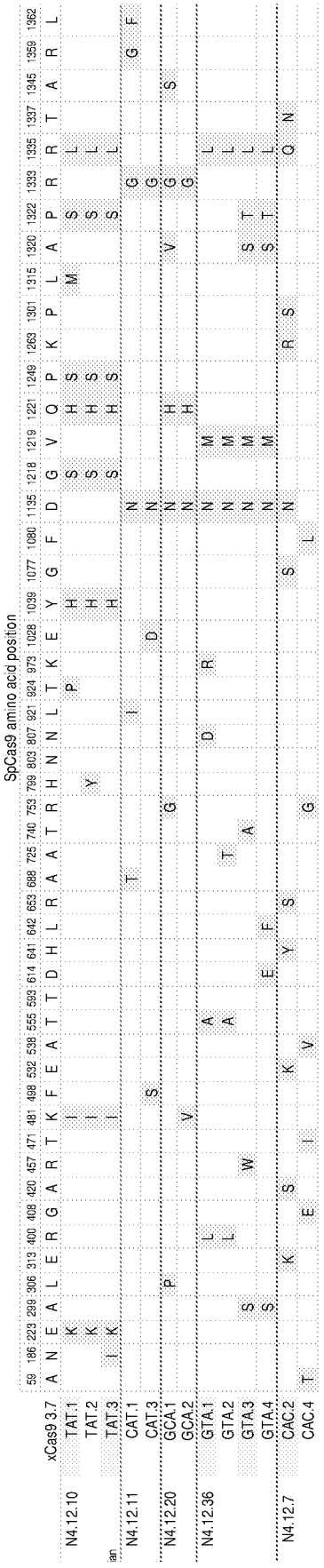

SpCas9 amino acid position
xCas9 3.7
N4.12.10
TAT.1
TAT.2
TAT.3
N4.12.11
CAT.1
CAT.3
N4.12.20
GCA.1
GCA.2
N4.12.36
GTA.1
GTA.2
GTA.3
GTA.4
N4.12.7
CAC.2
CAC.4

FIG. 5A

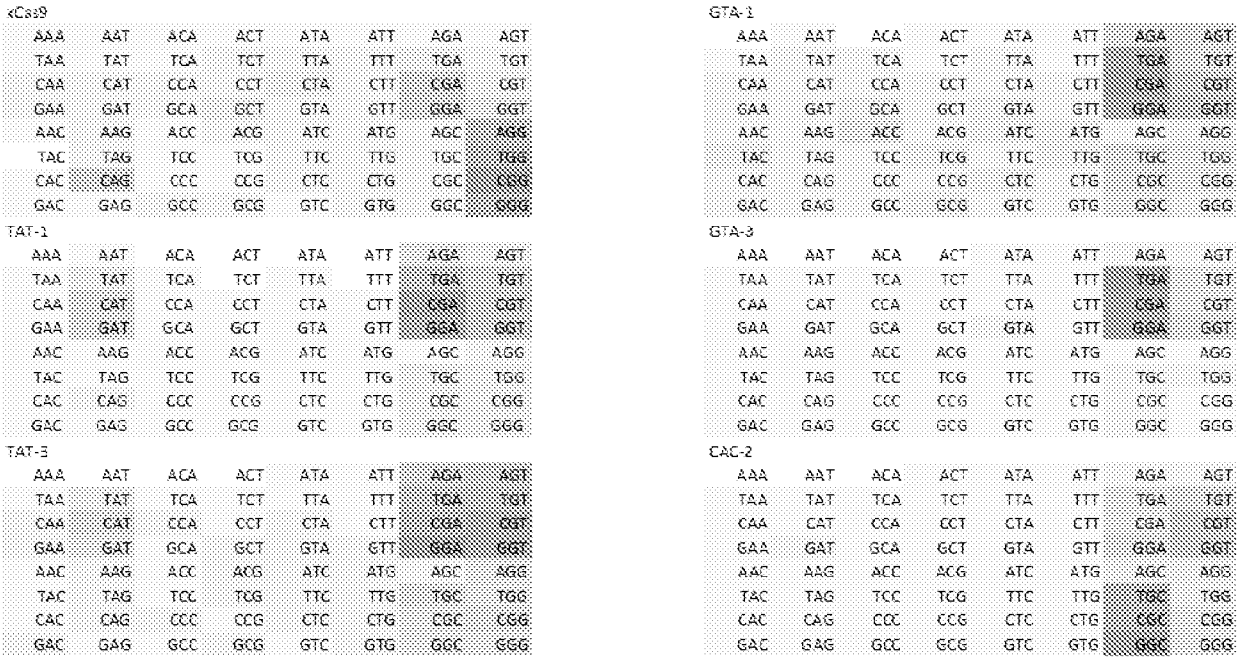


Fig. 5B

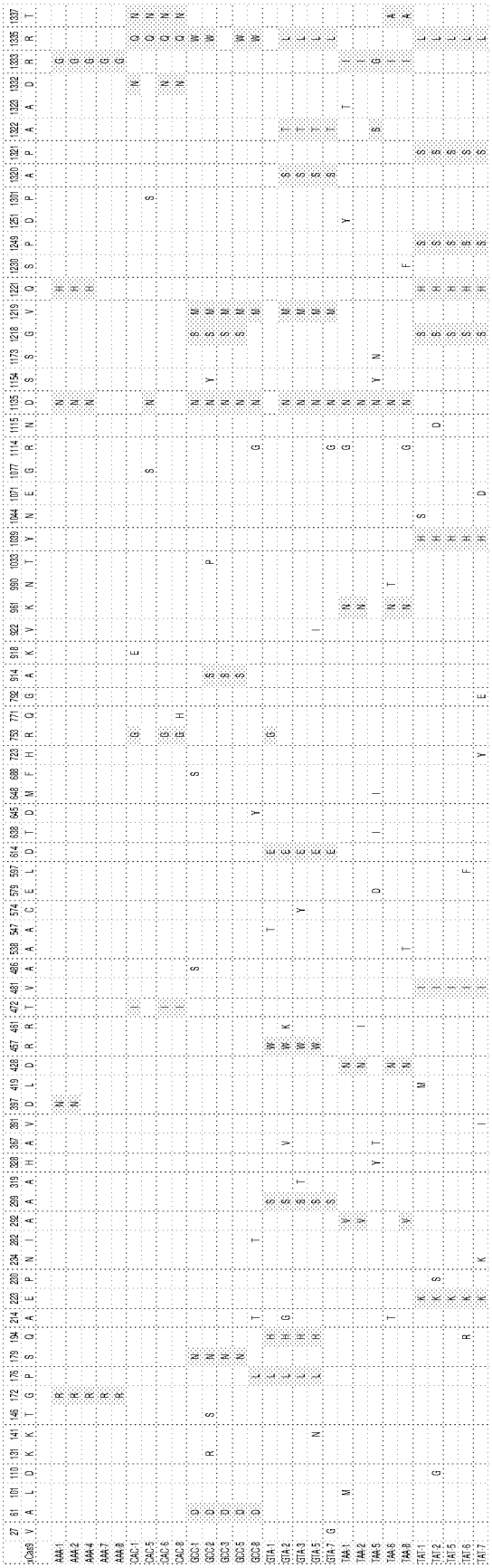


FIG. 6A

N4.19.AAA1

AAA	AAT	ACA	ACT	ATA	ATT	AGA	AGT
TAA	TAT	TCA	TCT	TTA	TTT	TGA	TGT
CAA	CAT	CCA	CCT	CTA	CTT	CGA	CGT
GAA	GAT	GCA	GCT	GTA	GTT	GGA	GGT
AAC	AAG	ACC	ACG	ATC	ATG	AGC	AGG
TAC	TAG	TCC	TCG	TTC	TTG	TGC	TGG
CAC	CAG	CCC	CCG	CTC	CTG	CGC	CGG
GAC	GAG	GCC	GCG	GTC	GTG	GGC	GGG

N4.19.TAA2

AAA	AAT	ACA	ACT	ATA	ATT	AGA	AGT
TAA	TAT	TCA	TCT	TTA	TTT	TGA	TGT
CAA	CAT	CCA	CCT	CTA	CTT	CGA	CGT
GAA	GAT	GCA	GCT	GTA	GTT	GGA	GGT
AAC	AAG	ACC	ACG	ATC	ATG	AGC	AGG
TAC	TAG	TCC	TCG	TTC	TTG	TGC	TGG
CAC	CAG	CCC	CCG	CTC	CTG	CGC	CGG
GAC	GAG	GCC	GCG	GTC	GTG	GGC	GGG

N4.19.TAA5

AAA	AAT	ACA	ACT	ATA	ATT	AGA	AGT
TAA	TAT	TCA	TCT	TTA	TTT	TGA	TGT
CAA	CAT	CCA	CCT	CTA	CTT	CGA	CGT
GAA	GAT	GCA	GCT	GTA	GTT	GGA	GGT
AAC	AAG	ACC	ACG	ATC	ATG	AGC	AGG
TAC	TAG	TCC	TCG	TTC	TTG	TGC	TGG
CAC	CAG	CCC	CCG	CTC	CTG	CGC	CGG
GAC	GAG	GCC	GCG	GTC	GTG	GGC	GGG

N4.19.TAT-5

AAA	AAT	ACA	ACT	ATA	ATT	AGA	AGT
TAA	TAT	TCA	TCT	TTA	TTT	TGA	TGT
CAA	CAT	CCA	CCT	CTA	CTT	CGA	CGT
GAA	GAT	GCA	GCT	GTA	GTT	GGA	GGT
AAC	AAG	ACC	ACG	ATC	ATG	AGC	AGG
TAC	TAG	TCC	TCG	TTC	TTG	TGC	TGG
CAC	CAG	CCC	CCG	CTC	CTG	CGC	CGG
GAC	GAG	GCC	GCG	GTC	GTG	GGC	GGG

N4.19.CAC5

AAA	AAT	ACA	ACT	ATA	ATT	AGA	AGT
TAA	TAT	TCA	TCT	TTA	TTT	TGA	TGT
CAA	CAT	CCA	CCT	CTA	CTT	CGA	CGT
GAA	GAT	GCA	GCT	GTA	GTT	GGA	GGT
AAC	AAG	ACC	ACG	ATC	ATG	AGC	AGG
TAC	TAG	TCC	TCG	TTC	TTG	TGC	TGG
CAC	CAG	CCC	CCG	CTC	CTG	CGC	CGG
GAC	GAG	GCC	GCG	GTC	GTG	GGC	GGG

N4.19.CAC6

AAA	AAT	ACA	ACT	ATA	ATT	AGA	AGT
TAA	TAT	TCA	TCT	TTA	TTT	TGA	TGT
CAA	CAT	CCA	CCT	CTA	CTT	CGA	CGT
GAA	GAT	GCA	GCT	GTA	GTT	GGA	GGT
AAC	AAG	ACC	ACG	ATC	ATG	AGC	AGG
TAC	TAG	TCC	TCG	TTC	TTG	TGC	TGG
CAC	CAG	CCC	CCG	CTC	CTG	CGC	CGG
GAC	GAG	GCC	GCG	GTC	GTG	GGC	GGG

N4.19.GTA2

AAA	AAT	ACA	ACT	ATA	ATT	AGA	AGT
TAA	TAT	TCA	TCT	TTA	TTT	TGA	TGT
CAA	CAT	CCA	CCT	CTA	CTT	CGA	CGT
GAA	GAT	GCA	GCT	GTA	GTT	GGA	GGT
AAC	AAG	ACC	ACG	ATC	ATG	AGC	AGG
TAC	TAG	TCC	TCG	TTC	TTG	TGC	TGG
CAC	CAG	CCC	CCG	CTC	CTG	CGC	CGG
GAC	GAG	GCC	GCG	GTC	GTG	GGC	GGG

N4.19.GTA7

AAA	AAT	ACA	ACT	ATA	ATT	AGA	AGT
TAA	TAT	TCA	TCT	TTA	TTT	TGA	TGT
CAA	CAT	CCA	CCT	CTA	CTT	CGA	CGT
GAA	GAT	GCA	GCT	GTA	GTT	GGA	GGT
AAC	AAG	ACC	ACG	ATC	ATG	AGC	AGG
TAC	TAG	TCC	TCG	TTC	TTG	TGC	TGG
CAC	CAG	CCC	CCG	CTC	CTG	CGC	CGG
GAC	GAG	GCC	GCG	GTC	GTG	GGC	GGG

N4.19.GCC2

AAA	AAT	ACA	ACT	ATA	ATT	AGA	AGT
TAA	TAT	TCA	TCT	TTA	TTT	TGA	TGT
CAA	CAT	CCA	CCT	CTA	CTT	CGA	CGT
GAA	GAT	GCA	GCT	GTA	GTT	GGA	GGT
AAC	AAG	ACC	ACG	ATC	ATG	AGC	AGG
TAC	TAG	TCC	TCG	TTC	TTG	TGC	TGG
CAC	CAG	CCC	CCG	CTC	CTG	CGC	CGG
GAC	GAG	GCC	GCG	GTC	GTG	GGC	GGG

N4.19.GCC5

AAA	AAT	ACA	ACT	ATA	ATT	AGA	AGT
TAA	TAT	TCA	TCT	TTA	TTT	TGA	TGT
CAA	CAT	CCA	CCT	CTA	CTT	CGA	CGT
GAA	GAT	GCA	GCT	GTA	GTT	GGA	GGT
AAC	AAG	ACC	ACG	ATC	ATG	AGC	AGG
TAC	TAG	TCC	TCG	TTC	TTG	TGC	TGG
CAC	CAG	CCC	CCG	CTC	CTG	CGC	CGG
GAC	GAG	GCC	GCG	GTC	GTG	GGC	GGG

N4.19.GCC8

AAA	AAT	ACA	ACT	ATA	ATT	AGA	AGT
TAA	TAT	TCA	TCT	TTA	TTT	TGA	TGT
CAA	CAT	CCA	CCT	CTA	CTT	CGA	CGT
GAA	GAT	GCA	GCT	GTA	GTT	GGA	GGT
AAC	AAG	ACC	ACG	ATC	ATG	AGC	AGG
TAC	TAG	TCC	TCG	TTC	TTG	TGC	TGG
CAC	CAG	CCC	CCG	CTC	CTG	CGC	CGG
GAC	GAG	GCC	GCG	GTC	GTG	GGC	GGG

Fig. 6B

Base Editing:

	sgCAG	sgATG	sgCAT	sgCAA	sgTAT	sgCGT	sgCGG
CAA3-BE3	1.0	x	0.4	0.2	x	x	7.5
ATG2-BE3	6.6	6.9	x	x	x	3.4	33.0
TAT1-BE3	1.1	x	4.7	0.4	1.0	27.0	19.5
xCas9-BE3	1.1	1.1	0.6	0.3	0.0	34.2	35.1
control	0.0	0.0	0.0	0.0	x	x	0.0

Indel formation:

	sgCAG	sgATG	sgCAT	sgCAA	sgTAT	sgCGT	sgCGG
CAA3	2.7	x	8.5	0.5	x	x	13.6
ATG2	15.4	1.1	x	x	x	2.1	26.9
TAT1	0.0	x	2.9	0.0	0.0	0.0	0.2
xCas9	3.9	3.8	13.2	0.7	0.1	46.8	30.7
control	0.0	x	1.9	0.0	0.0	x	0.0

FIG. 7

N6.2

AAA	AAT	ACA	ACT	ATA	ATT
TAA	TAT	TCA	TCT	TTA	TTT
CAA	CAT	CCA	CCT	CTA	CTT
GAA	GAT	GCA	GCT	GTA	GTT
AAC	AAG	ACC	ACG	ATC	ATG
TAC	TAG	TCC	TCG	TTC	TTG
CAC	CAG	CCC	CCG	CTC	CTG
GAC	GAG	GCC	GCG	GTC	GTG

N6.12

AAA	AAT	ACA	ACT	ATA	ATT
TAA	TAT	TCA	TCT	TTA	TTT
CAA	CAT	CCA	CCT	CTA	CTT
GAA	GAT	GCA	GCT	GTA	GTT
AAC	AAG	ACC	ACG	ATC	ATG
TAC	TAG	TCC	TCG	TTC	TTG
CAC	CAG	CCC	CCG	CTC	CTG
GAC	GAG	GCC	GCG	GTC	GTG

N6.16

AAA	AAT	ACA	ACT	ATA	ATT
TAA	TAT	TCA	TCT	TTA	TTT
CAA	CAT	CCA	CCT	CTA	CTT
GAA	GAT	GCA	GCT	GTA	GTT
AAC	AAG	ACC	ACG	ATC	ATG
TAC	TAG	TCC	TCG	TTC	TTG
CAC	CAG	CCC	CCG	CTC	CTG
GAC	GAG	GCC	GCG	GTC	GTG

FIG. 8

S. pyogenes Cas9 amino acid
223 262 324 409 480 481 543 694 924 1039 1218 1219 1221 1249 1315 1321 1335
E A R S E V E M T Y G E Q P L P R
K T L I K I D I P H S V H S M S L
Sequence 7777777

FIG. 9A

TAT-1
AAA AAT ACA ACT ATA ATT AGA AGT
TAA TAT TCA TCT TTA TTT TGA TGT
CAA CAT CCA CCT CTA CTT CGA CGT
GAA GAT GCA GCT GTA GTT GGA GGT
AAC AAG ACC ACG ATC ATG AGC AGG
TAC TAG TCC TCG TTC TTG TGC TGG
CAC CAG CCC CCG CTC CTG CGC CGG
GAC GAG GCC GCG GTC GTG GGC GGG

FIG. 9B

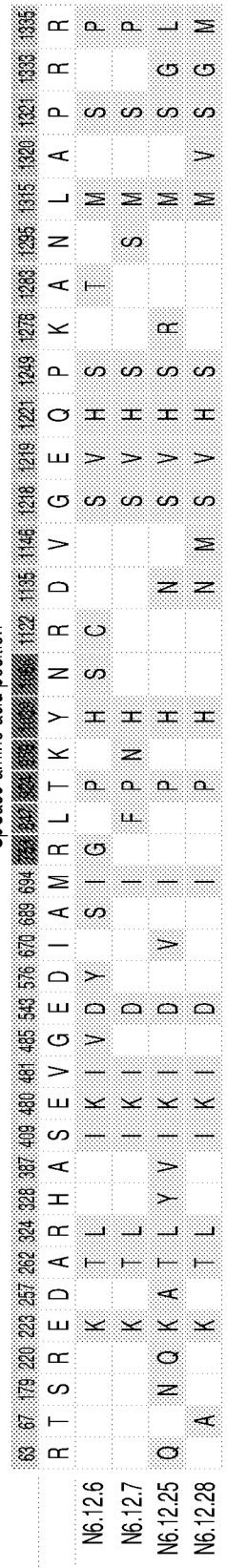

SpCas9 amino acid position
63 67 179 220 223 257 262 324 328 367 409 430 481 485 543 576 670 689 694 1122 1135 1145 1218 1219 1221 1249 1278 1288 1295 1315 1320 1321 1333 1335
R T S R E D A R H A S E V G E D I A M R L T K Y N R D V G E Q P K A N L A P R R
N6.12.6 K T L I K I V D Y S I G P H S C S V H S T M S P
N6.12.7 K T L I K I D I F P N H S V H S S M S P
N6.12.25 Q N Q K A T L Y V I K I D V I P H N S V H S R M S G L
N6.12.28 A K T L I K I D I P H N M S V H S M V S G M

FIG. 10

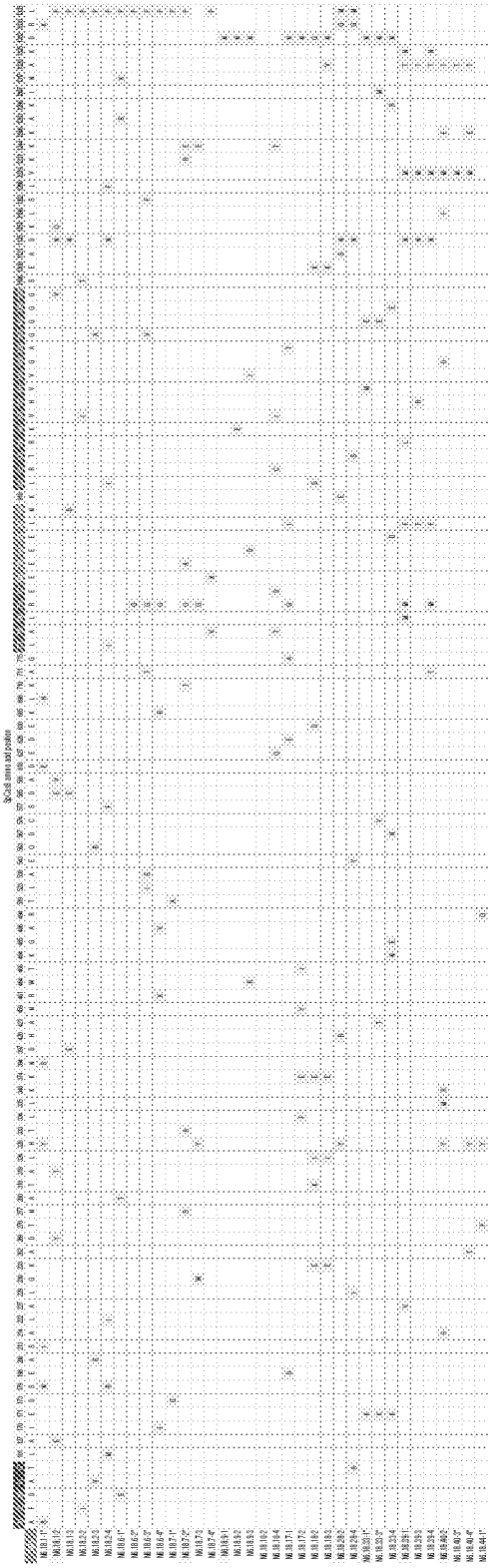


FIG. 11

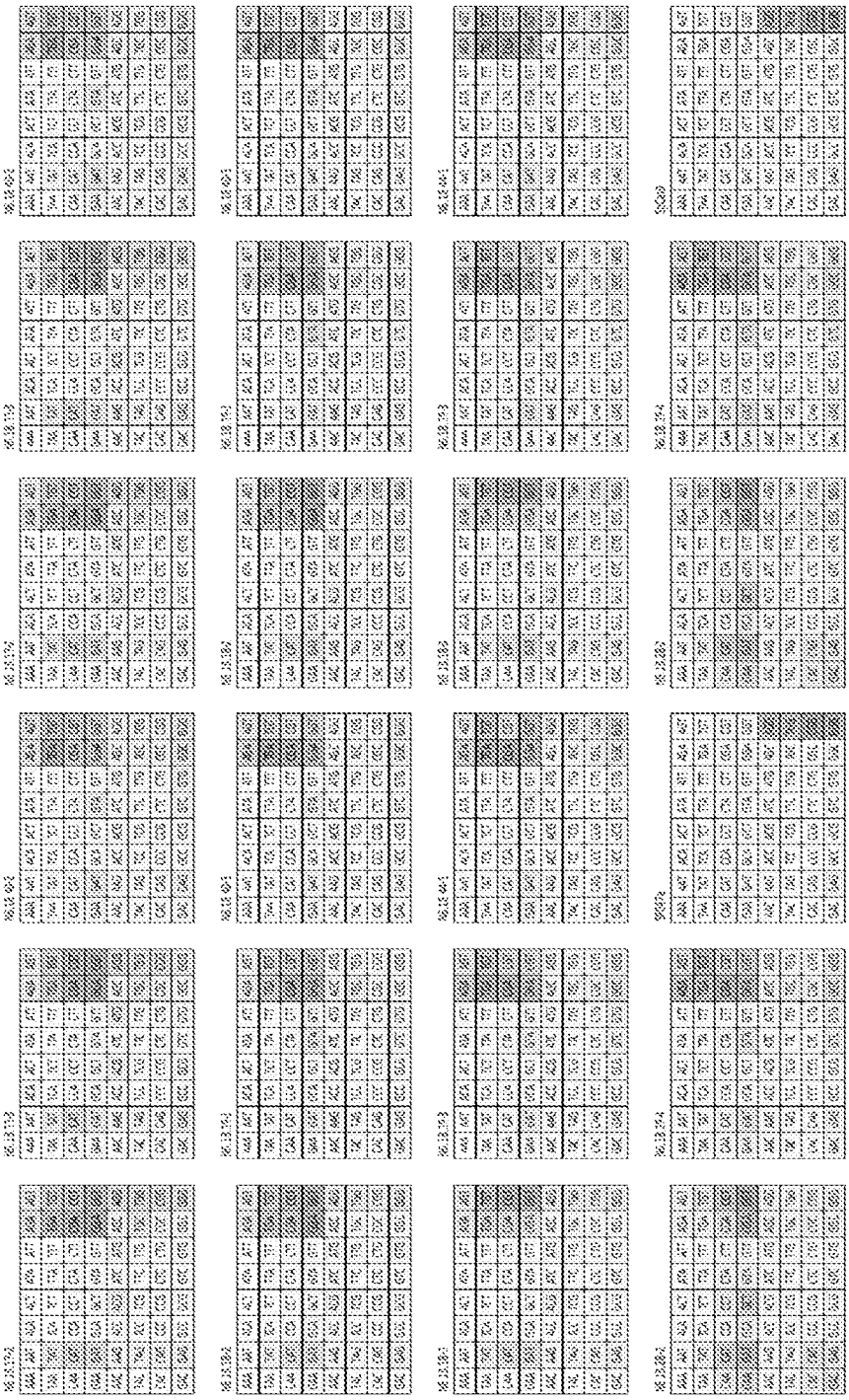


FIG. 12

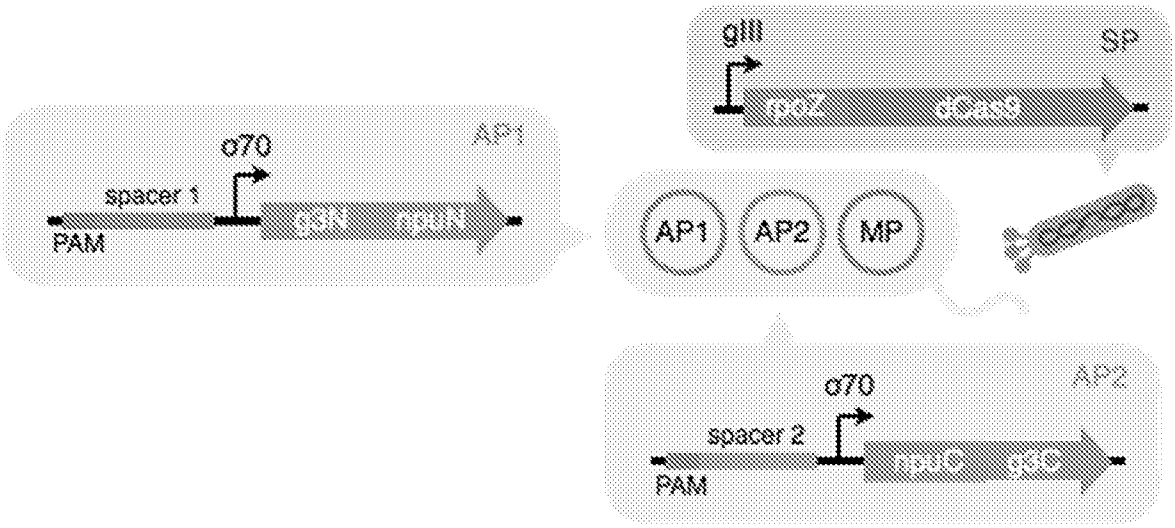


FIG. 13A

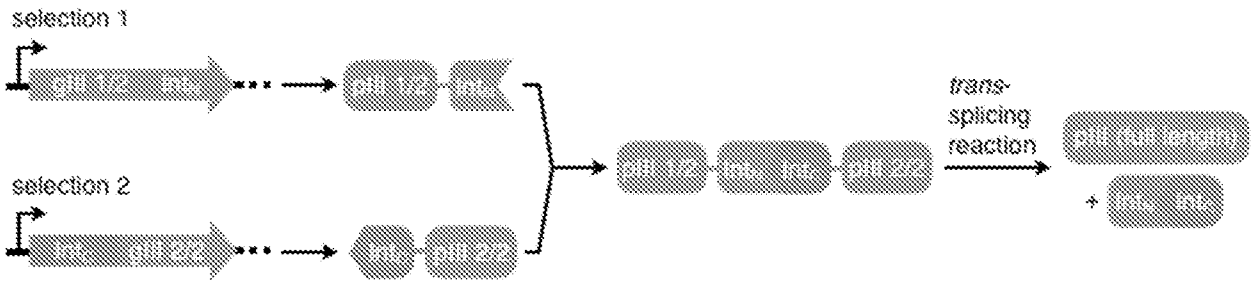


FIG. 13B

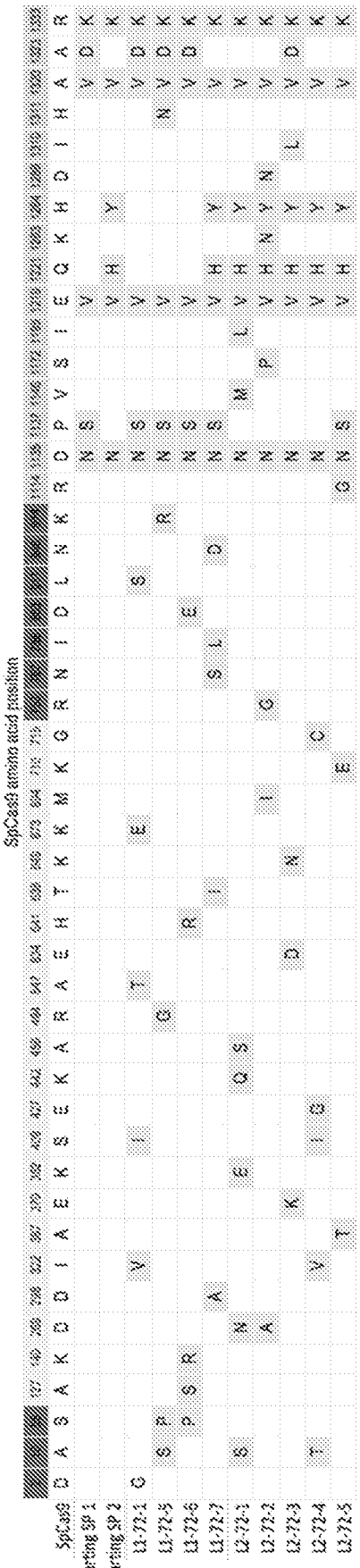

SpCas9 amino acid position
SpCas9
starting gR 1
starting gR 2
U1-72-1
U1-72-5
U1-72-6
U1-72-7
U2-72-1
U2-72-2
U2-72-3
U2-72-4
U2-72-5

FIG. 14A

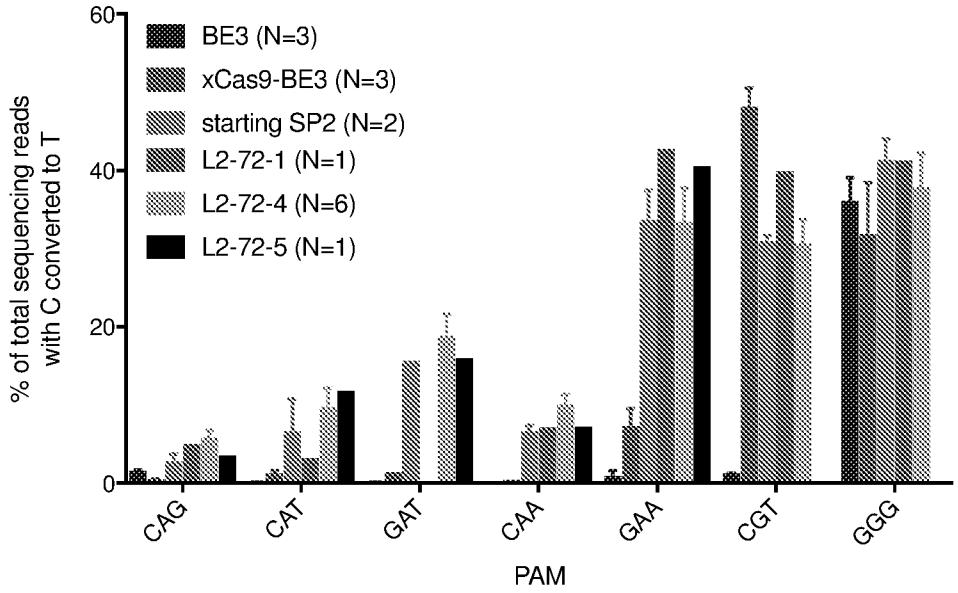


FIG. 14B

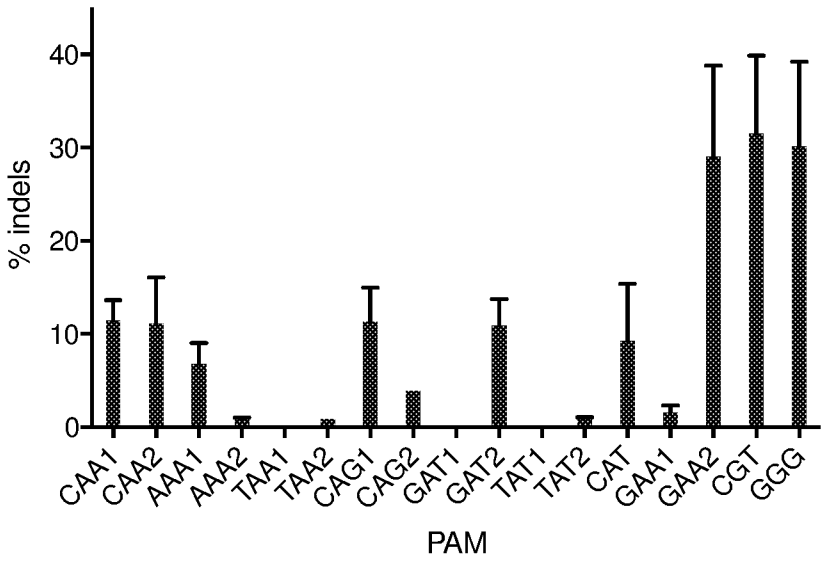


FIG. 14C

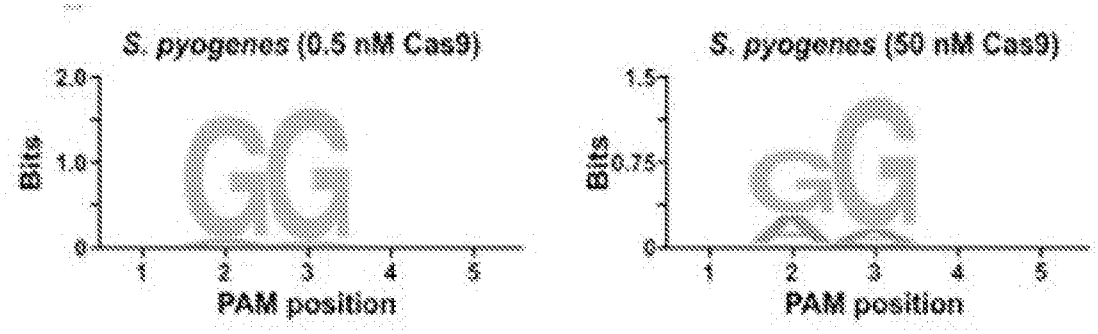


FIG. 15A

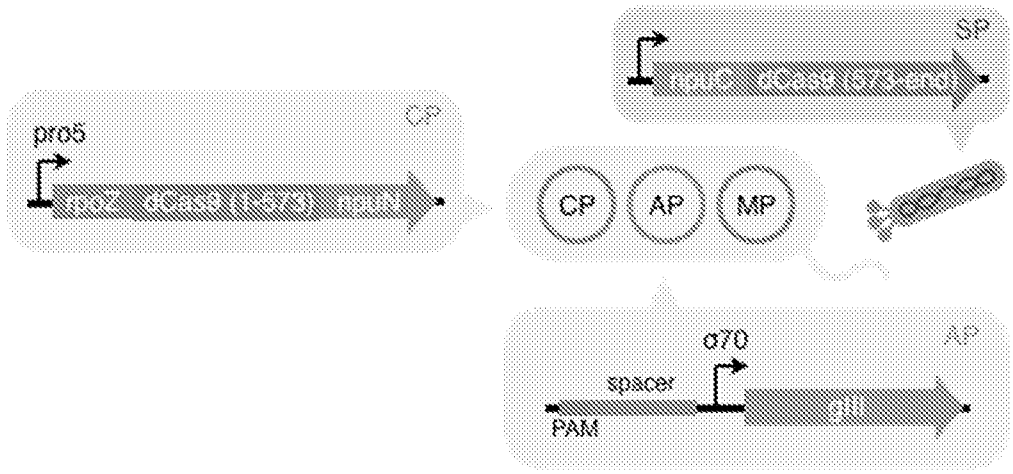


FIG. 15B

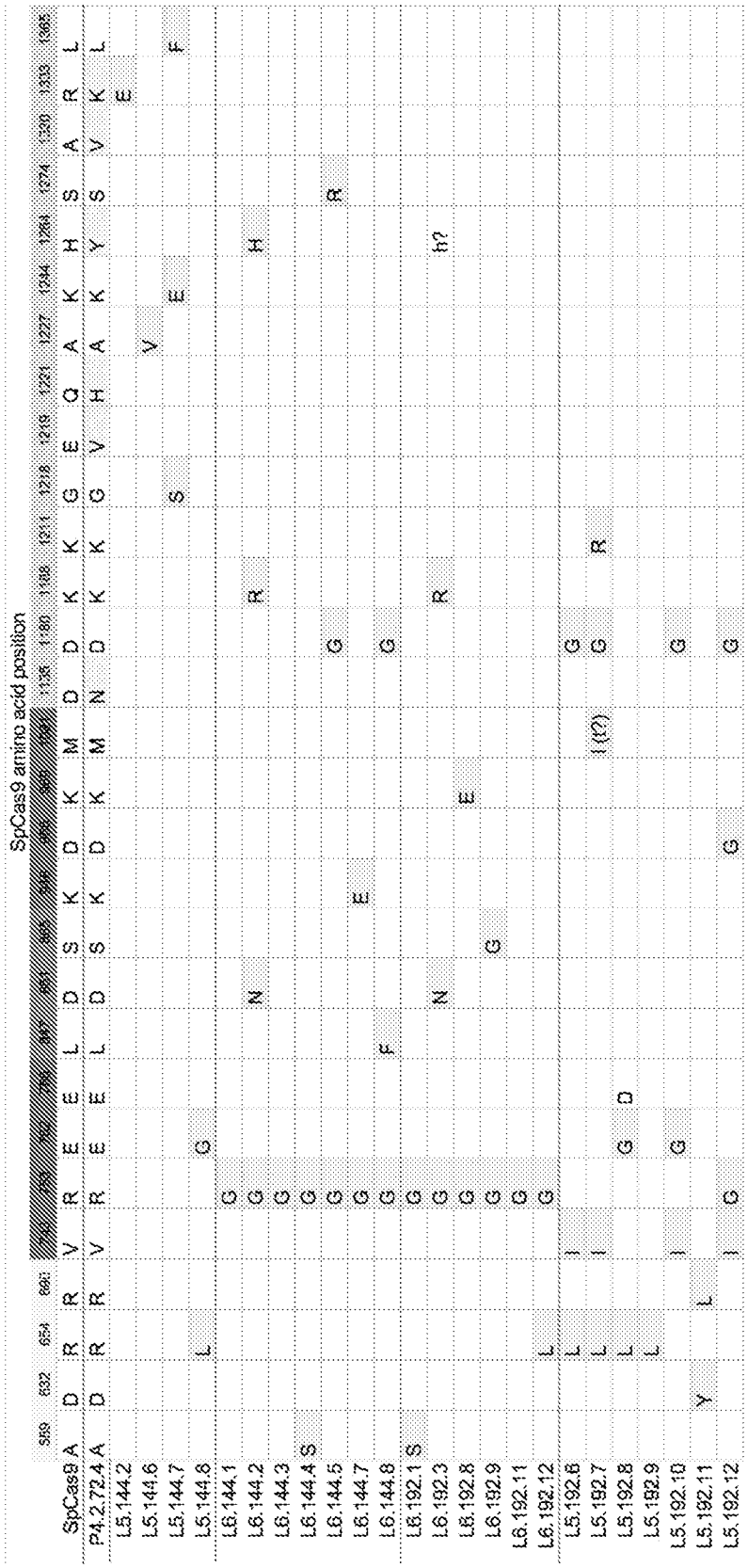

SpCas9 amino acid position
SpCas9 A
P4.2.72.4 A
L5.144.2
L5.144.6
L5.144.7
L5.144.8
L6.144.1
L6.144.2
L6.144.3
L6.144.4 S
L6.144.5
L6.144.7
L6.144.8
L6.192.1 S
L6.192.3
L6.192.8
L6.192.9
L6.192.11
L6.192.12
L5.192.6
L5.192.7
L5.192.8
L5.192.9
L5.192.10
L5.192.11
L5.192.12

FIG. 16

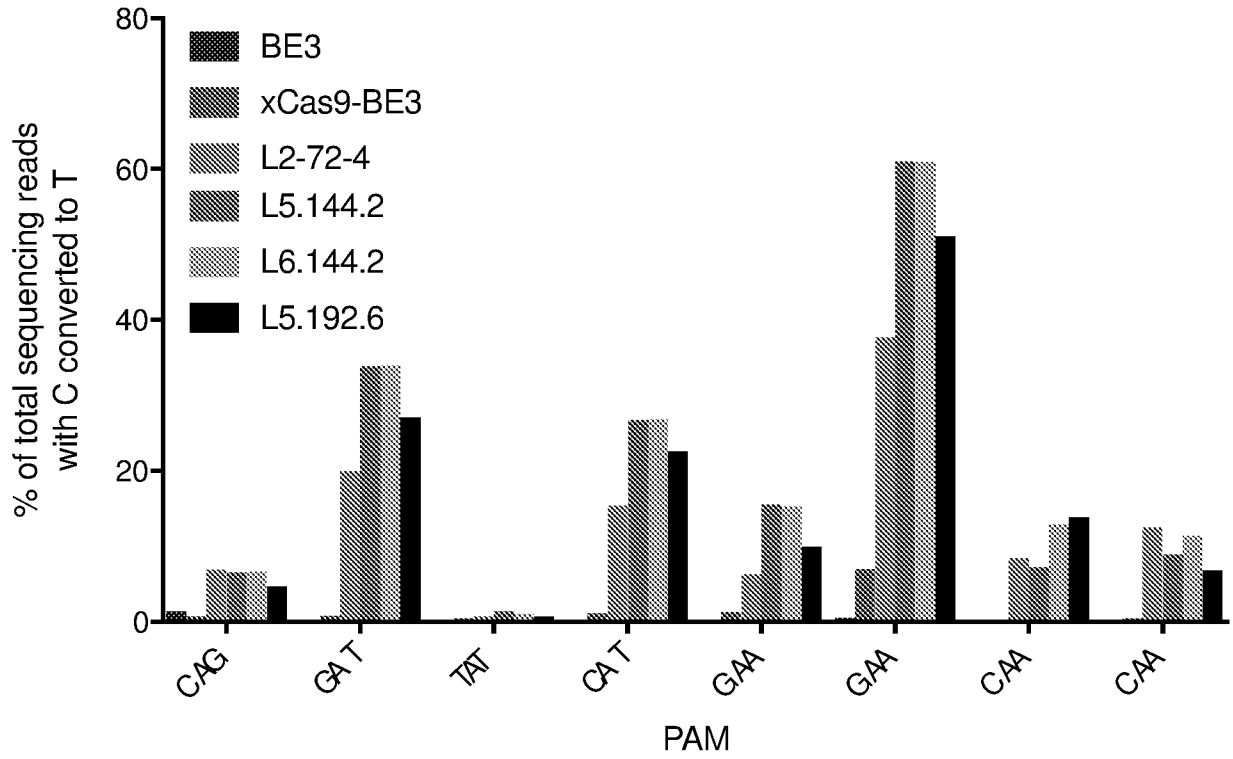

80
60
40
20
0
% of total sequencing reads with C converted to T
BE3
xCas9-BE3
L2-72-4
L5.144.2
L6.144.2
L5.192.6
CAG
GAT
TAT
CAT
GAA
GAA
CAA
CAA
PAM

FIG. 17

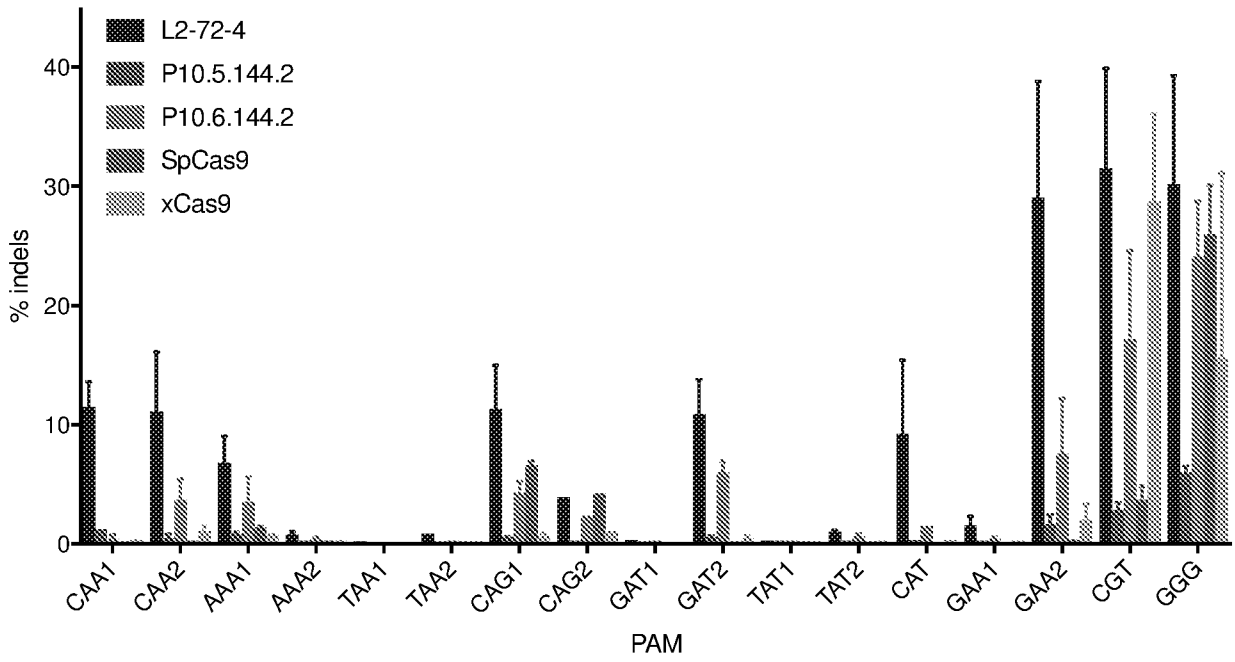

L2-72-4
P10.5.144.2
P10.6.144.2
SpCas9
xCas9
40
30
20
10
0
% indels
CAA1
CAA2
AAA1
AAA2
TAA1
TAA2
CAG1
CAG2
GAT1
GAT2
TAT1
TAT2
CAT
GAA1
GAA2
CGT
GGG
PAM

FIG. 18

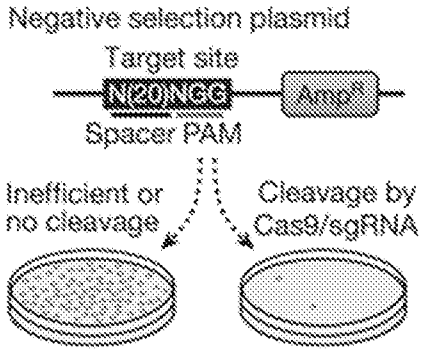


FIG. 19A

depletion on NNN

AAA	AAT	ACA	ACT	ATA	ATT	AGA	AGT
TAA	TAT	TCA	TCT	TTA	TTT	TGA	TGT
CAA	CAT	CCA	CCT	CTA	CTT	CGA	CGT
GAA	GAT	GCA	GCT	GTA	GTT	GGA	GGT
AAC	AAG	ACC	ACG	ATC	ATG	AGC	AGG
TAC	TAG	TCC	TCG	TTC	TTG	TGC	TGG
CAC	CAG	CCC	CCG	CTC	CTG	CGC	CGG
GAC	GAG	GCC	GCG	GTC	GTG	GGC	GGG

luciferase assay

AAA	AAT	ACA	ACT	ATA	ATT	AGA	AGT
TAA	TAT	TCA	TCT	TTA	TTT	TGA	TGT
CAA	CAT	CCA	CCT	CTA	CTT	CGA	CGT
GAA	GAT	GCA	GCT	GTA	GTT	GGA	GGT
AAC	AAG	ACC	ACG	ATC	ATG	AGC	AGG
TAC	TAG	TCC	TCG	TTC	TTG	TGC	TGG
CAC	CAG	CCC	CCG	CTC	CTG	CGC	CGG
GAC	GAG	GCC	GCG	GTC	GTG	GGC	GGG

FIG. 19B

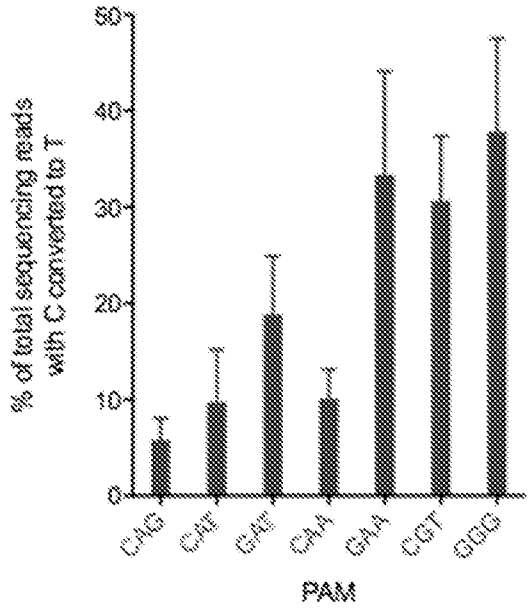


FIG. 19C

depletion on NNNN shows preference of C > T > A > G

	4th nucleotide			
PAM	A	C	G	T
AAA	-0.067	-1.124	-0.168	-0.378
TAA	-0.732	-1.271	-0.753	-1.210
CAA	-0.228	-1.169	-0.232	-0.566
GAA	-1.250	-2.657	-1.418	-1.632
AAG	-1.009	-1.643	-0.639	-0.978
TAG	-0.497	-1.216	-0.252	-0.856
CAG	-1.055	-2.013	-0.845	-1.327
GAG	-2.199	-2.750	-1.797	-2.411
AGA	-1.041	-1.881	-0.890	-1.216
TGA	-0.455	-1.580	-0.348	-1.056
CGA	-1.011	-2.406	-0.793	-1.315
GGA	-1.543	*	-1.333	-2.030
AGG	-1.505	*	-1.215	-1.752
TGG	-0.835	-1.507	-0.513	-1.179
CGG	-1.366	-2.442	-0.897	-1.415
GGG	-2.423	*	-1.503	-2.862

* no score calculated because 100% of plasmids were cut

Interestingly, xCas9 has the same 3'-PAM preference

	4th nucleotide			
PAM	A	C	G	T
AGA	-0.396	-1.677	-0.444	-0.553
TGA	-0.638	-1.926	-0.574	-0.899
CGA	-0.708	-2.066	-0.632	-1.018
GGA	-0.877	-2.006	-0.838	-1.197
AGT	-0.327	-1.696	-0.238	-0.321
TGT	-0.612	-1.855	-0.449	-0.723
CGT	-0.868	-2.254	-0.646	-0.995
GGT	-0.931	-1.975	-0.672	-1.023
AGC	-0.202	-1.036	-0.001	-0.238
TGC	-0.305	-1.202	-0.085	-0.391
CGC	-0.472	-1.305	-0.141	-0.614
GGC	-0.601	-1.33	-0.233	-0.666
AGG	-1.681	*	-1.697	-2.501
TGG	-1.96	*	-1.603	-2.866
CGG	-2.253	-3.341	-1.961	-3.11
GGG	-2.595	-2.829	-1.983	-3.159

...But wtCas9 does not

	4th nucleotide			
PAM	A	C	G	T
AAG	-0.859	-0.974	-1.714	-0.707
TAG	-0.785	-1.116	-1.027	-0.824
CAG	-1.007	-1.21	-0.844	-1.13
GAG	-1.402	-1.873	-1.297	-1.647
AGA	-0.402	-0.485	-0.826	-0.29
TGA	-0.569	-0.852	-0.979	-0.548
CGA	-0.551	-0.697	-0.754	-0.382
GGA	-0.715	-0.854	-1.046	-0.586
AGG	-2.762	-2.181	-2.931	-3.31
TGG	-2.71	*	-2.453	-3.197
CGG	-3.238	*	-2.423	-2.743
GGG	-2.529	-3.035	-2.433	-2.792

FIG. 20

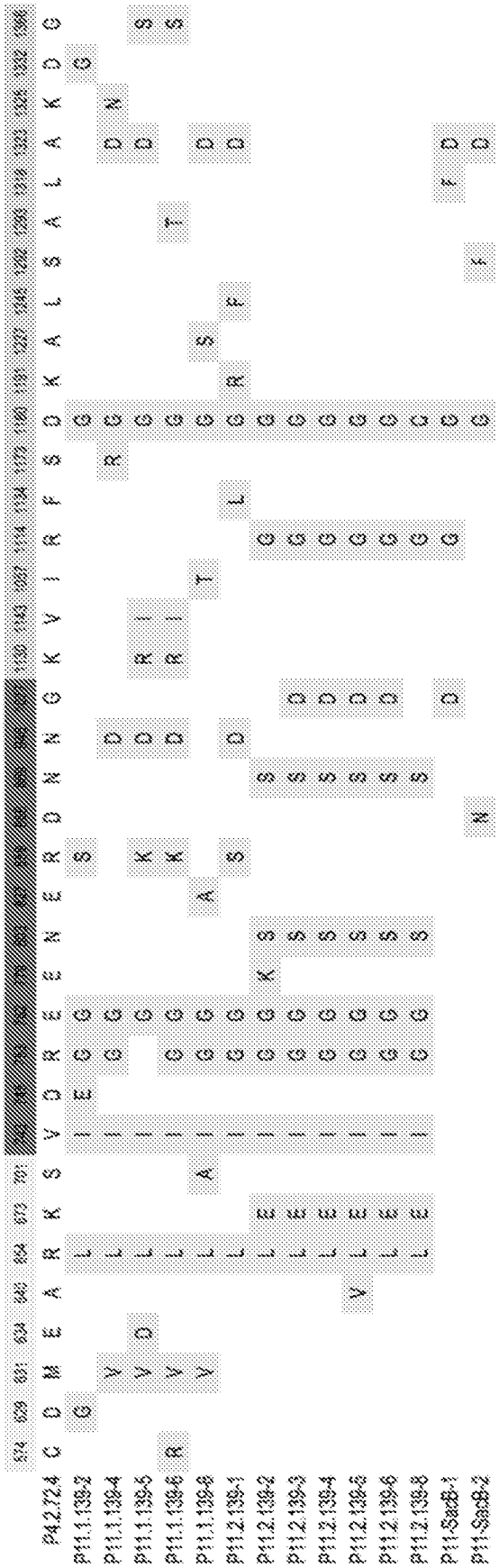

574 C
629 D
631 M
634 E
640 A
654 R
672 K
701 S
707 V
717 D
718 R
719 E
720 E
721 N
722 E
723 R
724 D
725 N
726 N
727 G
728 K
1139 K
1140 V
1141 I
1142 R
1143 F
1144 S
1145 D
1146 K
1147 A
1148 L
1149 S
1150 A
1151 L
1152 A
1153 K
1154 D
1155 G
P42.272-4
P11.1.139-2
P11.1.139-4
P11.1.139-5
P11.1.139-6
P11.1.139-8
P11.2.139-1
P11.2.139-2
P11.2.139-3
P11.2.139-4
P11.2.139-6
P11.2.139-6
P11.2.139-8
P11-SacB-1
P11-SacB-2

FIG. 21

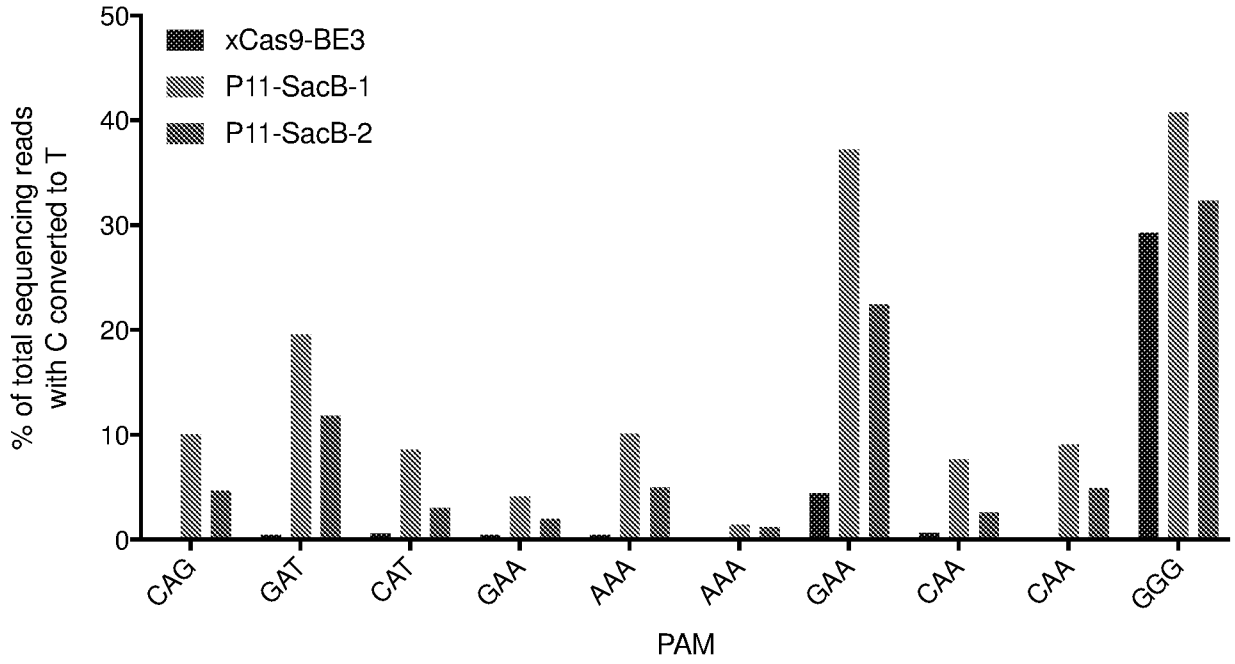

xCas9-BE3
P11-SacB-1
P11-SacB-2
% of total sequencing reads with C converted to T
50
40
30
20
10
0
CAG
GAT
CAT
GAA
AAA
AAA
GAA
CAA
CAA
GGG
PAM

FIG. 22

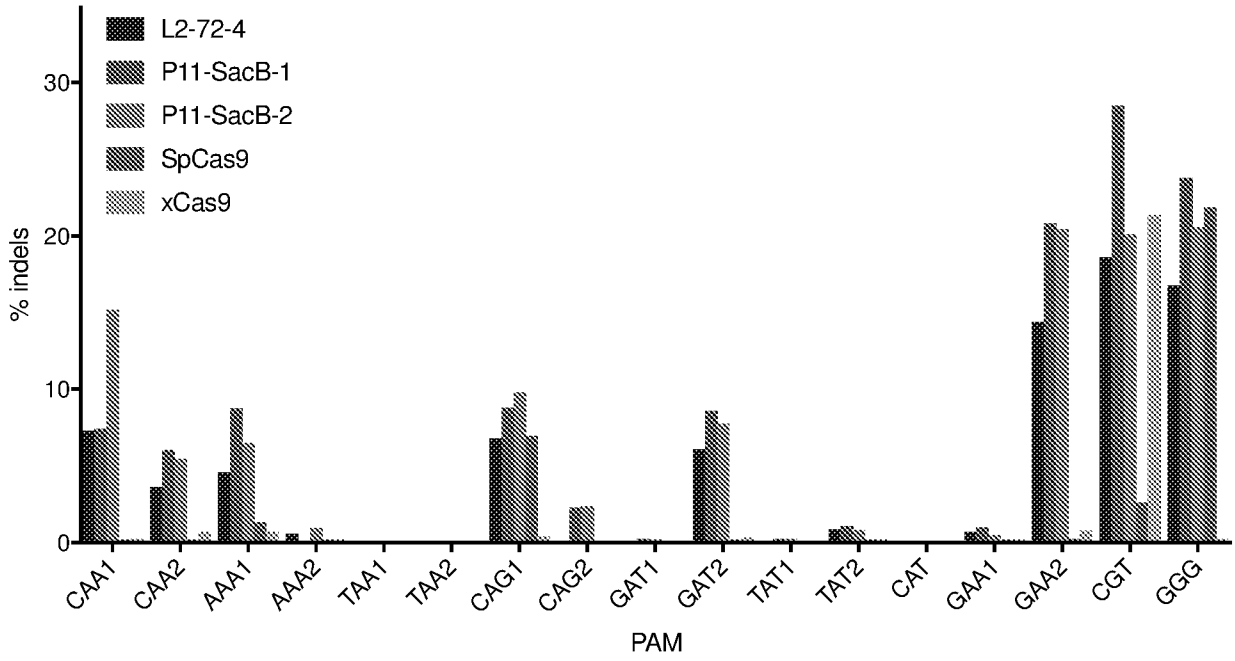

L2-72-4
P11-SacB-1
P11-SacB-2
SpCas9
xCas9
% indels
30
20
10
0
CAA1
CAA2
AAA1
AAA2
TAA1
TAA2
CAG1
CAG2
GAT1
GAT2
TAT1
TAT2
CAT
GAA1
GAA2
CGT
GGG
PAM

FIG. 23

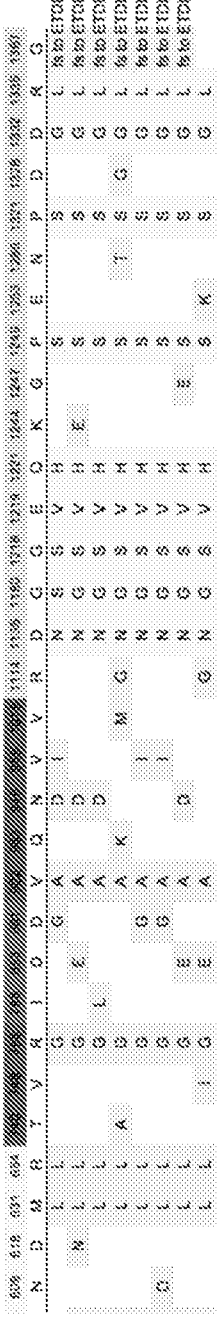


FIG. 24A

FIG. 24B

% editing

site	s893	s1073	s1081	s1140	b3	e1	e2	f1	f2	s33	s34	s35	s36	s37	s38	s39	s40	s41	s43	s44	s45	s46
	GGG	GAAC	GGTC	TGTC	GATG	TATG	TATT	CATC	CATT	AATA	AATT	AATC	AATG	TATA	TATT	TATC	TATG	CATA	CATC	CATG	GATA	GATT
pSM060ax	44.9	21.7	39.2	43.4	45.9	3.5	2.3	33.1	32.4	8.0	1.4	0.9	2.4	17.3	1.7	22.5	3.4	1.0	10.4	0.6	21.0	15.2
pSM060ax2	44.9	31.0	43.4	48.9	48.8	4.3	3.6	40.4	35.6	13.8	2.2	1.4	4.5	23.3	2.5	25.8	5.1	1.8	11.3	2.3	26.7	18.1
pSM060ay	40.7	18.4	35.8	35.5	38.1	2.8	1.6	29.1	20.2	7.0	1.1	0.7	2.0	13.1	1.4	19.0	2.5	0.7	7.8	0.5	16.6	13.6
pSM060ay2	46.3	30.3	49.6	41.1	45.6	3.9	4.2	40.8	34.9	14.7	2.7	1.5	3.7	18.7	2.2	26.6	4.8	1.8	10.9	1.8	24.0	15.4
SpCas9-NG	37.8	11.1	32.6	18.3	39.0	1.6	0.1	8.7	3.3	0.2	0.4	0.2	20.4	2.0	0.2	2.5	5.2	0.4	1.8	0.9	1.6	2.3
xCas9	31.6	4.3	17.8	5.4	0.5	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.1	0.0	0.2	0.0	0.0	0.1	0.0	0.1	0.1

FIG. 25A

Averages:

PAM	NATA	NATT	NATC	NATG
pSM060ax	11.8	10.6	16.7	11.2
pSM060ax2	16.4	12.4	19.7	13.0
pSM060ay	9.3	7.6	14.2	9.2
pSM060ay2	14.8	11.9	20.0	12.0
SpCas9-NG	1.1	1.3	3.3	13.4
xCas9	0.1	0.1	0.1	0.1

FIG. 25B

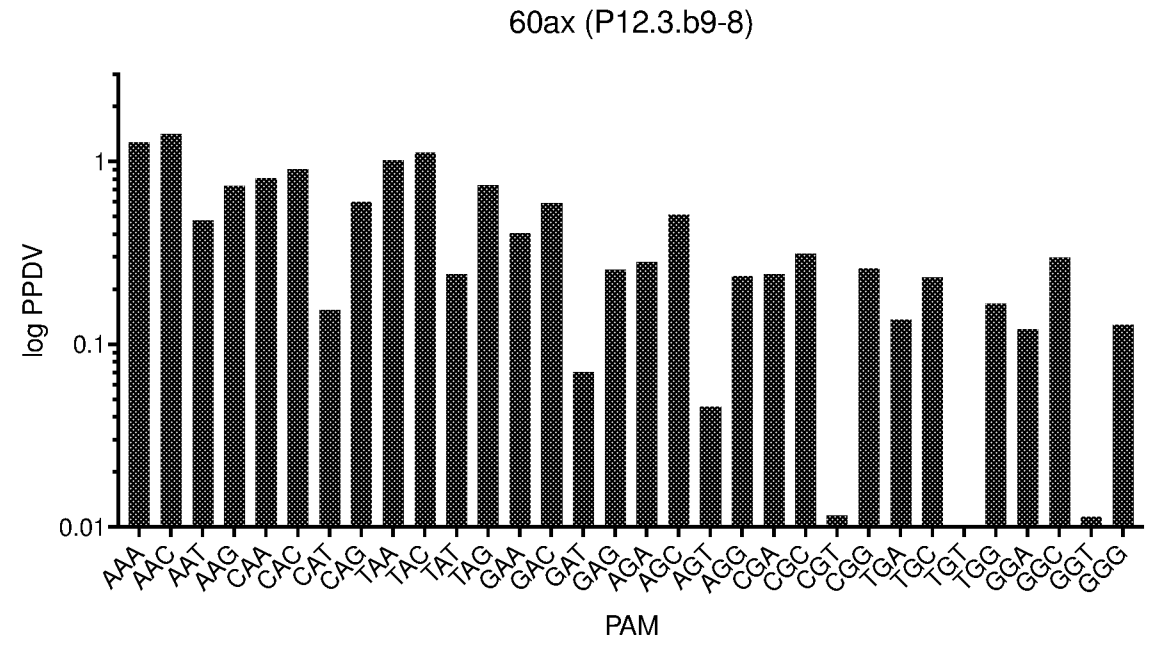


FIG. 26A

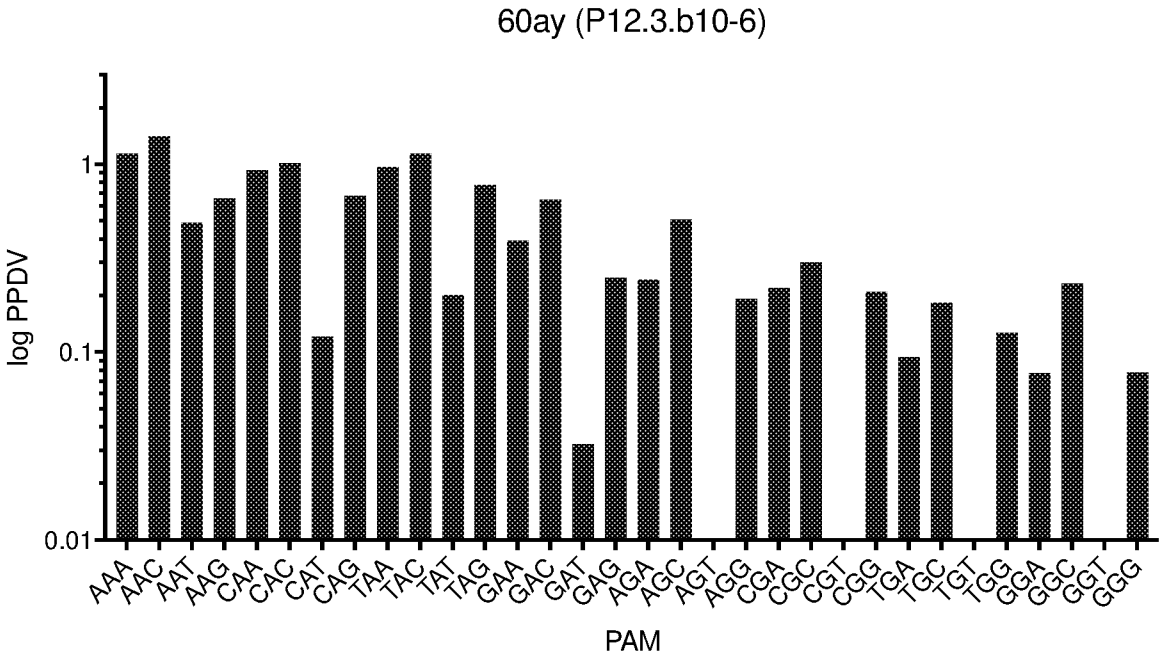


FIG. 26B

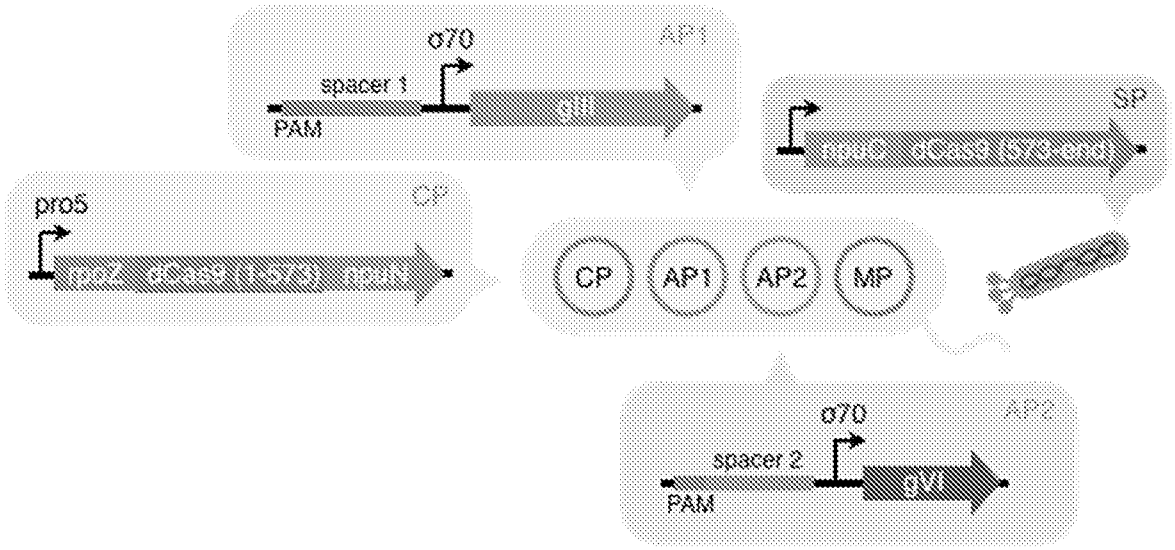


FIG. 27A

	930	942	954	973	993										1114	1135	1161	1189	1219	1221	1223	1264	1280	1318	1320	1323	1335	1366
SpCas9	E	L	R	K	F	V	R	E	N	N	H	Y	G	F	R	D	K	D	E	Q	G	H	V	L	A	A	R	D
L2-120-1			L	E		I	G	G	S	S		D	D		G	N		G	V	H		Y	S	S	V	D	K	
L2-120-2			L	E		I	G	G	S	S		D	D		G	N	E	G	V	H		Y		S	V		K	
L2-120-3			L	E		I	G	G	S	S		D	D		G	N		G	V	H		Y		S	V	D	K	
L2-120-4			L	E	L	I	G	G	S	S	P	D	D	S	G	N		G	V	H		Y		S	V	D	K	
L2-120-5	K		L	E		I	G	G	S	S		D	D		G	N		G	V	H		Y		S	V		K	
L2-120-6			L	E	L	I	G	G	S	S		D	D		G	N		G	V	H	S	Y		S	V		K	
L2-120-7			L	E		I	G	G	S	S		C			G	N		G	V	H		Y			V		K	E
L2-120-8		R	L	E		I	G	G	S	S		D	D		G	N		G	V	H		Y		S	V	D	K	

FIG. 27B

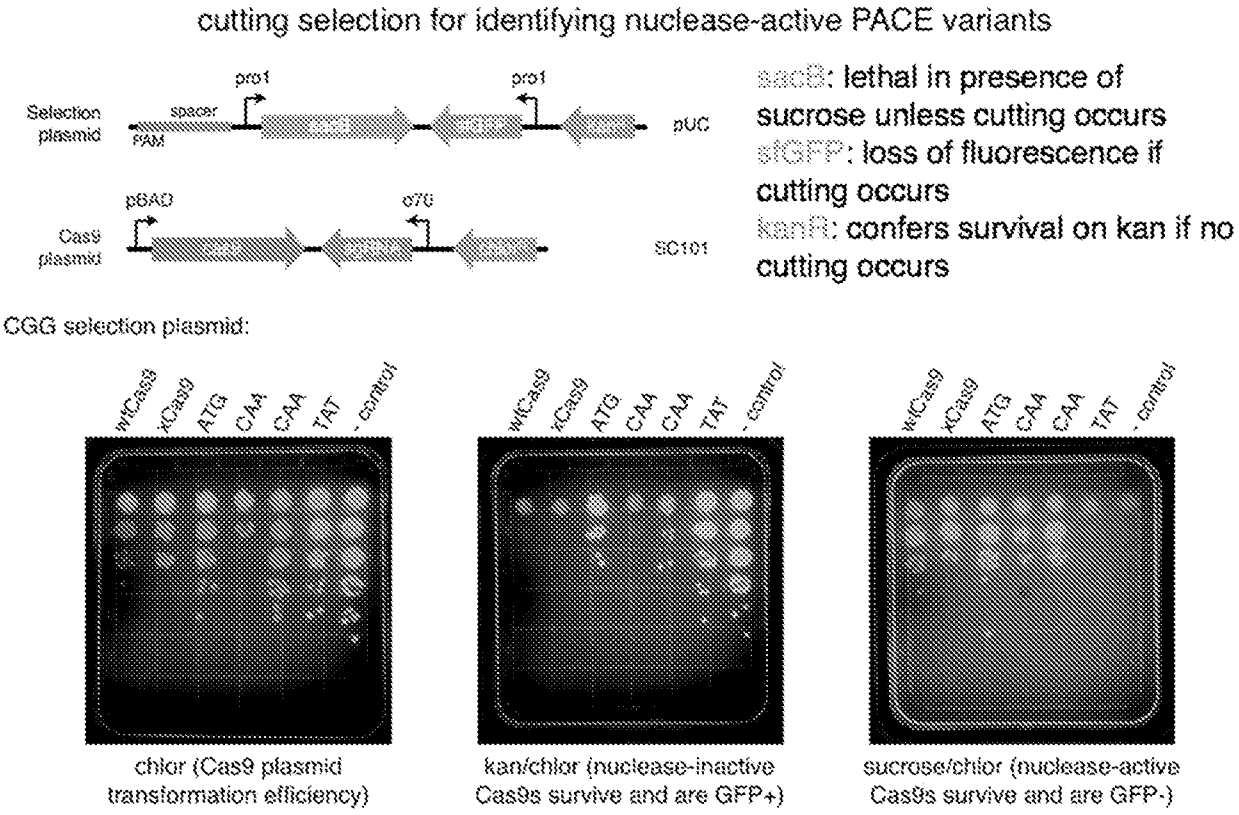


FIG. 28

	223	262	324	409	480	543	694	692	961	963	1003	1135	1191	1216	1219	1221	1249	1315	1320	1321	1332	1335
SpCas9	E	A	R	S	E	E	M	T	K	H	Y	D	K	G	E	Q	P	L	A	P	D	R
TAT	K	T	L	I	K	D	I	P			H			S	V	H	S	M		S		L
SacB-TAT-1									E	Y		N	N		V	H			V	S		L
SacB-TAT-2												N		S	V	H	S			S	G	L

FIG. 29A

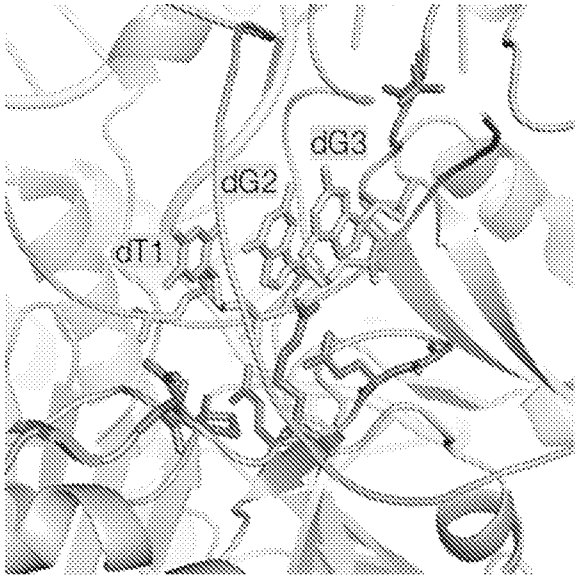


FIG. 29B

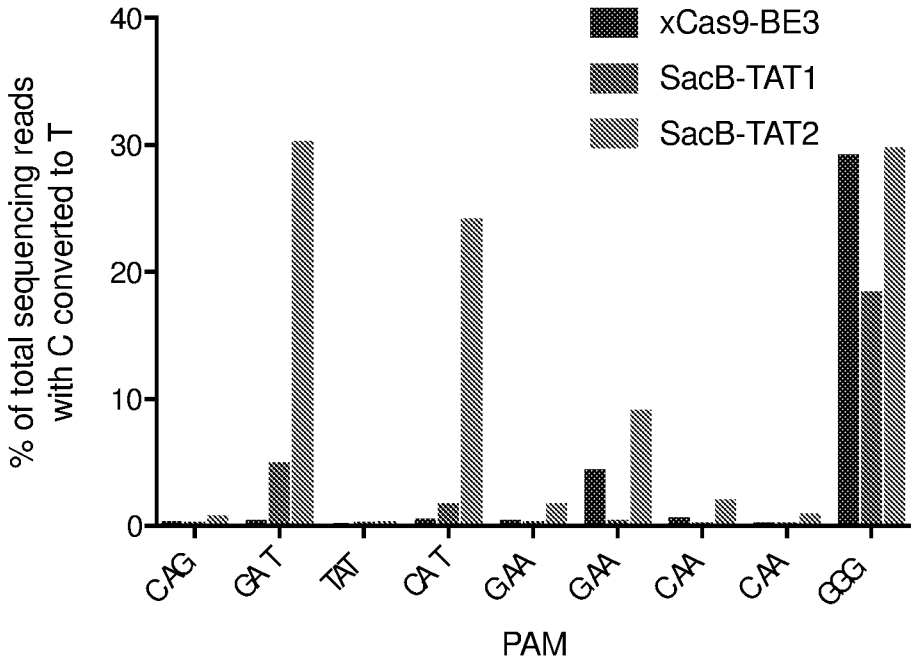


FIG. 30A

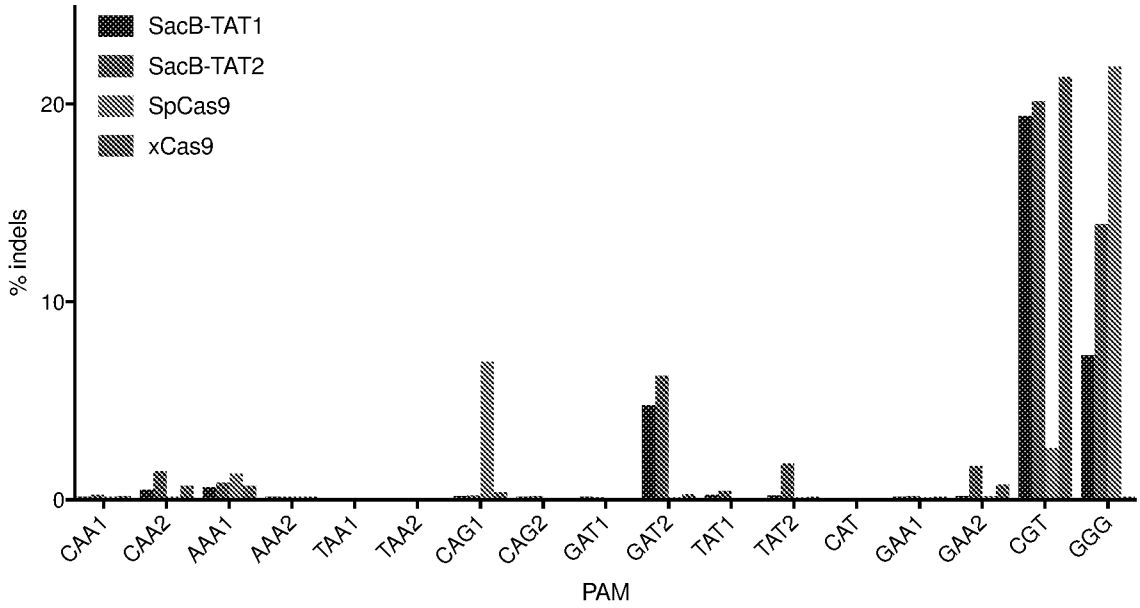


FIG. 30B

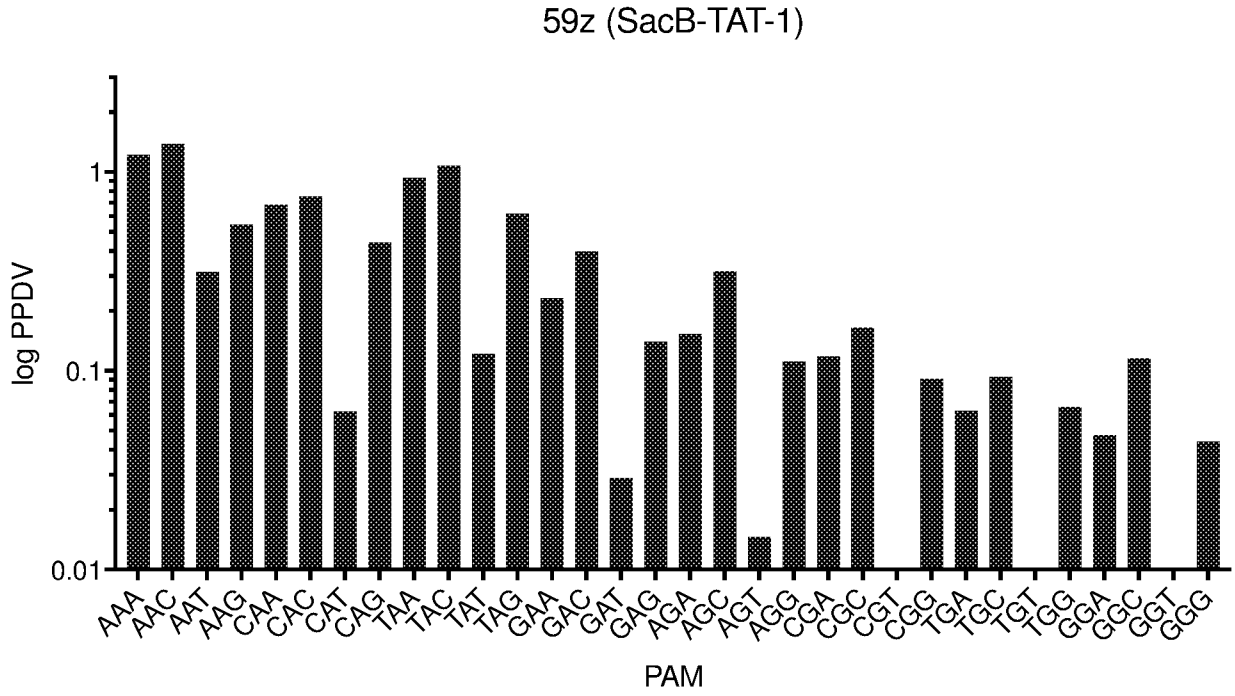
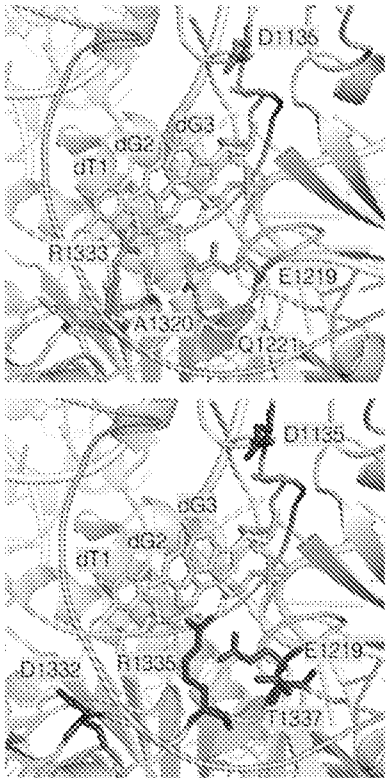
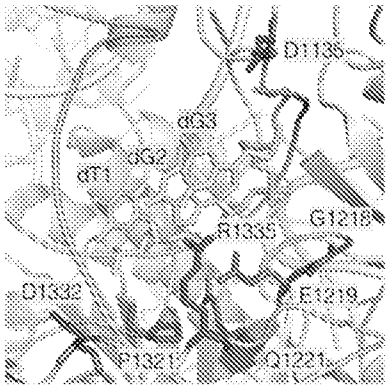


FIG. 31



[illegible]	AAT	ACA	ACT	ATA	ATT	[illegible]	[illegible]
TAA	TAT	TCA	TCT	TTA	TTT	[illegible]	[illegible]
[illegible]	[illegible]	CCA	CCT	CTA	CTT	[illegible]	[illegible]
[illegible]	[illegible]	GCA	GCT	GTA	GTT	[illegible]	[illegible]
AAC	AAG	ACC	ACG	ATC	ATG	[illegible]	AGG
TAC	TAG	TCC	TCG	TTC	TTG	[illegible]	TGG
[illegible]	CAG	CCC	CCG	CTC	CTG	[illegible]	CGG
[illegible]	GAG	GCC	GCG	GTC	GTG	[illegible]	GGG



Positions commonly mutated in variants obtained when evolving on NAN:

D1135
E1219
D1332

FIG. 32

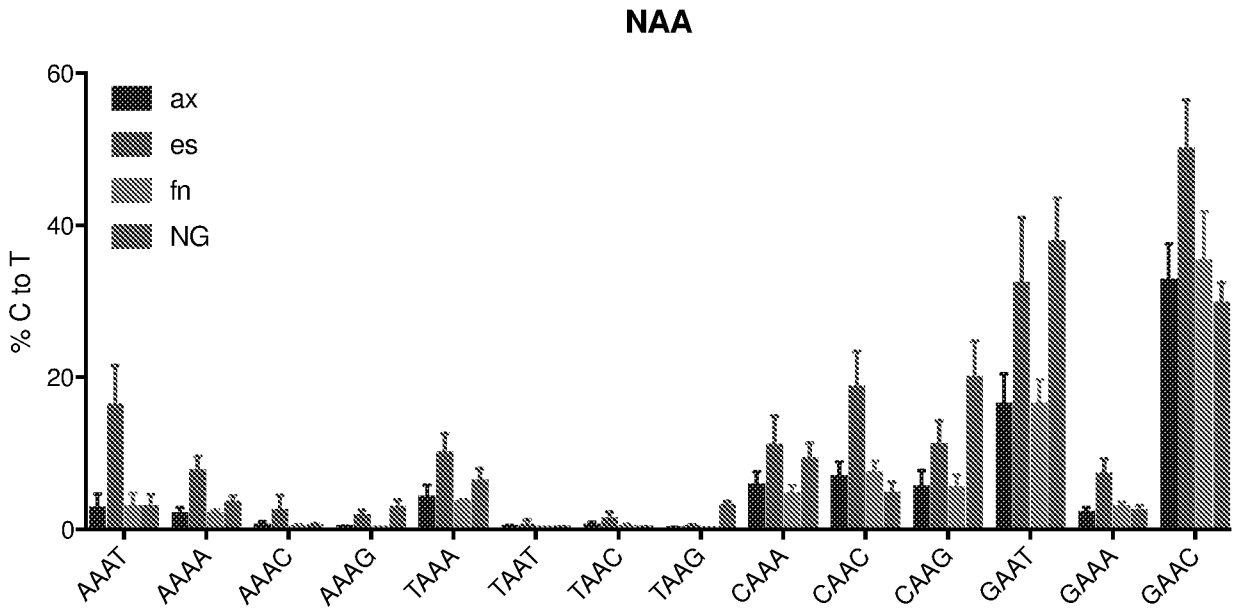


FIG. 33A

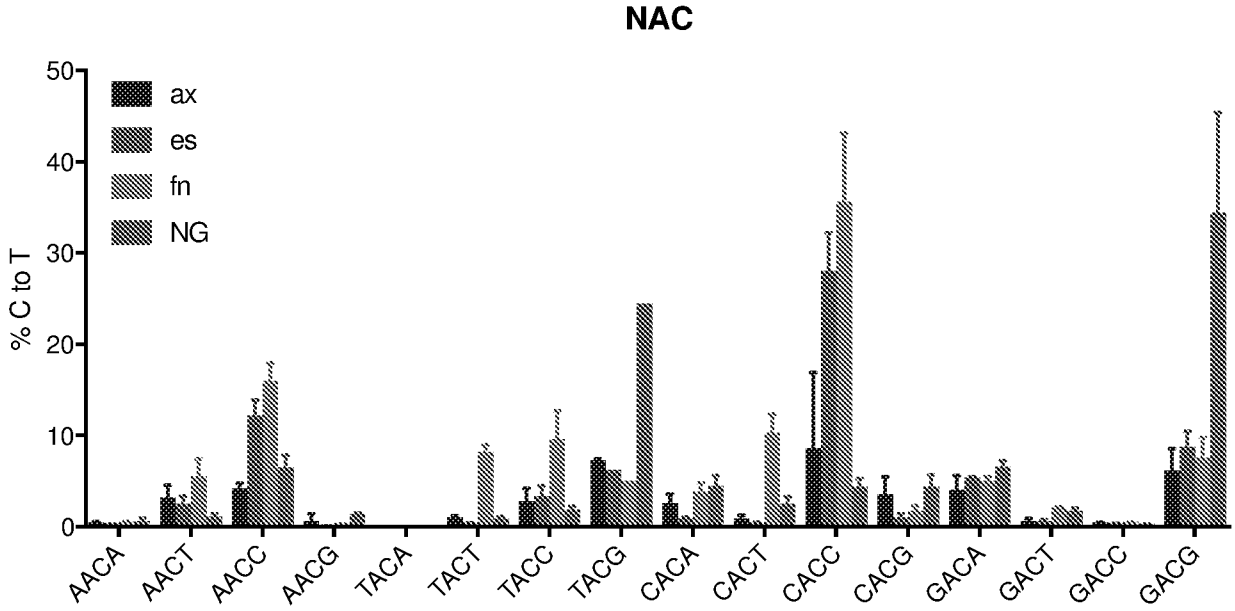


FIG. 33B

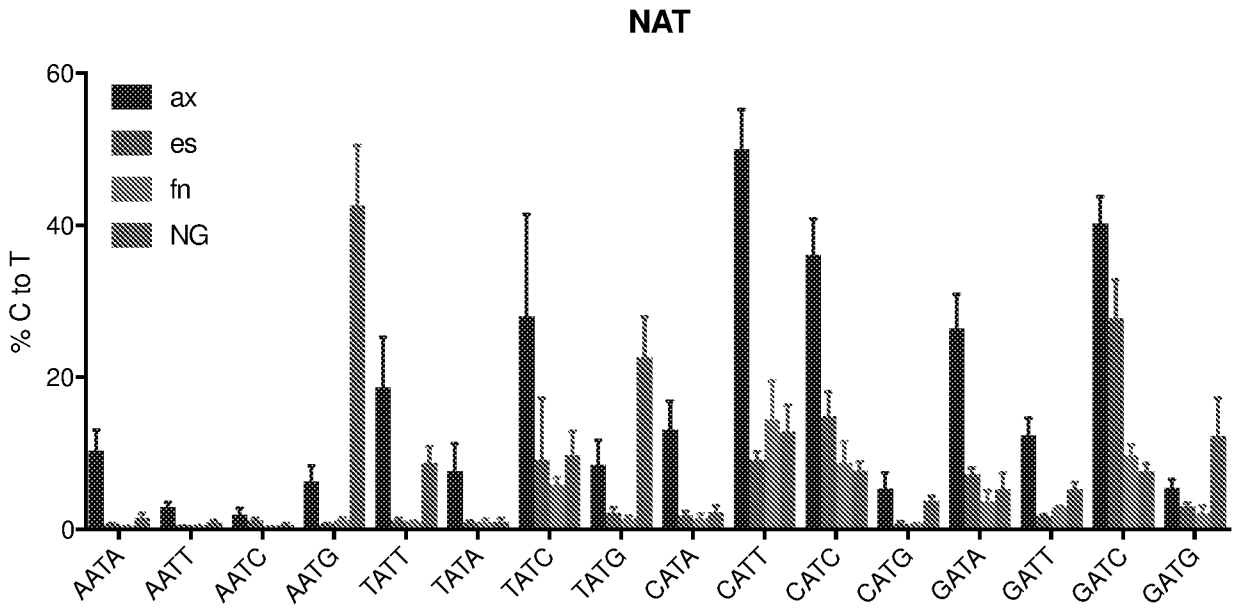


FIG. 33C

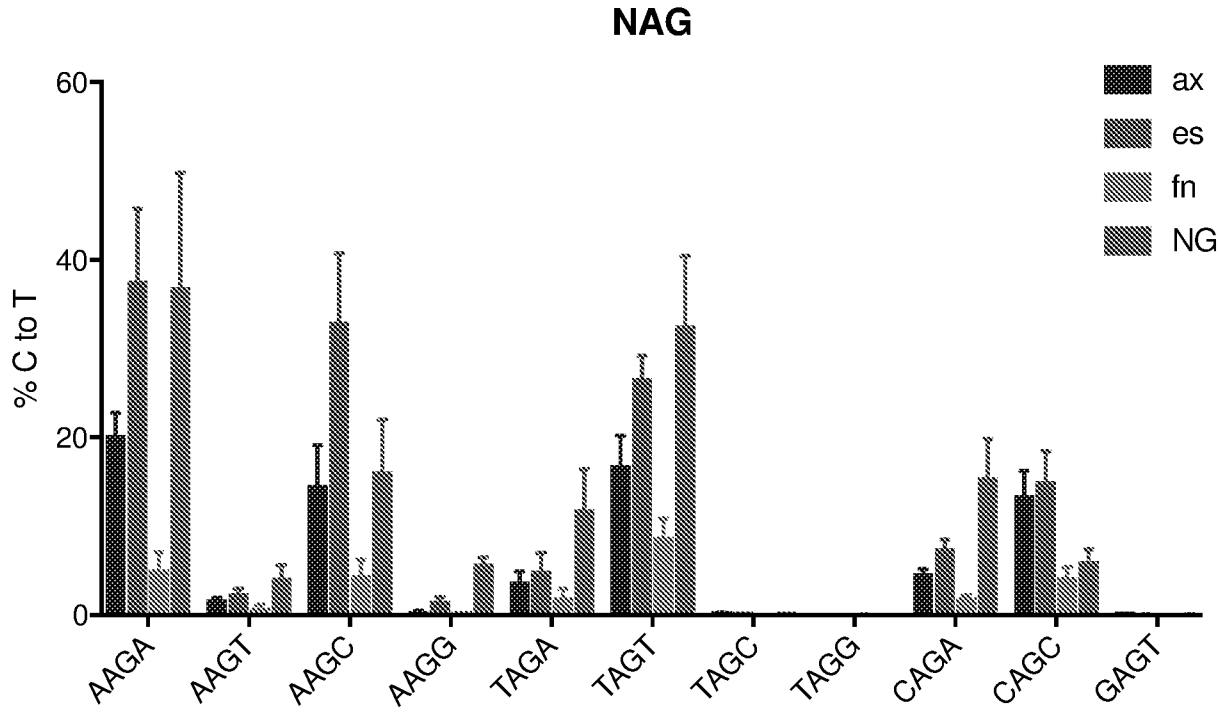


FIG. 33D

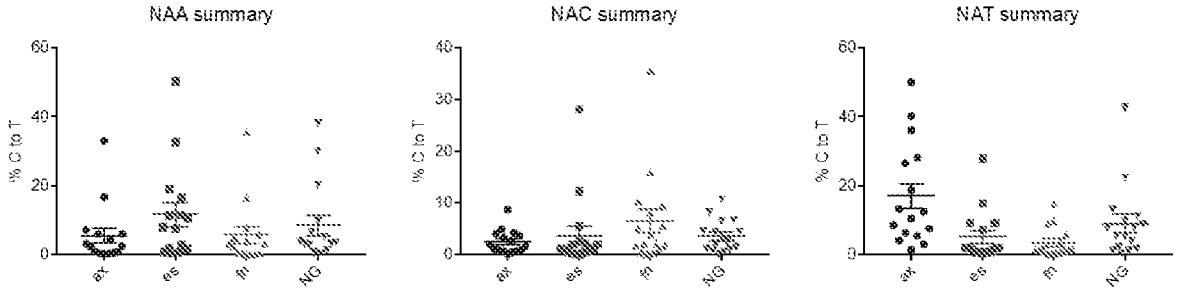


FIG. 34A

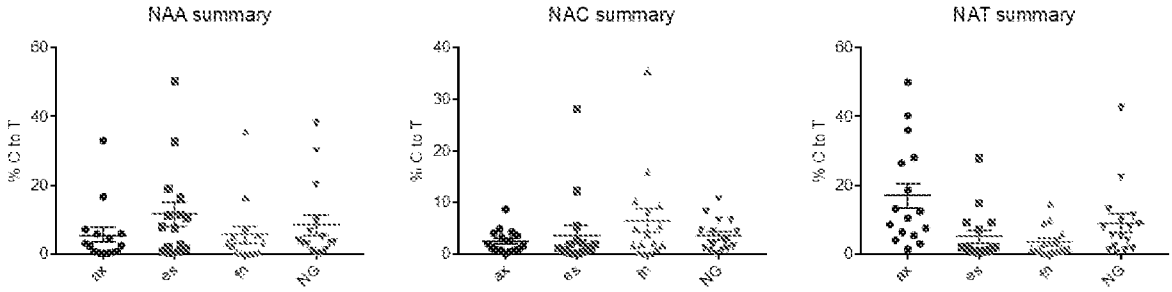


FIG. 34B

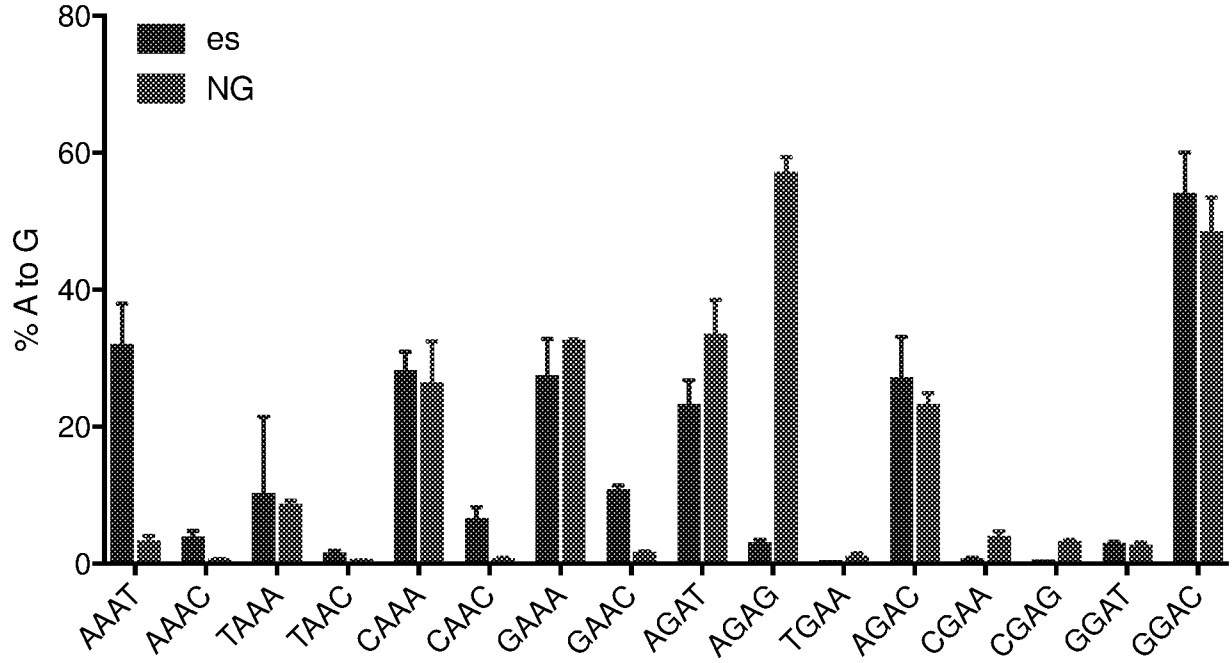


FIG. 35A

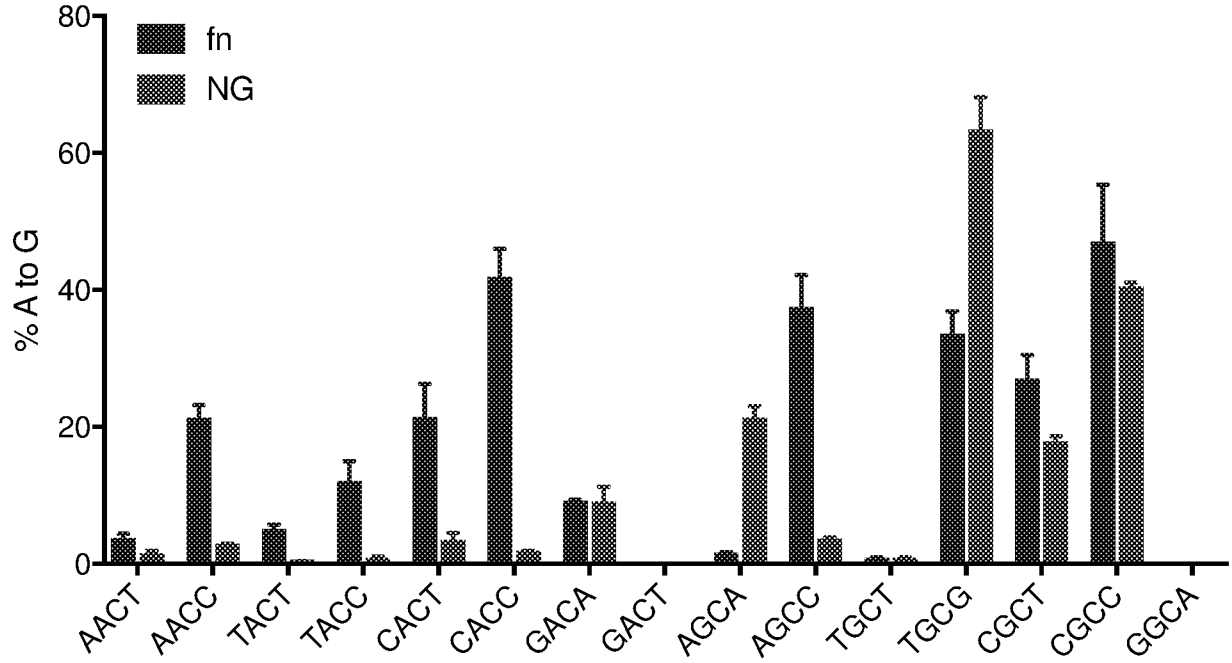


FIG. 35B

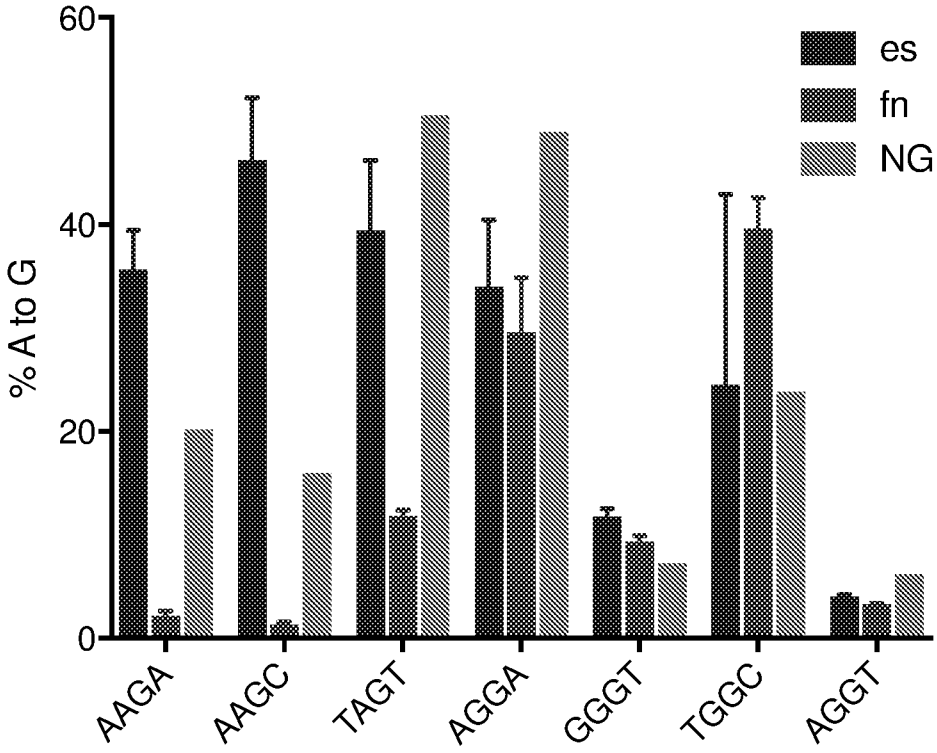


FIG. 35C

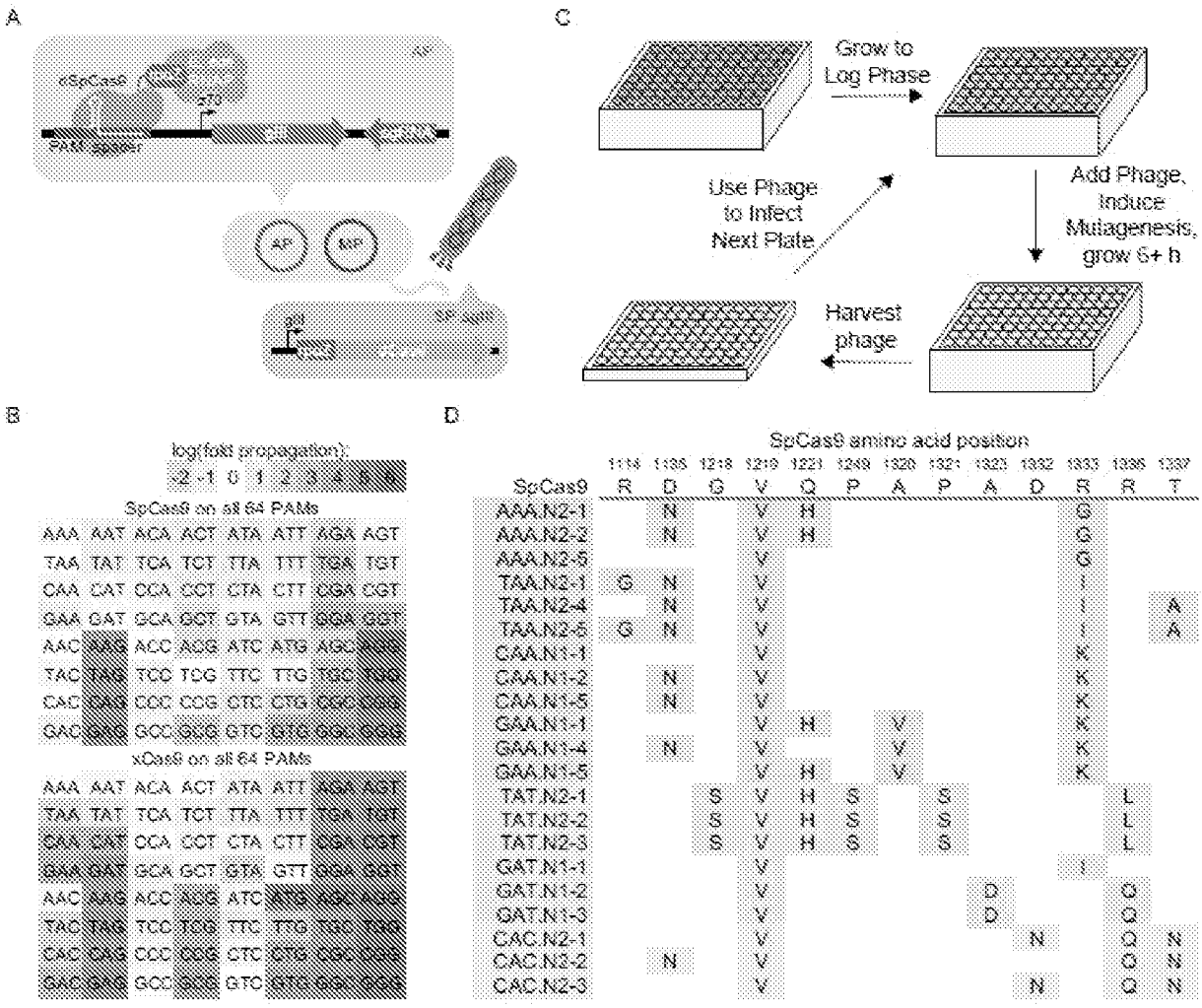

A
xSpCas9
sgRNA
PAM spacer
AP
MP
B
log(fold propagation):
-2 -1 0 1 2 3 4 5 6
SpCas9 on all 64 PAMs
xCas9 on all 64 PAMs
C
Grow to Log Phase
Add Phage, Induce Mutagenesis, grow 6+ h
Harvest phage
Use Phage to Infect Next Plate
D
SpCas9 amino acid position
SpCas9
AAA.N2-1
AAA.N2-2
AAA.N2-5
TAA.N2-1
TAA.N2-4
TAA.N2-5
CAA.N1-1
CAA.N1-2
CAA.N1-5
GAA.N1-1
GAA.N1-4
GAA.N1-5
TAT.N2-1
TAT.N2-2
TAT.N2-3
GAT.N1-1
GAT.N1-2
GAT.N1-3
CAC.N2-1
CAC.N3-2
CAC.N3-3

FIG. 36

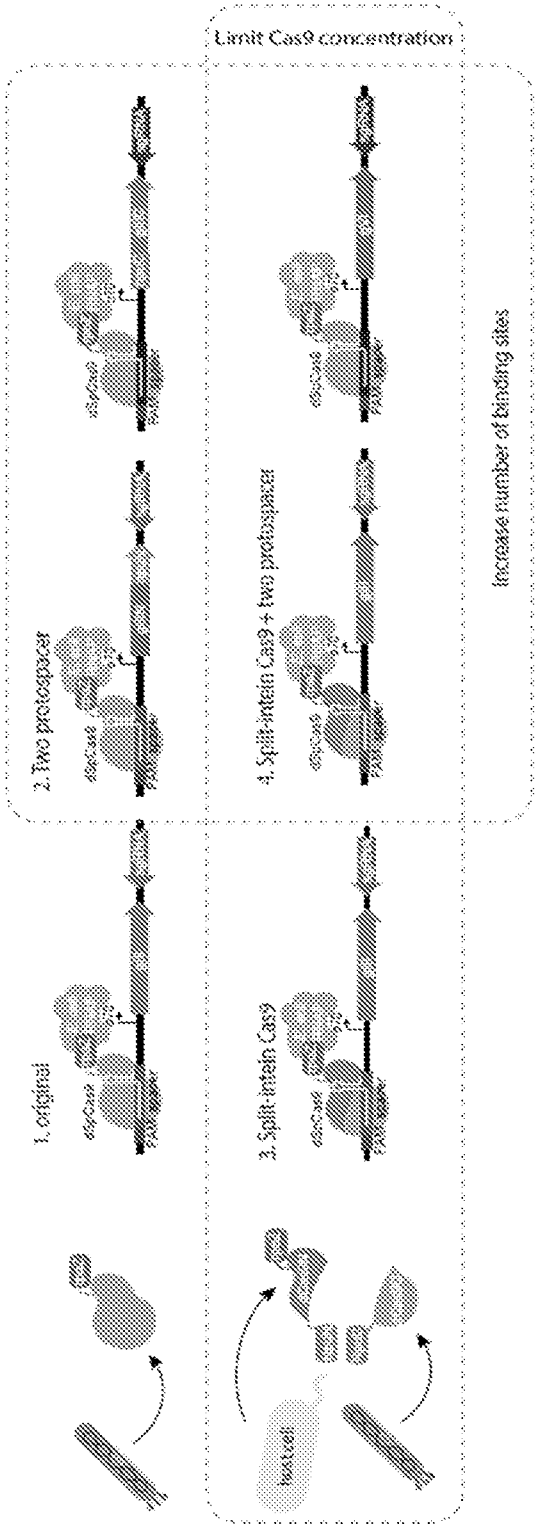

Limit Cas9 concentration
Increase number of binding sites
1. original
2. Two protospacer
3. Split-intein Cas9
4. Split-intein Cas9 + two protospacer
Intein

FIG. 37A

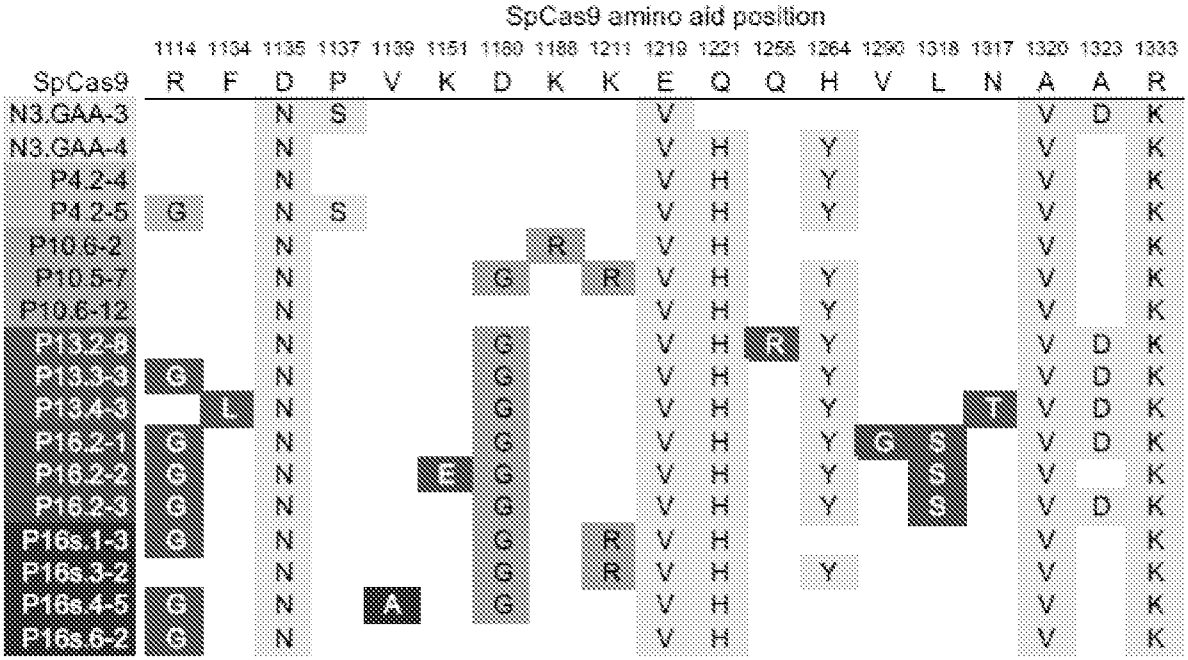


FIG. 37B

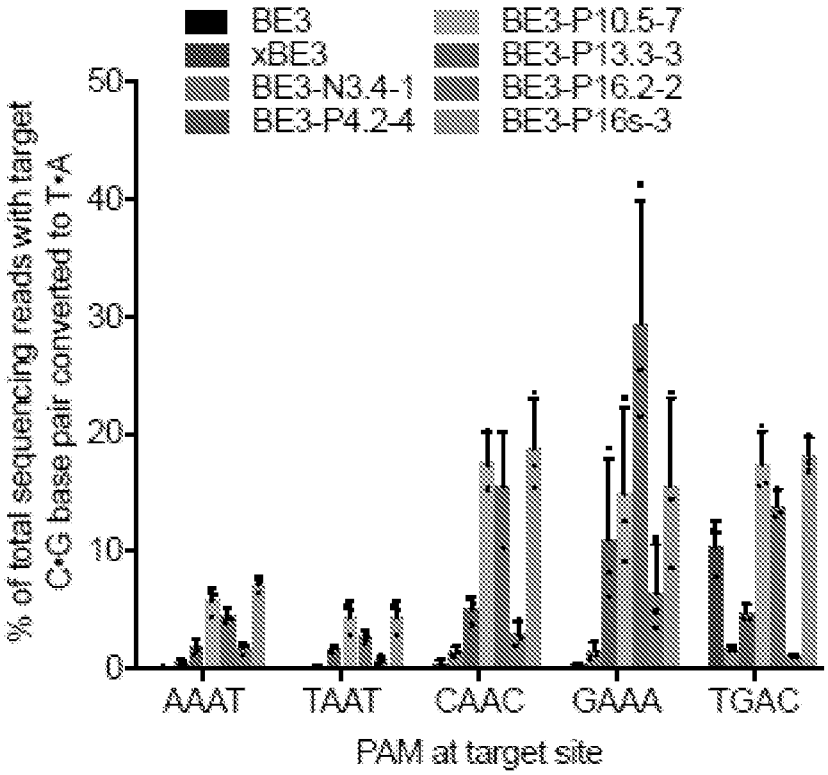
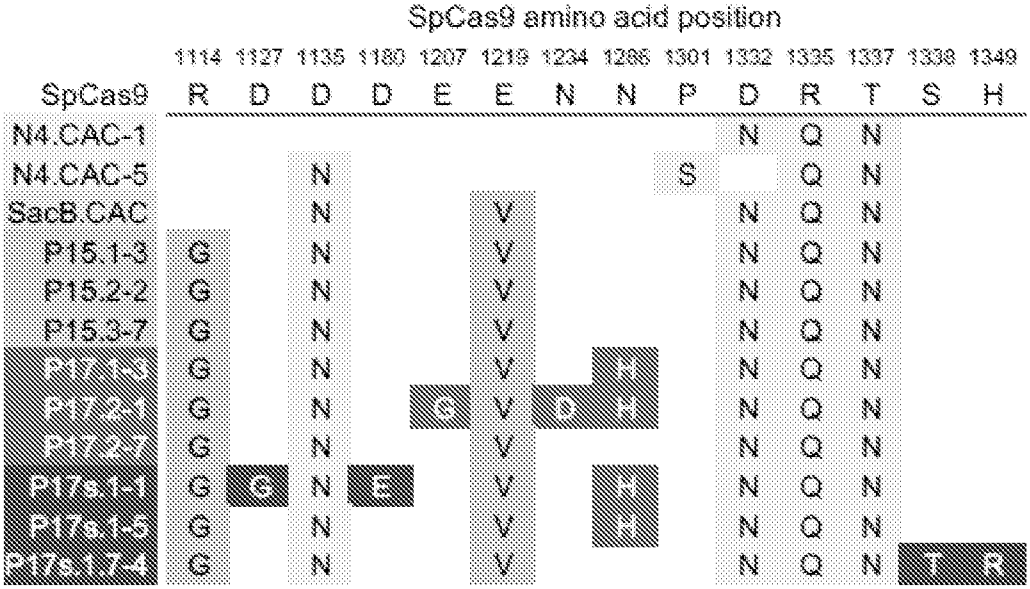


FIG. 37C


SpCas9 amino acid position
1114 1127 1135 1180 1207 1219 1234 1286 1301 1332 1335 1337 1338 1349
SpCas9 R D D D E E N N P D R T S H
N4.CAC-1 N Q N
N4.CAC-5 N S Q N
SacB.CAC N V N Q N
P15.1-3 G N V N Q N
P15.2-2 G N V N Q N
P15.3-7 G N V N Q N
P17.1-3 G N V N Q N
P17.2-1 G N G V D N Q N
P17.3-7 G N V N Q N
P17s.1-1 G G N E V N Q N
P17s.1-5 G N V N Q N
P17s.17-4 G N V N Q N T R

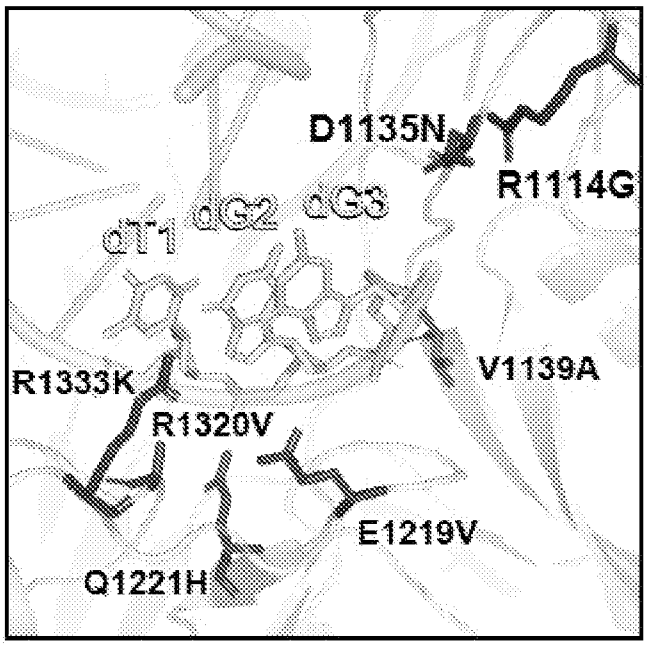

D1135N
R1114G
gT1
gG2
gG3
V1139A
R1333K
R1320V
E1219V
Q1221H

FIG. 37D

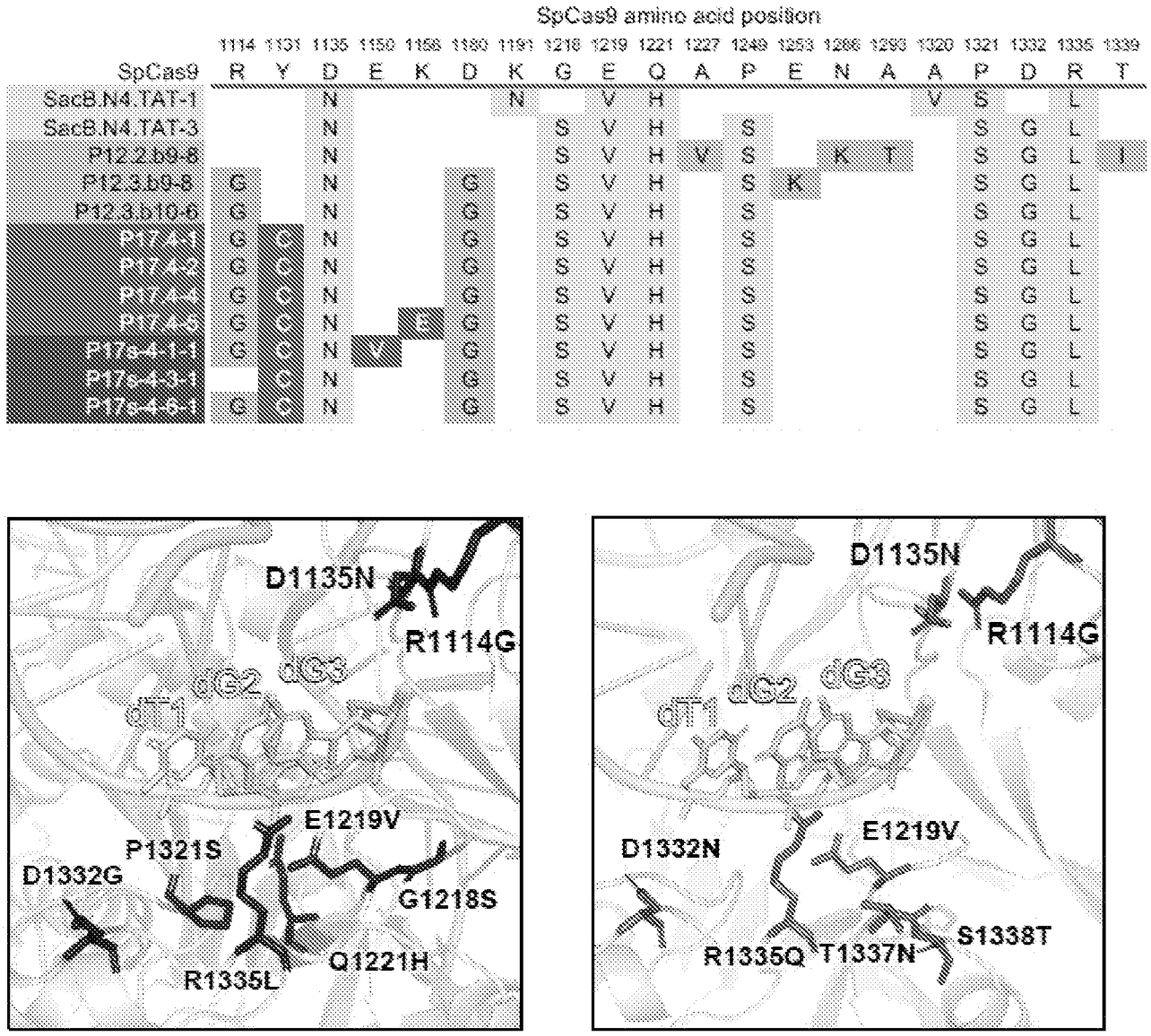


FIG. 37E

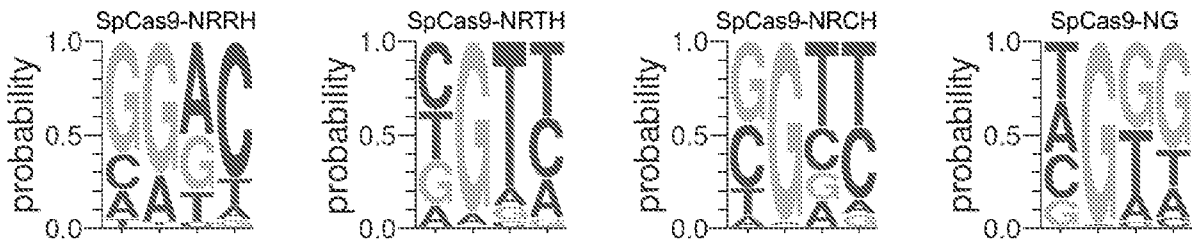


FIG. 38A

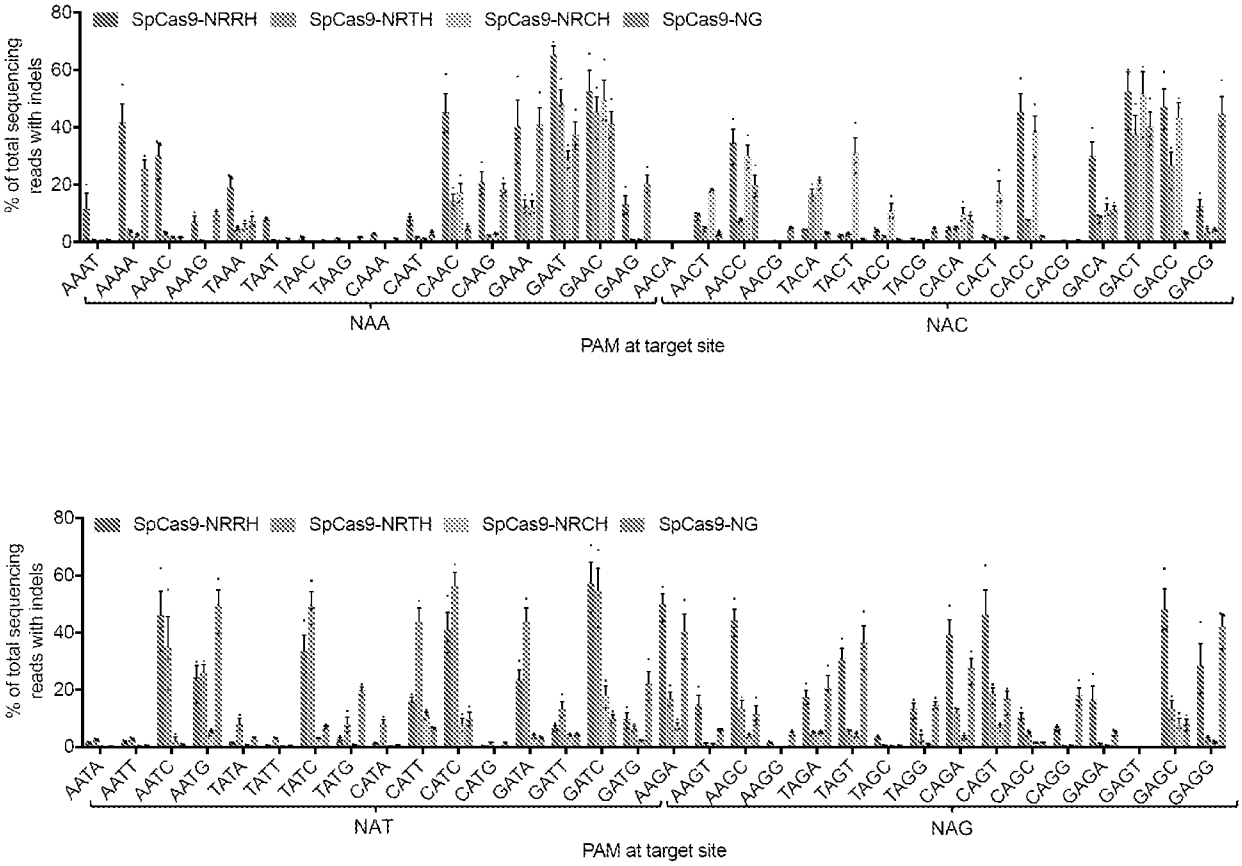

SpCas9-NRRH
SpCas9-NRTH
SpCas9-NRCH
SpCas9-NG
% of total sequencing reads with indels
0
20
40
60
80
AAAT AAAA AAAC AAAG TAAA TAAT TAAC TAAG CAAA CAAT CAAC CAAG GAAA GAAT GAAC GAAG AACA AACT AACC AACG TACA TACT TACC TACG CACA CACT CACC CACG GACA GACT GACC GACG
NAA
NAC
PAM at target site
SpCas9-NRRH
SpCas9-NRTH
SpCas9-NRCH
SpCas9-NG
% of total sequencing reads with indels
0
20
40
60
80
AATA AATT AATC AATG TATA TATT TATC TATG CATA CATT CATC CATG GATA GATT GATC GATG AAGA AAGT AAGC AAGG TAGA TAGT TAGC TAGG CAGA CAGT CAGC CAGG GAGA GAGT GAGC GAGG
NAT
NAG
PAM at target site

FIG. 38B

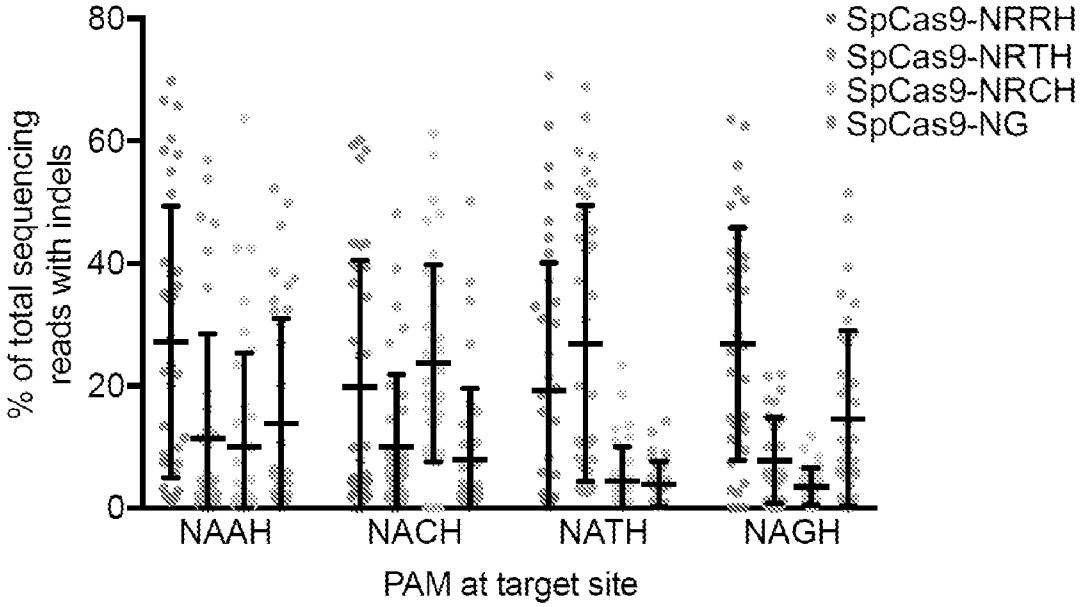


FIG. 38C

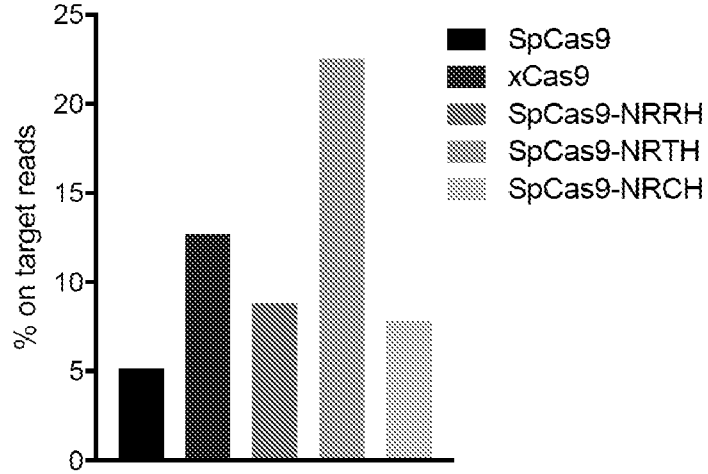


FIG. 38D

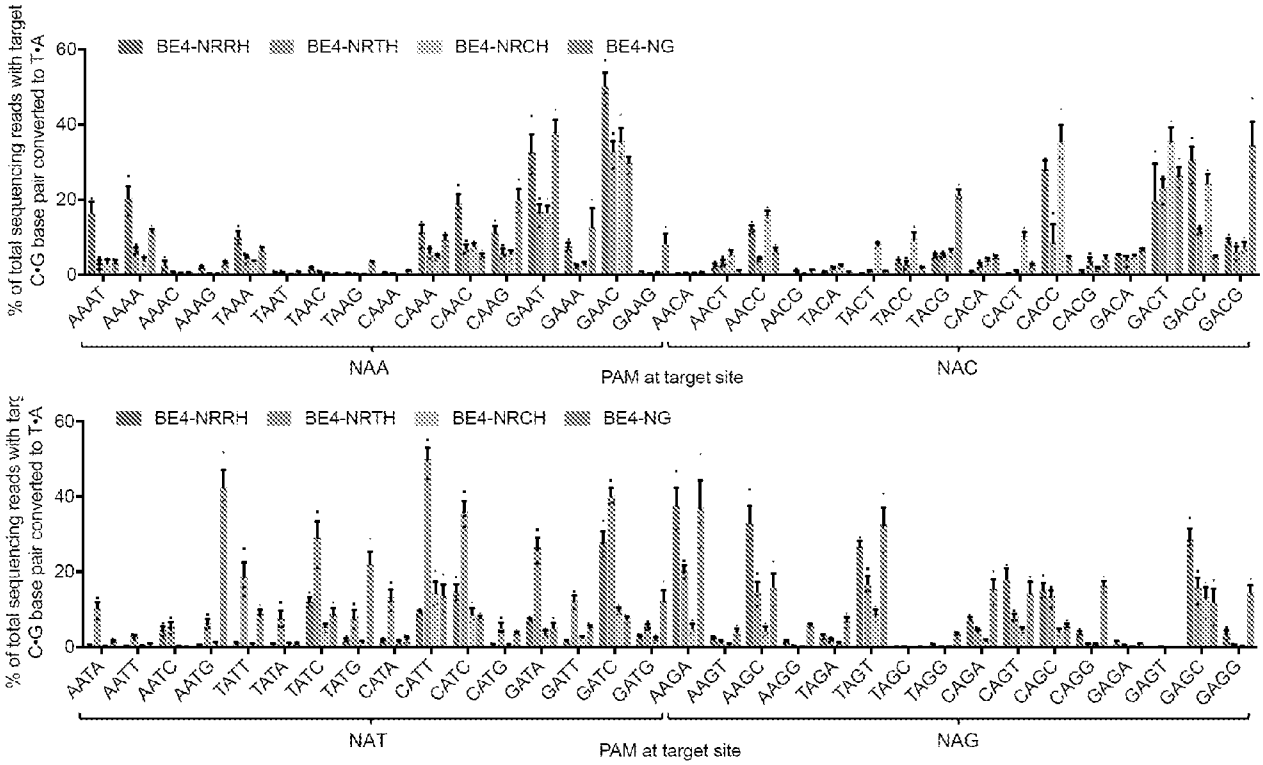

% of total sequencing reads with target C•G base pair converted to T•A
BE4-NRRH
BE4-NRTH
BE4-NRCH
BE4-NG
60
40
20
0
AAAT AAAA AAAC AAAG TAAA TAAT TAAC TAAG CAAA CAAA CAAC CAAG GAAT GAAA GAAC GAAG AACA AACT AACC AACG TACA TACT TACC TACG CACA CACT CACC CACG GACA GACT GACC GACG
NAA
PAM at target site
NAC
% of total sequencing reads with targ C•G base pair converted to T•A
BE4-NRRH
BE4-NRTH
BE4-NRCH
BE4-NG
60
40
20
0
AATA AATT AATC AATG TATT TATA TATC TATG CATA CATT CATC CATG GATA GATT GATC GATG AAGA AAGT AAGC AAGG TAGA TAGT TAGC TAGG CAGA CAGT CAGC CAGG GAGA GAGT GAGC GAGG
NAT
PAM at target site
NAG

FIG. 39A

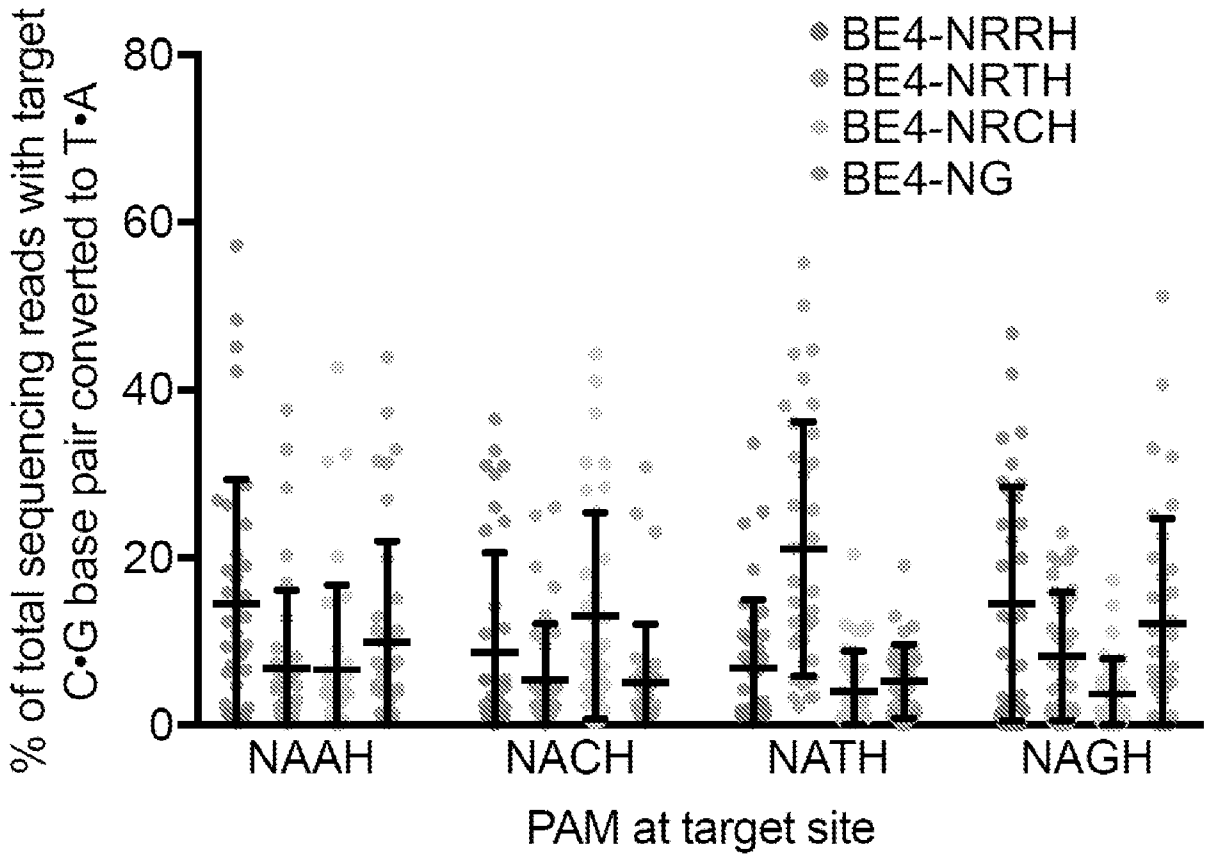


FIG. 39B

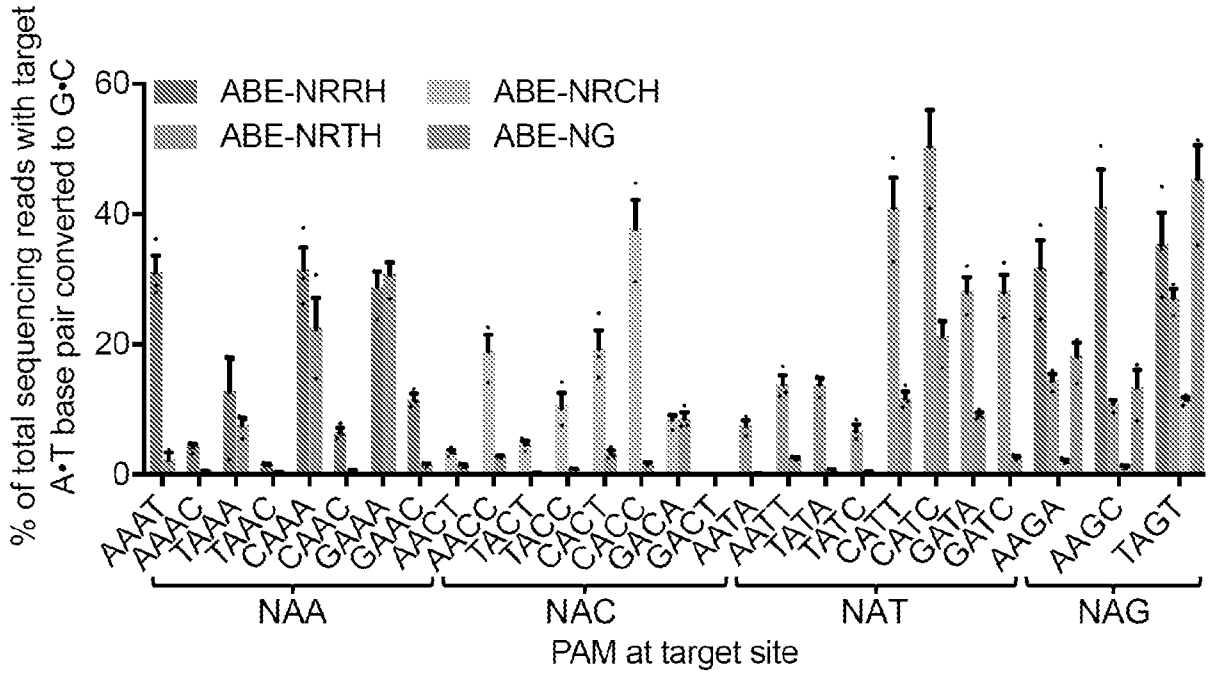


FIG. 39C

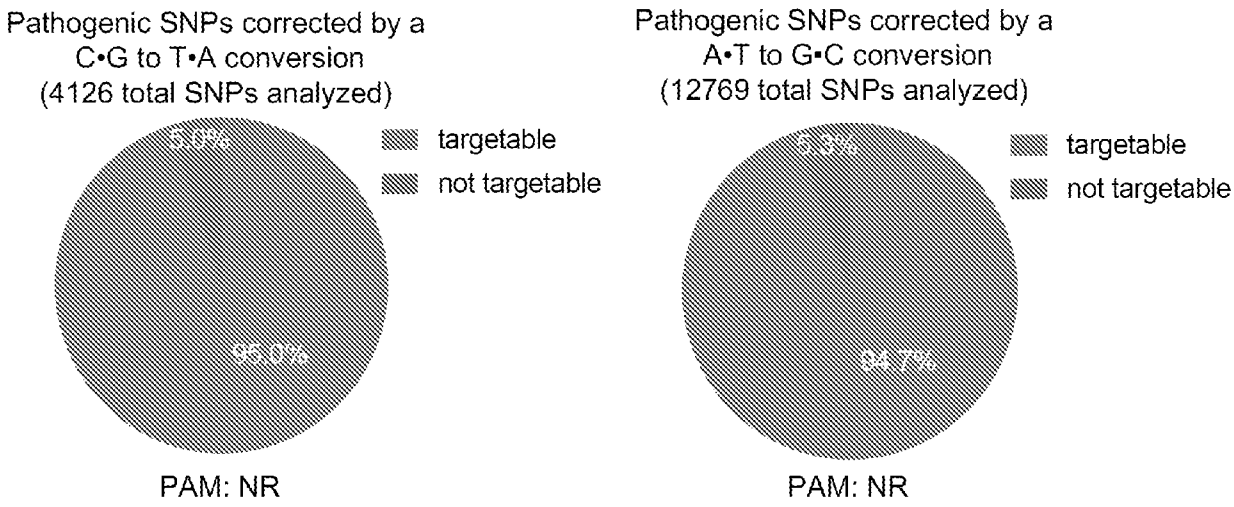


FIG. 39D

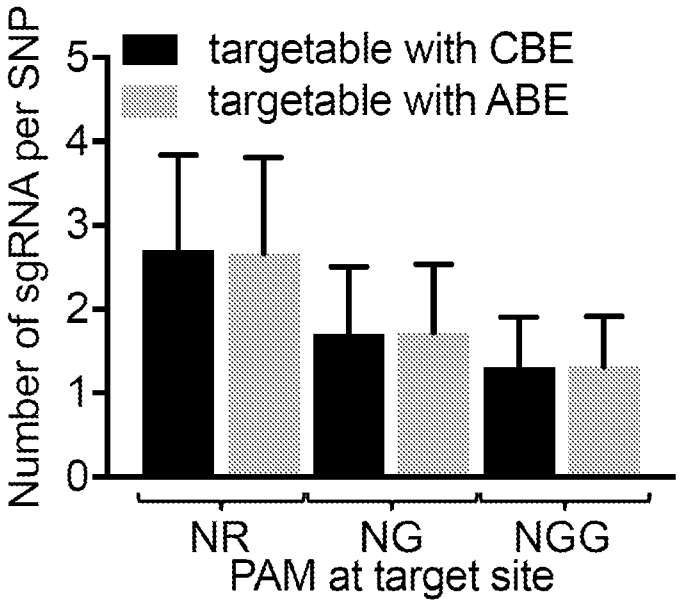


FIG. 39E

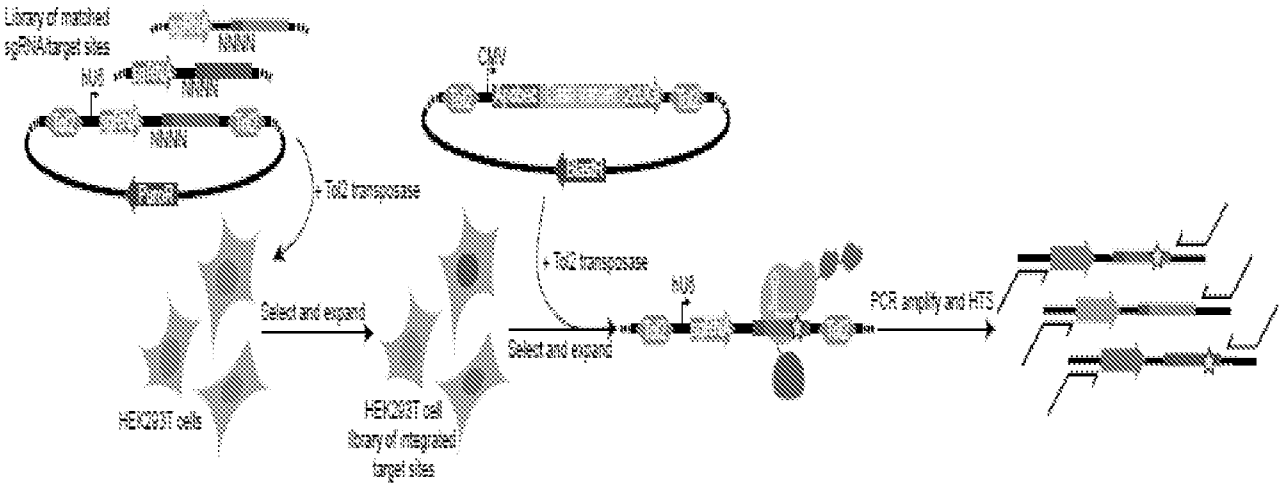

Library of matched sgRNA-target sites
NNNN
NNNN
NNNN
NNNN
NJS
CMV
+ Tol2 transposase
Select and expand
HEK293T cells
HEK293T cell library of integrated target sites
+ Tol2 transposase
Select and expand
NJS
PCR amplify and HTS

FIG. 40A

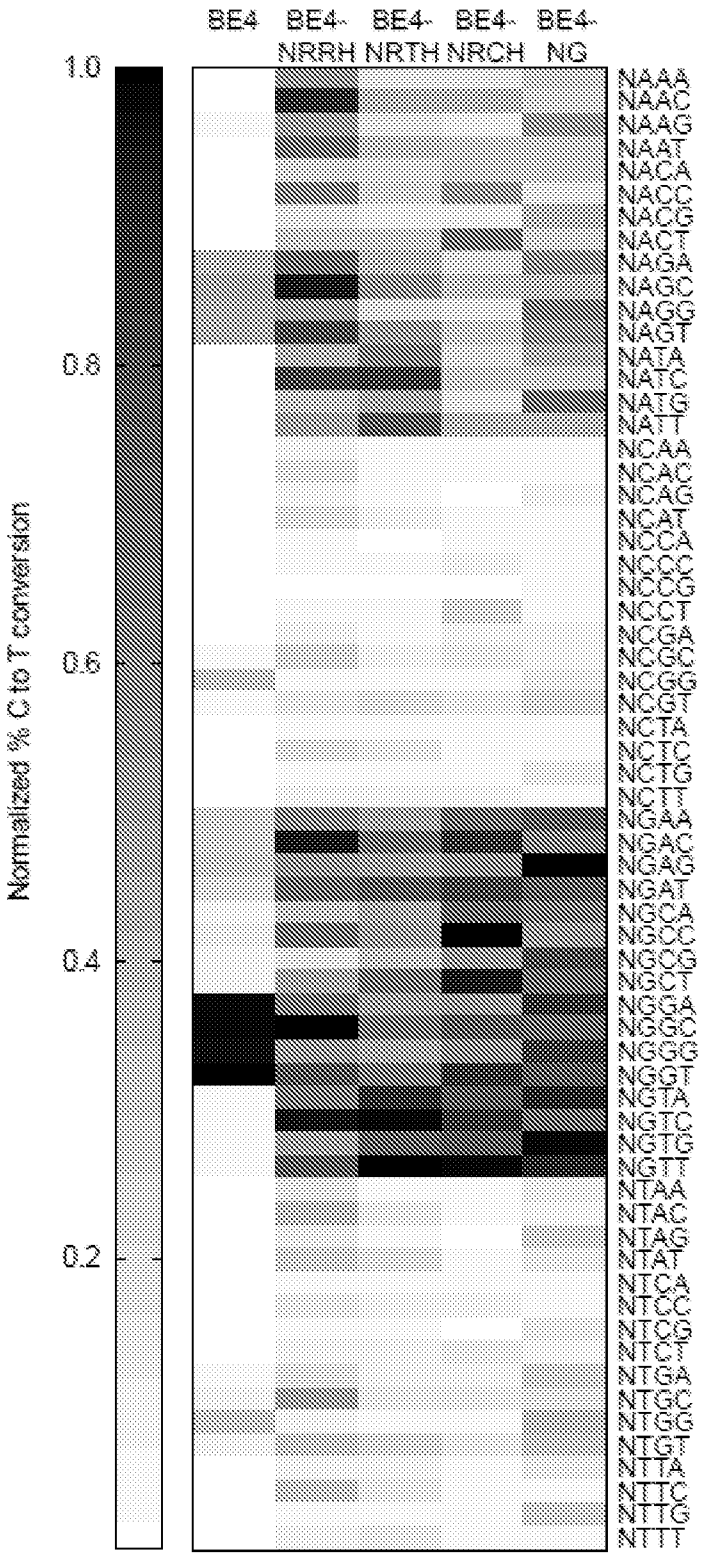

BE4
BE4-NRRH
BE4-NRTH
BE4-NRCH
BE4-NG
1.0
0.8
0.6
0.4
0.2
Normalized % C to T conversion

FIG. 40B

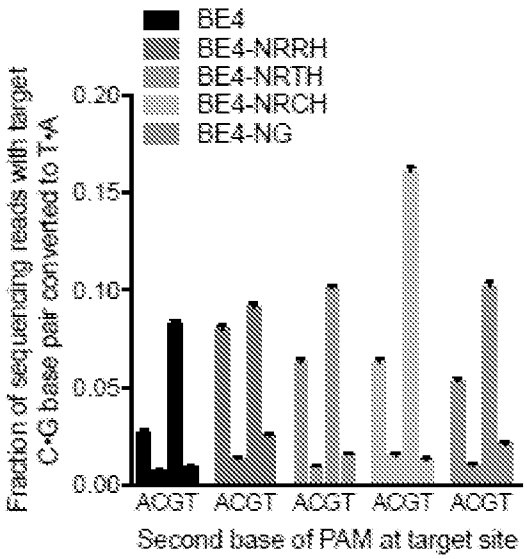


FIG. 40C

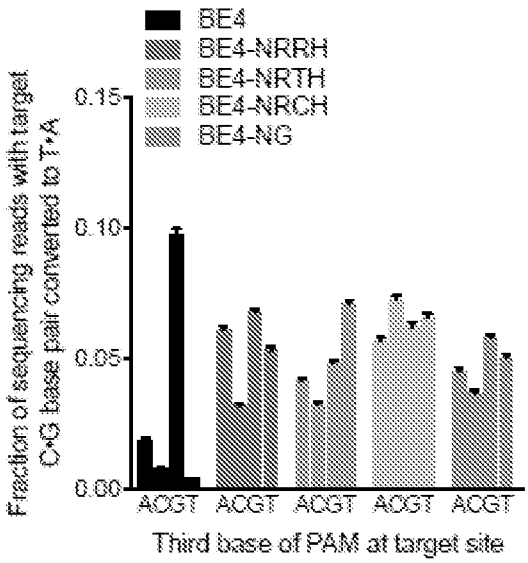


FIG. 40D

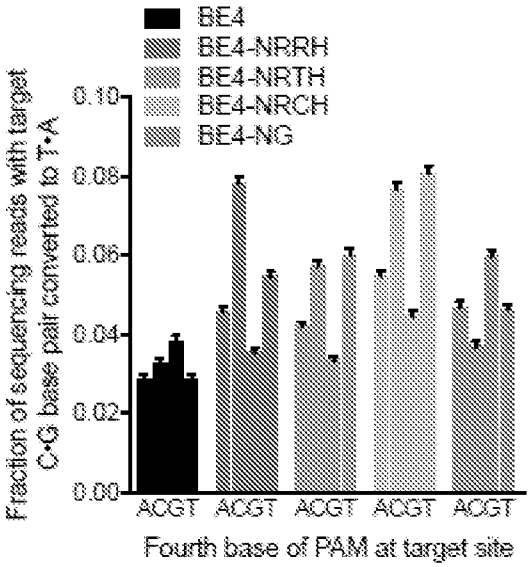


FIG. 40E

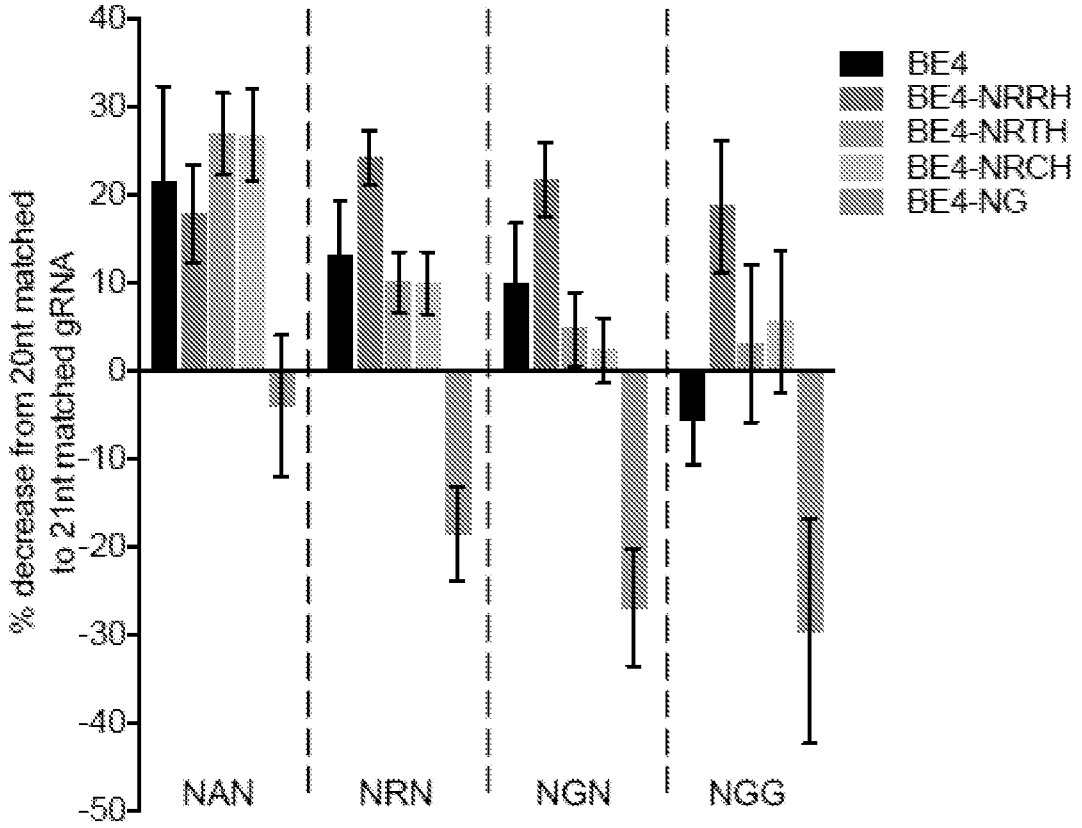

% decrease from 20nt matched to 21nt matched gRNA
40
30
20
10
0
-10
-20
-30
-40
-50
NAN
NRN
NGN
NGG
BE4
BE4-NRRH
BE4-NRTH
BE4-NRCH
BE4-NG

FIG. 40F

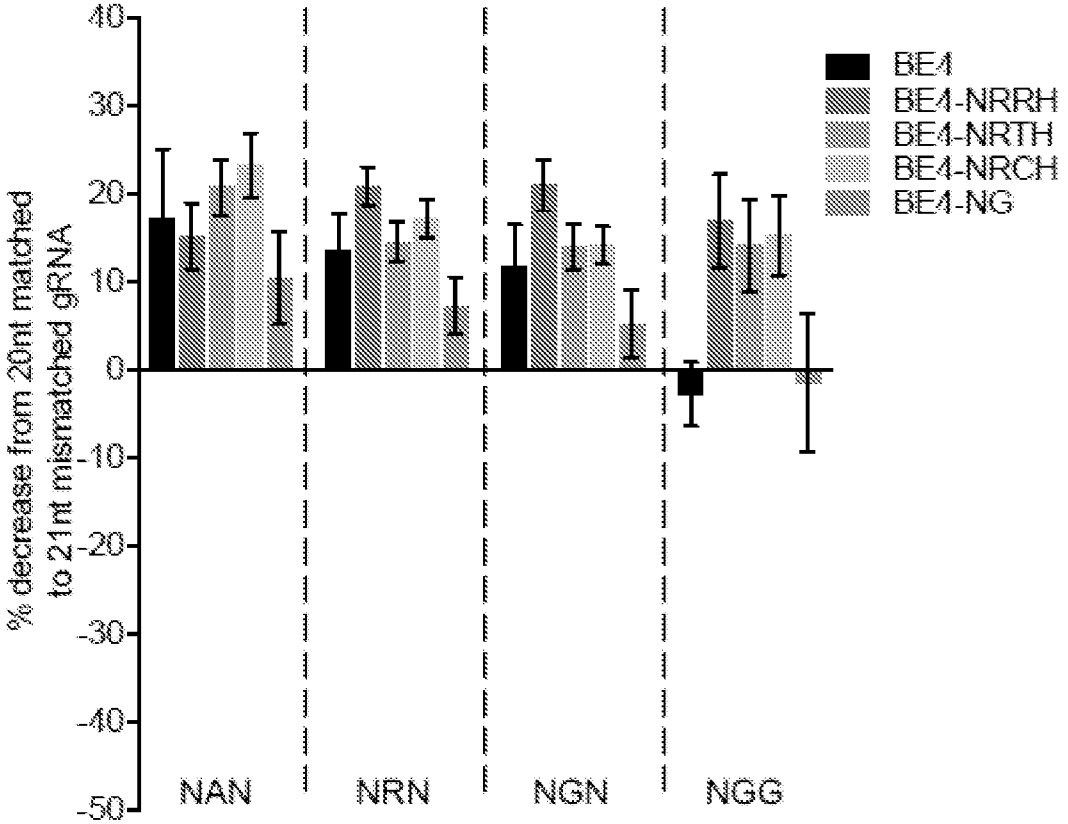


FIG. 40G

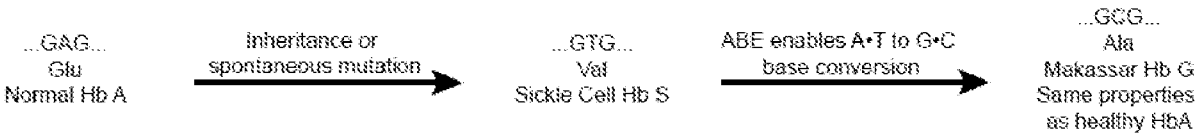


FIG. 41A

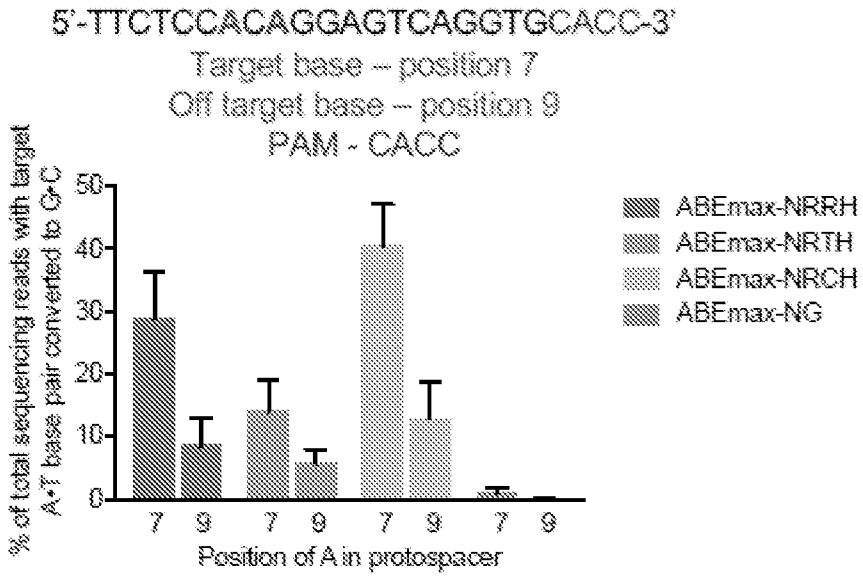


FIG. 41B

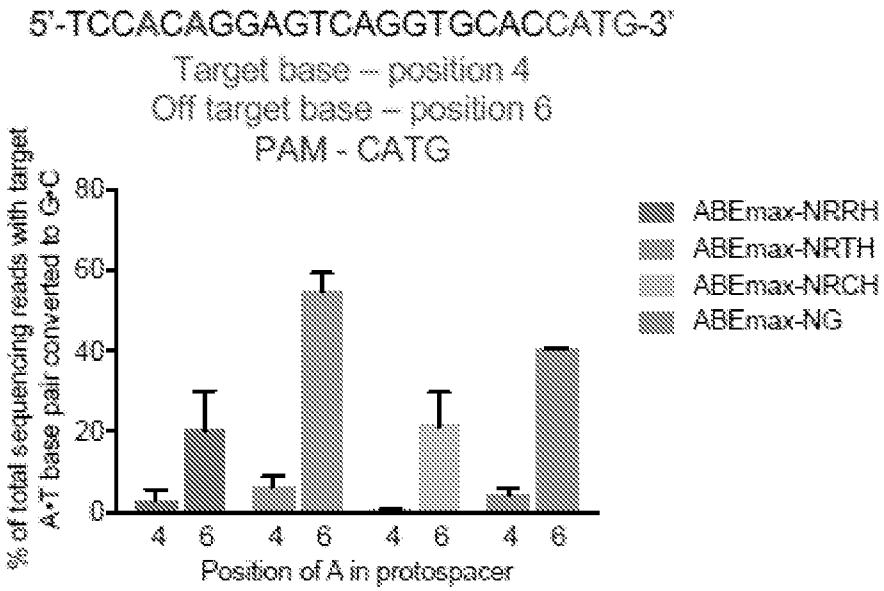
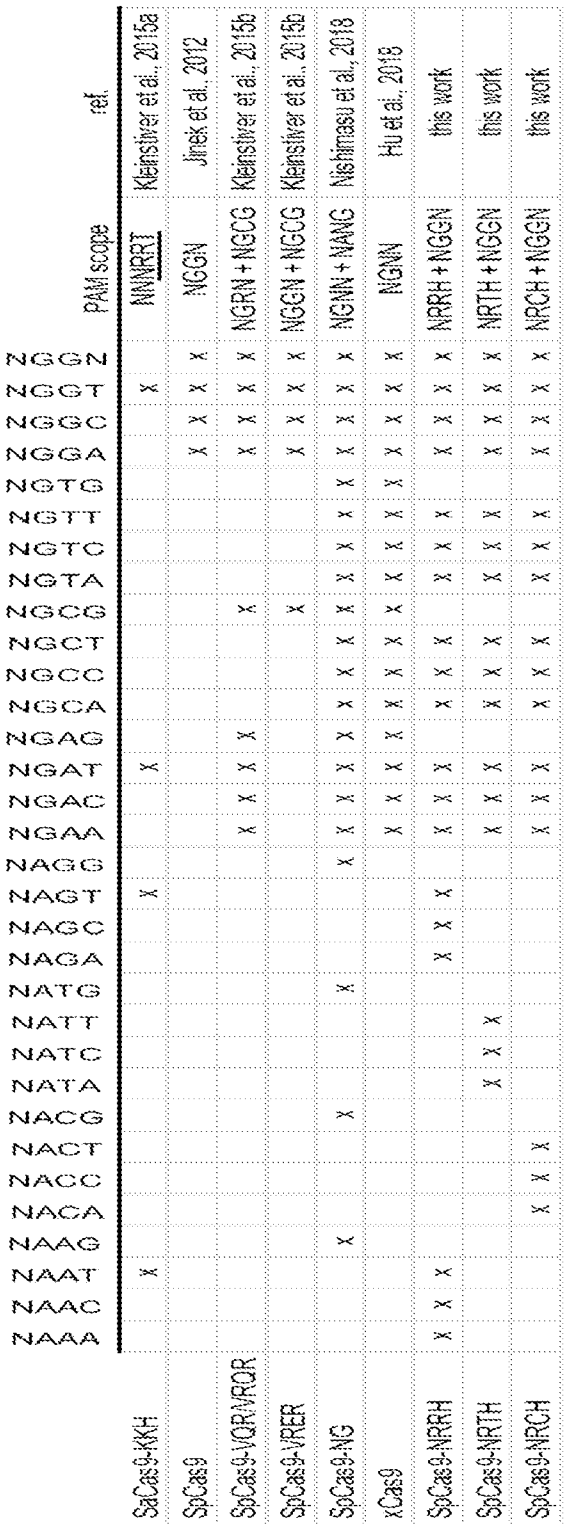


FIG. 41C



	PAM scope	ref.
SaCas9-KKH	NNNRRT	Kleinstiver et al., 2015a
SpCas9	NGGN	Jinek et al., 2012
SpCas9-VQR/VRQR	NGRN + NGCG	Kleinstiver et al., 2015b
SpCas9-VRER	NGCN + NGCG	Kleinstiver et al., 2015b
SpCas9-NG	NGNN + NANG	Nishimasu et al., 2018
xCas9	NGNN	Hu et al., 2018
SpCas9-NRRH	NRRH + NGGN	this work
SpCas9-NRTH	NRTH + NGGN	this work
SpCas9-NRCH	NRCH + NGGN	this work

FIG. 42

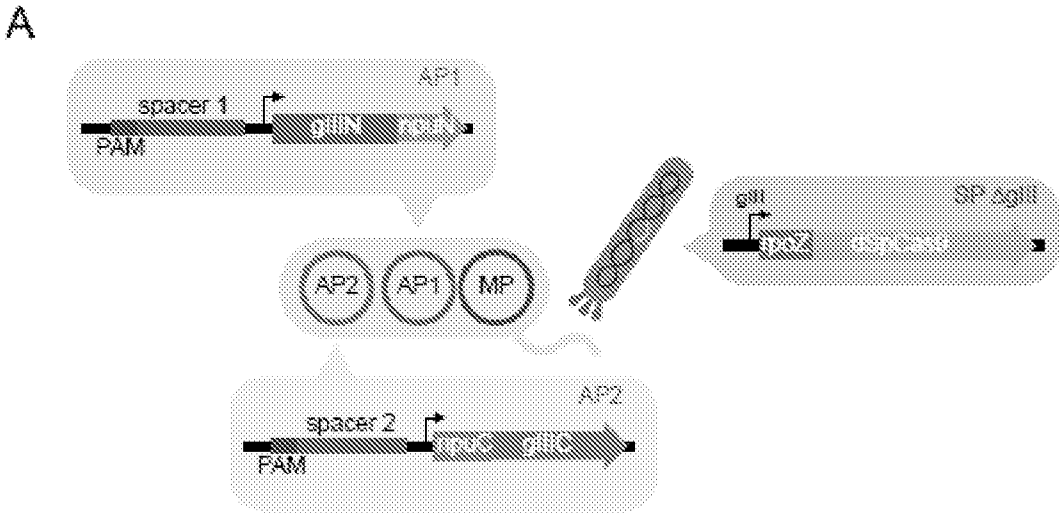


FIG. 43A

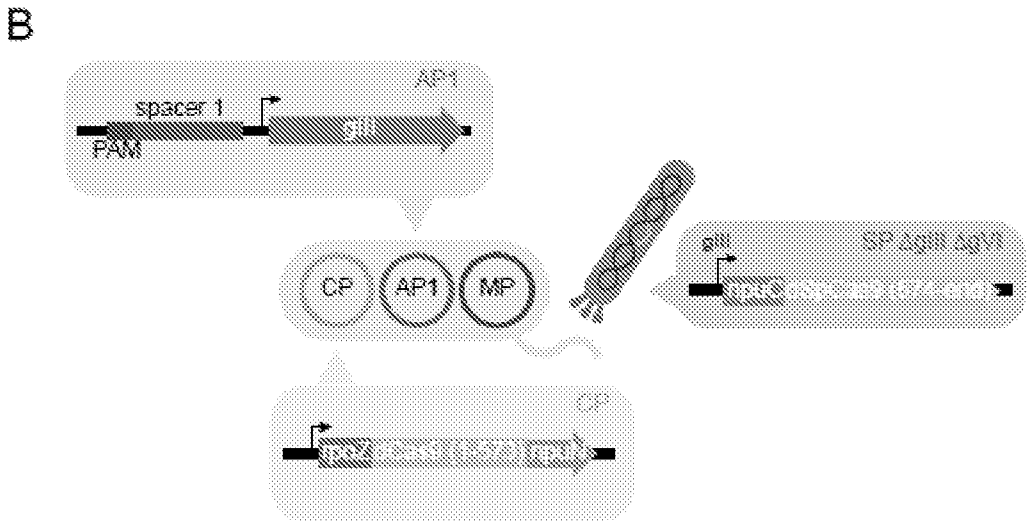


FIG. 43B

C

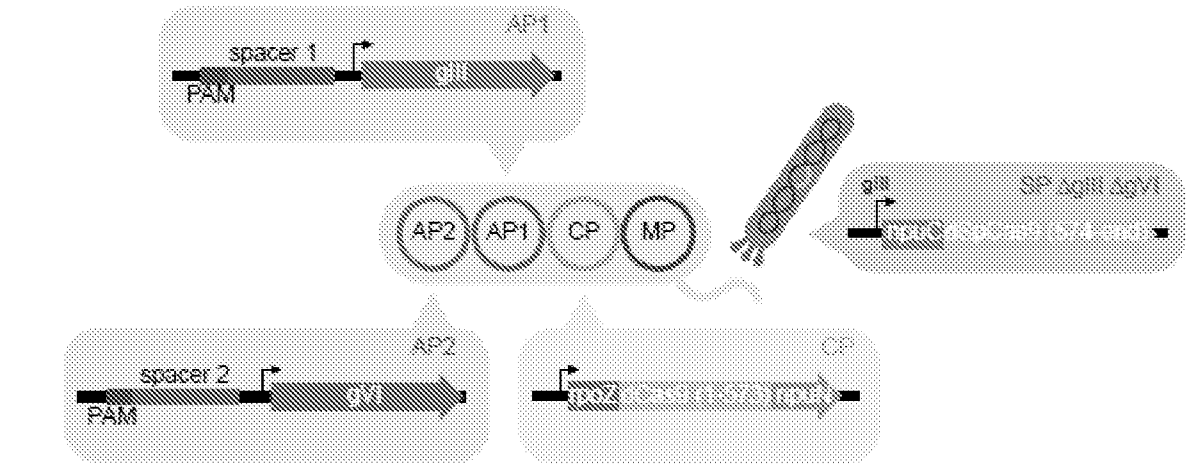


FIG. 43C

D

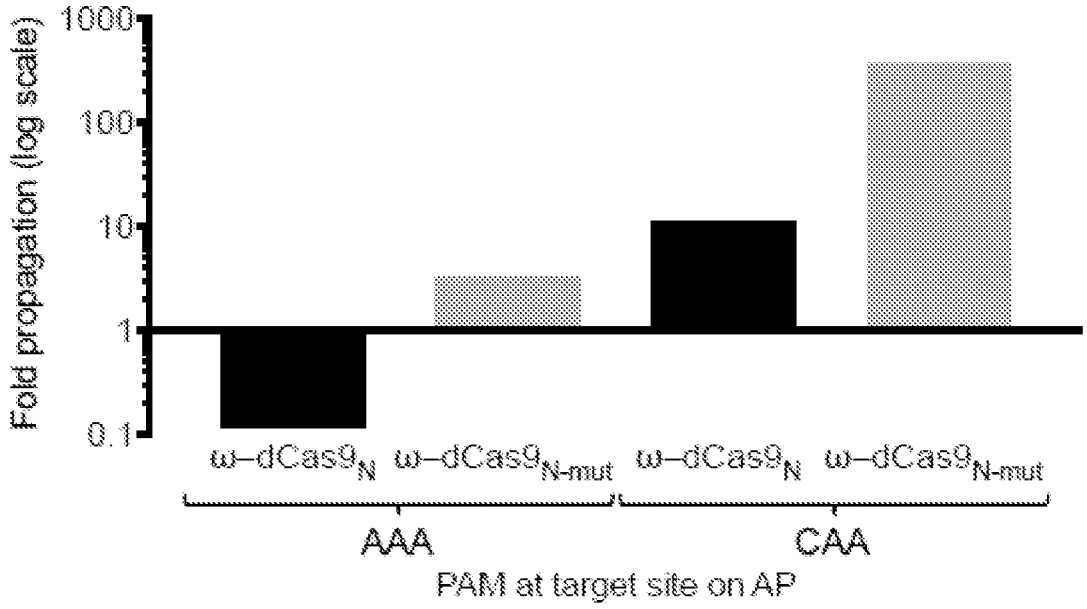


FIG. 43D

E

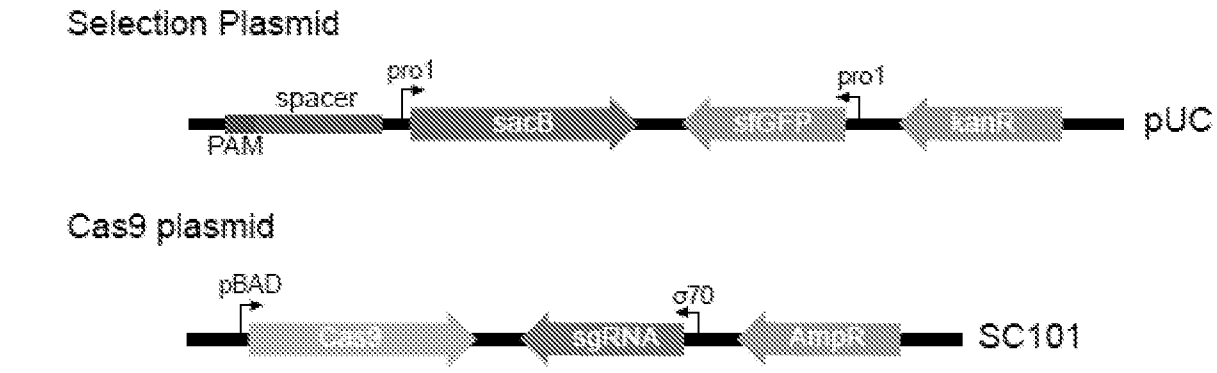


FIG. 43E

F

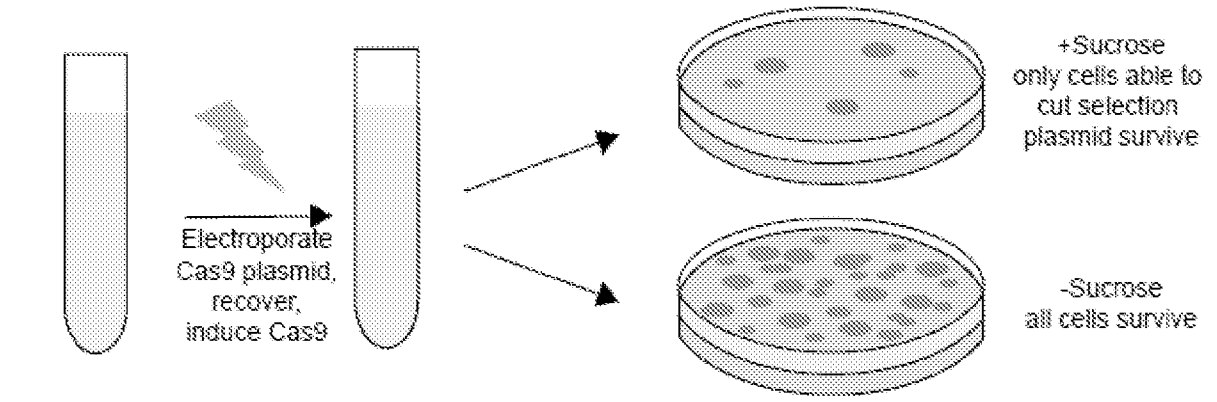


FIG. 43F

A

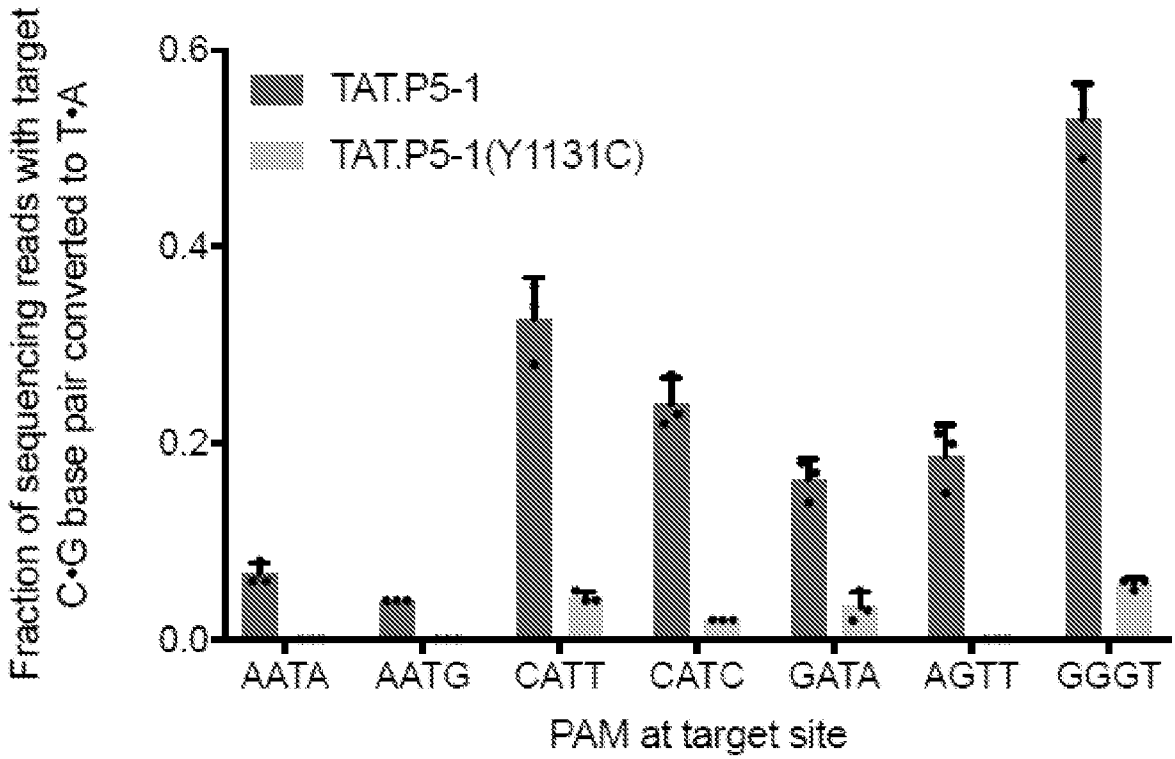

Fraction of sequencing reads with target C•G base pair converted to T•A
0.6
0.4
0.2
0.0
TAT.P5-1
TAT.P5-1(Y1131C)
AATA
AATG
CATT
CATC
GATA
AGTT
GGGT
PAM at target site

FIG. 44A

B

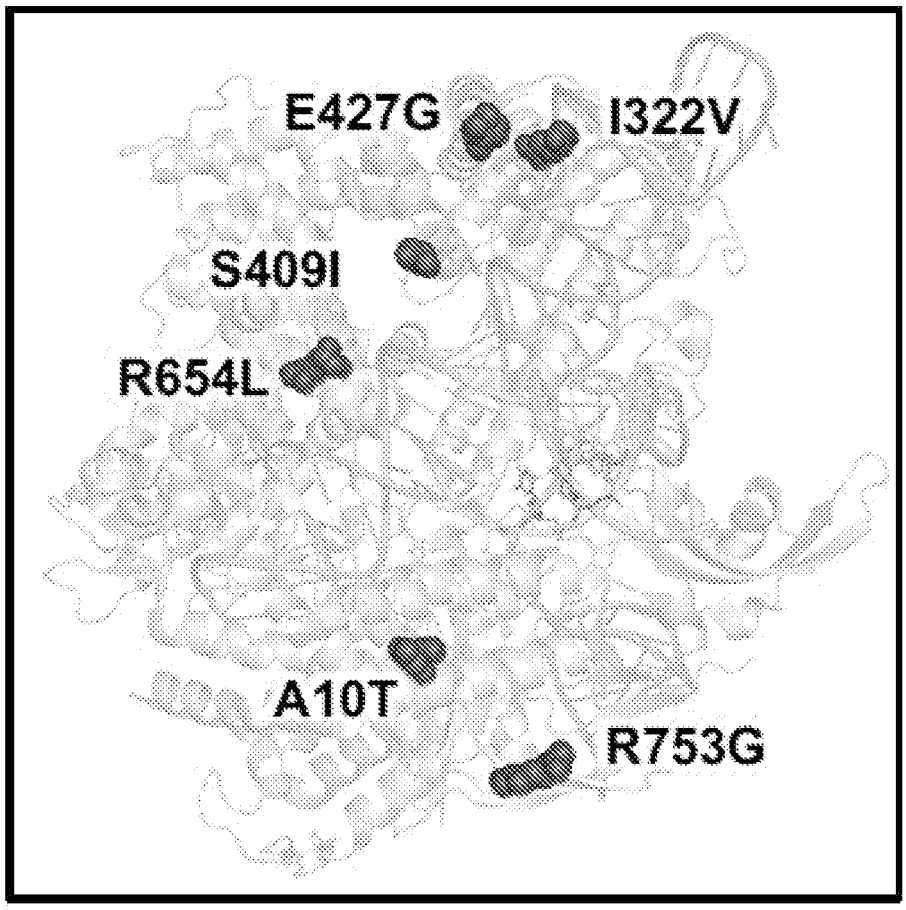


FIG. 44B

C

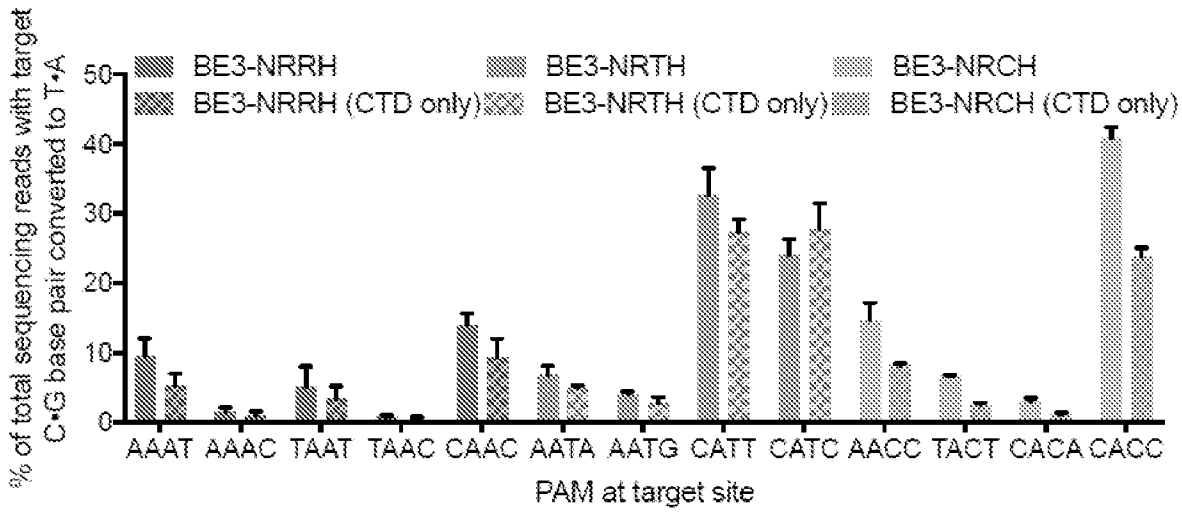


FIG. 44C

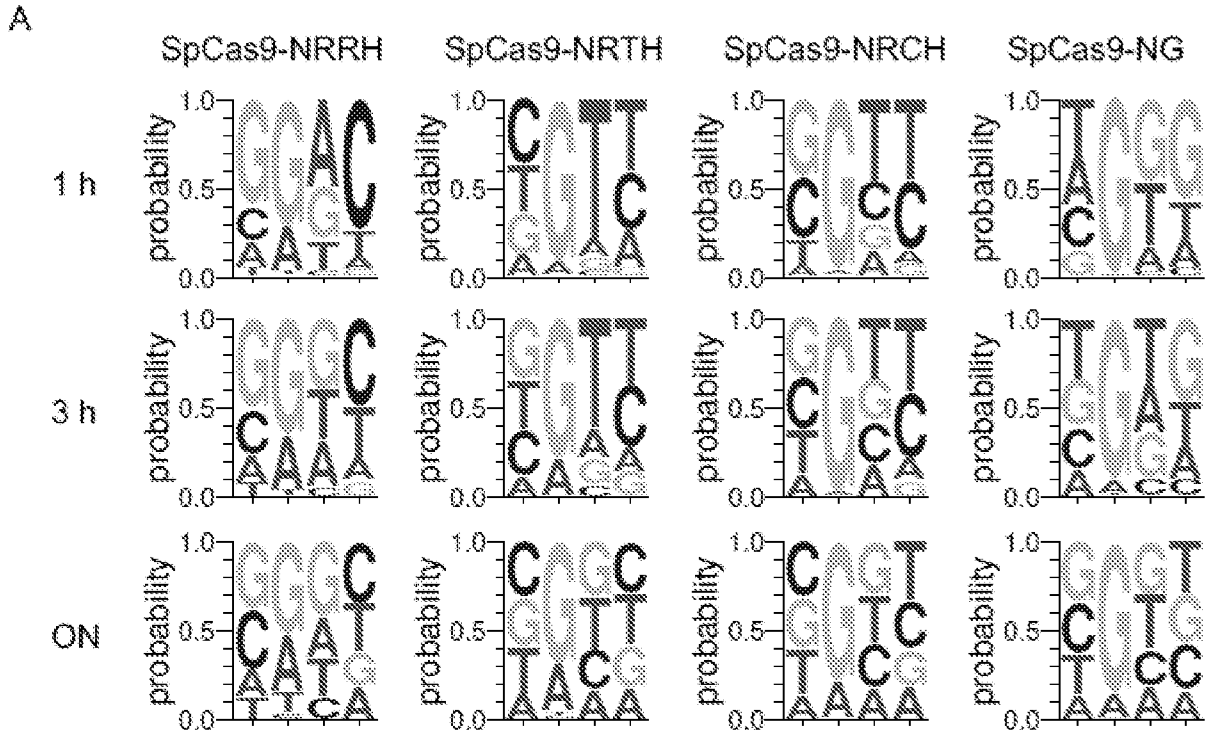


FIG. 45A

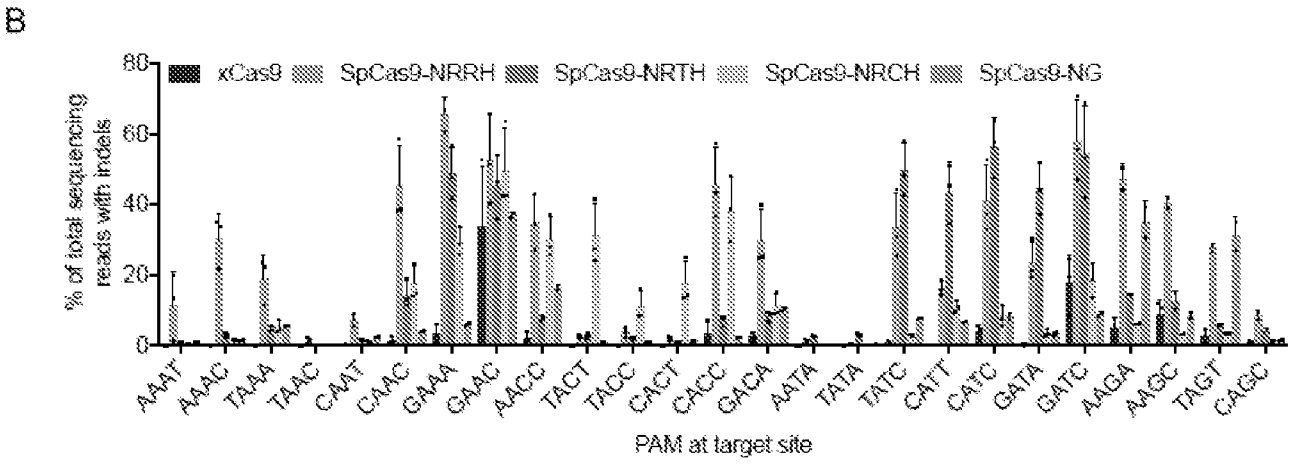


FIG. 45B

C

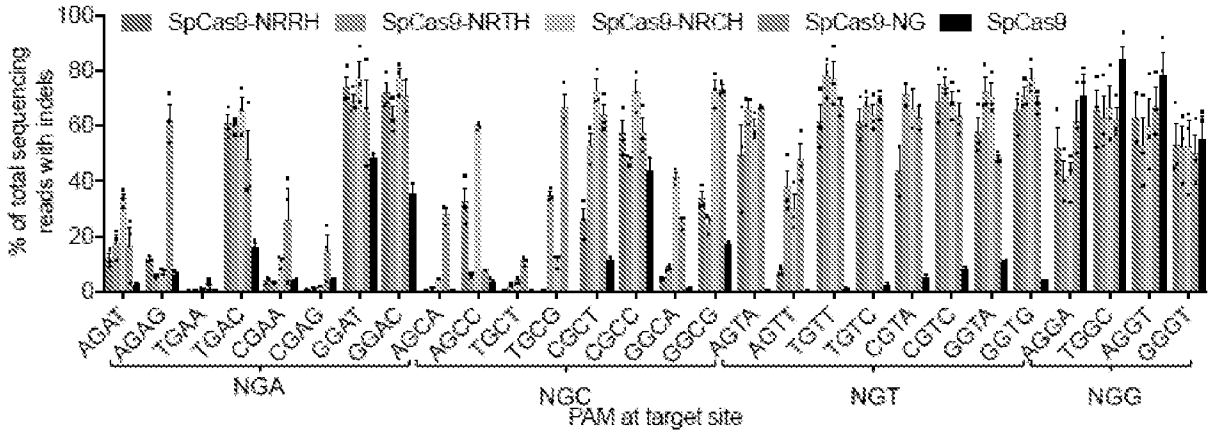

SpCas9-NRRH
SpCas9-NRTH
SpCas9-NRCH
SpCas9-NG
SpCas9
% of total sequencing reads with indels
100
80
60
40
20
0
AGAT
AGAG
TGAA
TGAC
CGAA
CGAG
GGAT
GGAC
AGCA
AGCC
TGCT
TGCG
CGCT
CGCC
GGCA
GGCG
AGTA
AGTT
TGTT
TGTC
CGTA
CGTC
GGTA
GGTG
AGGA
TGGC
AGGT
GGGT
NGA
NGC
NGT
NGG
PAM at target site

FIG. 45C

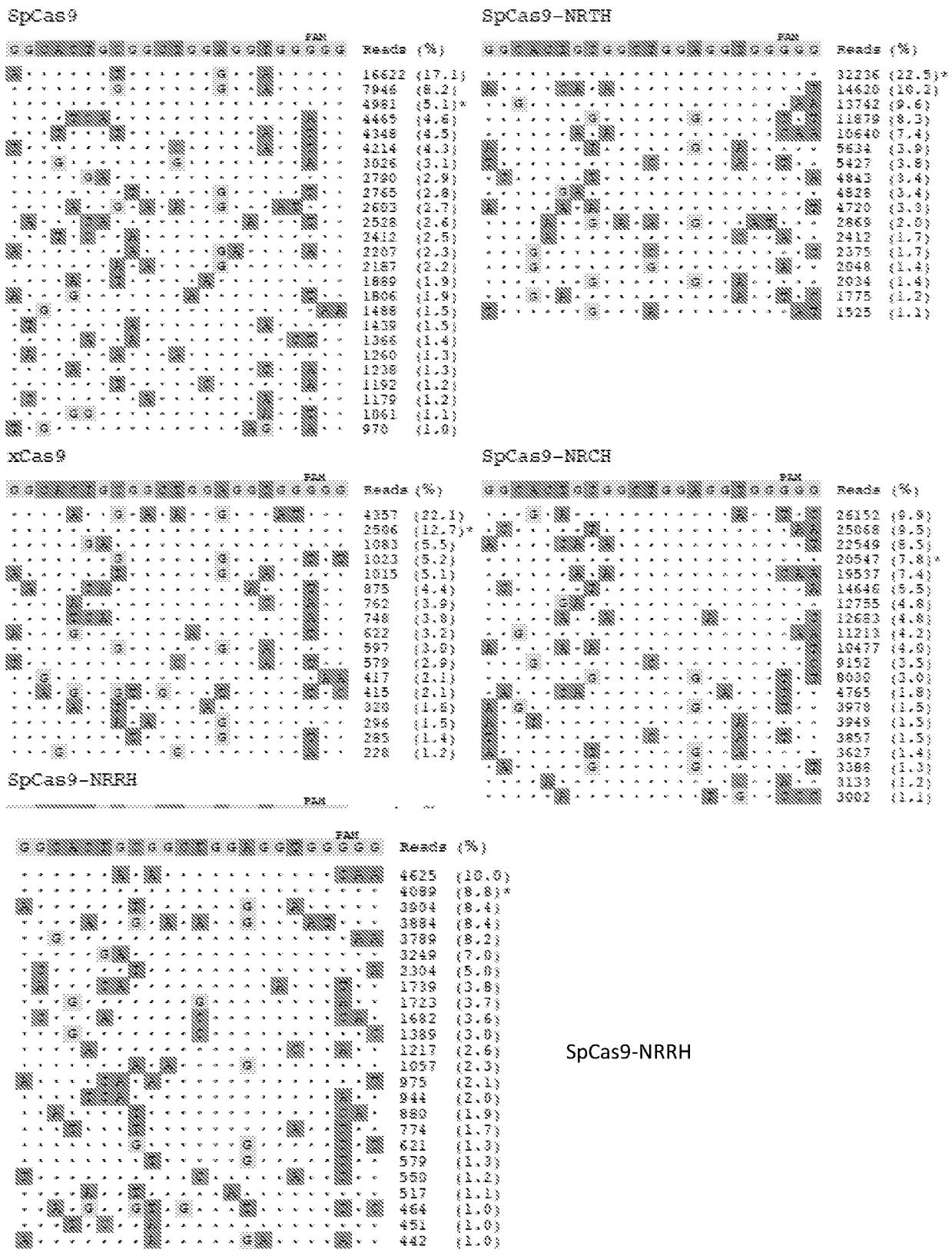

SpCas9
PAM
Reads (%)
16622 (17.1)
7946 (8.2)
4981 (5.1)*
4465 (4.6)
4348 (4.5)
4214 (4.3)
3026 (3.1)
2790 (2.9)
2765 (2.8)
2603 (2.7)
2528 (2.6)
2412 (2.5)
2207 (2.3)
2187 (2.2)
1889 (1.9)
1806 (1.9)
1488 (1.5)
1439 (1.5)
1368 (1.4)
1260 (1.3)
1238 (1.3)
1192 (1.2)
1179 (1.2)
1061 (1.1)
970 (1.0)
SpCas9-NRTH
PAM
Reads (%)
32236 (22.5)*
14620 (10.2)
13742 (9.6)
11879 (8.3)
10640 (7.4)
5634 (3.9)
5427 (3.8)
4843 (3.4)
4828 (3.4)
4720 (3.3)
2869 (2.0)
2412 (1.7)
2375 (1.7)
2048 (1.4)
2034 (1.4)
1775 (1.2)
1525 (1.1)
xCas9
PAM
Reads (%)
4357 (22.1)
2506 (12.7)*
1083 (5.5)
1023 (5.2)
1015 (5.1)
875 (4.4)
762 (3.9)
748 (3.8)
622 (3.2)
597 (3.0)
579 (2.9)
417 (2.1)
415 (2.1)
328 (1.6)
296 (1.5)
283 (1.4)
228 (1.2)
SpCas9-NRCH
PAM
Reads (%)
26152 (9.9)
25068 (9.5)
22549 (8.5)
20547 (7.8)*
19537 (7.4)
14646 (5.5)
12755 (4.8)
12683 (4.8)
11213 (4.2)
10477 (4.0)
9152 (3.5)
8030 (3.0)
4765 (1.8)
3978 (1.5)
3949 (1.5)
3857 (1.5)
3627 (1.4)
3386 (1.3)
3133 (1.2)
3002 (1.1)
SpCas9-NRRH
PAM
Reads (%)
4625 (10.6)
4089 (8.8)*
3904 (8.4)
3884 (8.4)
3789 (8.2)
3249 (7.0)
3304 (5.0)
1739 (3.8)
1723 (3.7)
1682 (3.6)
1389 (3.0)
1217 (2.6)
1057 (2.3)
975 (2.1)
944 (2.0)
880 (1.9)
774 (1.7)
621 (1.3)
579 (1.3)
550 (1.2)
517 (1.1)
464 (1.0)
451 (1.0)
442 (1.0)
SpCas9-NRRH

FIG. 45D

A

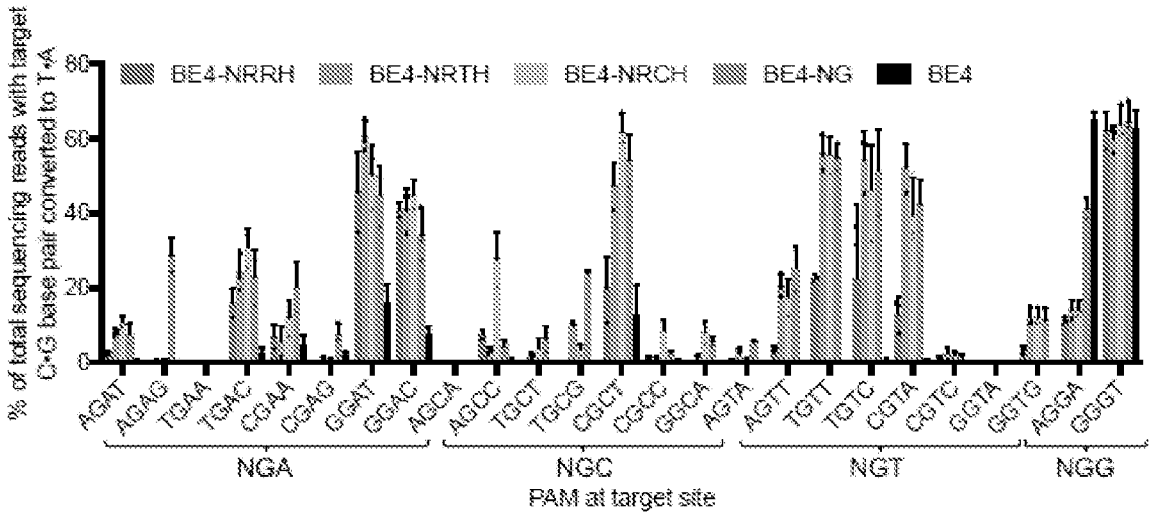


FIG. 46A

B

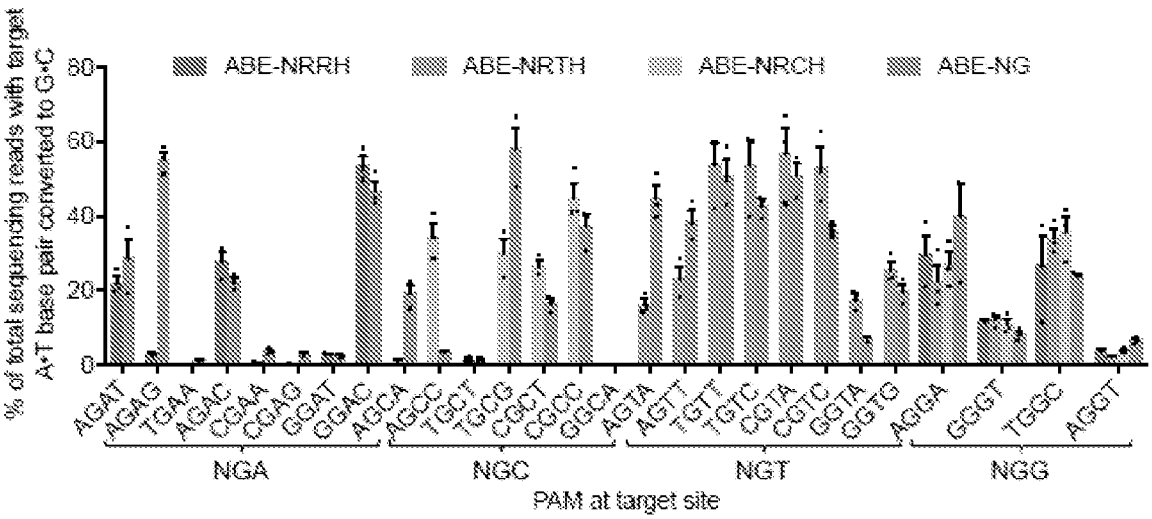


FIG. 46B

C

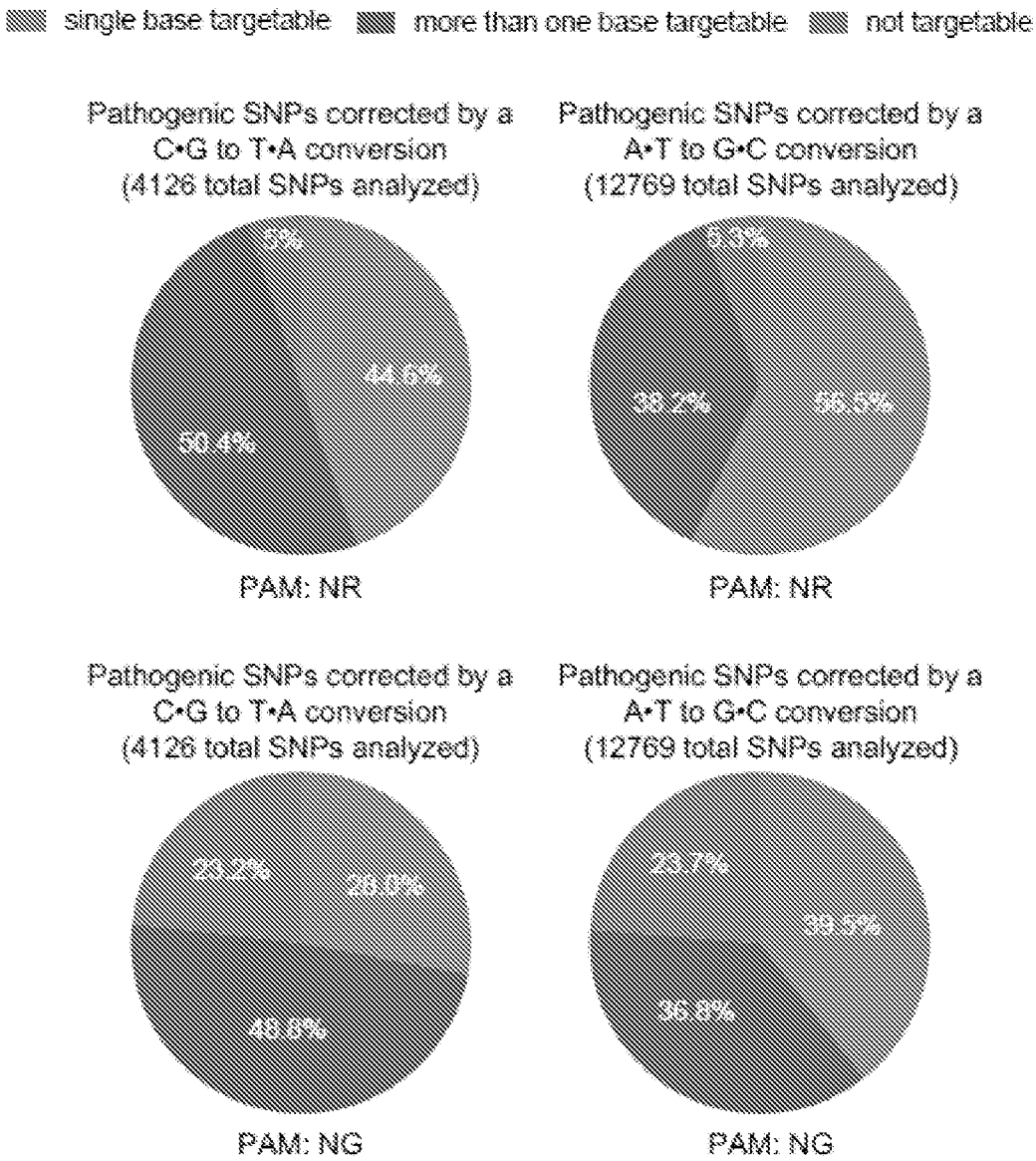

single base targetable
more than one base targetable
not targetable
Pathogenic SNPs corrected by a C•G to T•A conversion (4126 total SNPs analyzed)
5%
44.6%
50.4%
PAM: NR
Pathogenic SNPs corrected by a A•T to G•C conversion (12769 total SNPs analyzed)
5.3%
39.2%
55.6%
PAM: NR
Pathogenic SNPs corrected by a C•G to T•A conversion (4126 total SNPs analyzed)
23.2%
30.8%
46.0%
PAM: NG
Pathogenic SNPs corrected by a A•T to G•C conversion (12769 total SNPs analyzed)
23.7%
39.5%
36.8%
PAM: NG

FIG. 46C

A

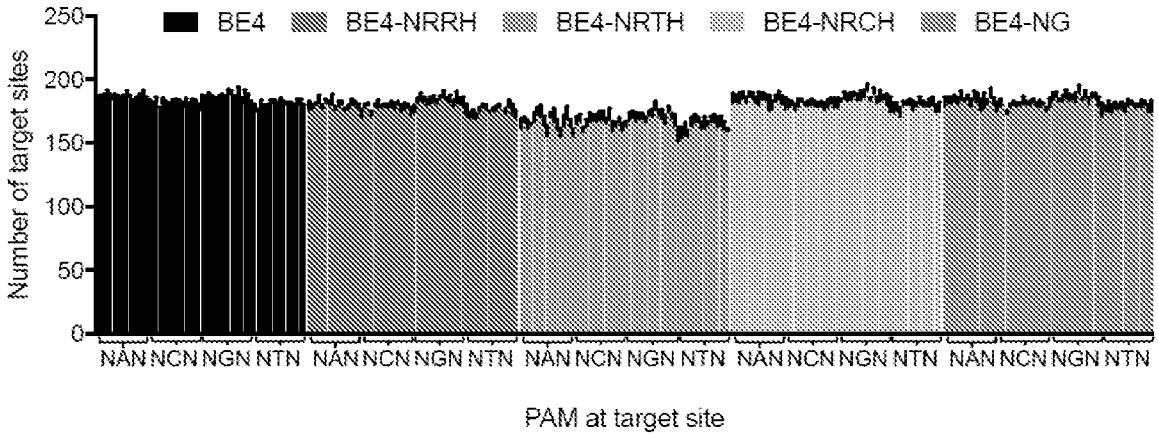


FIG. 47A

B

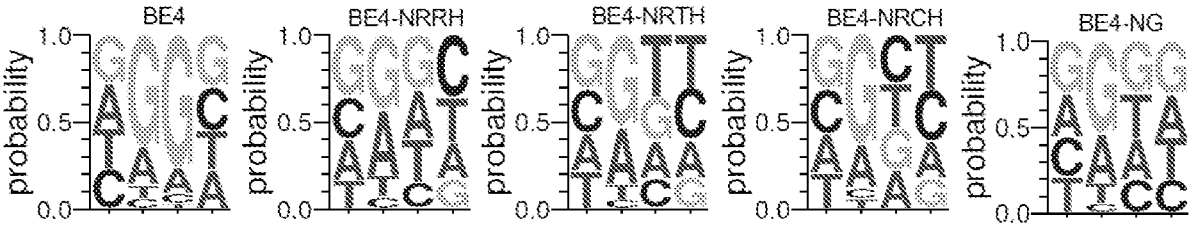


FIG. 47B

C

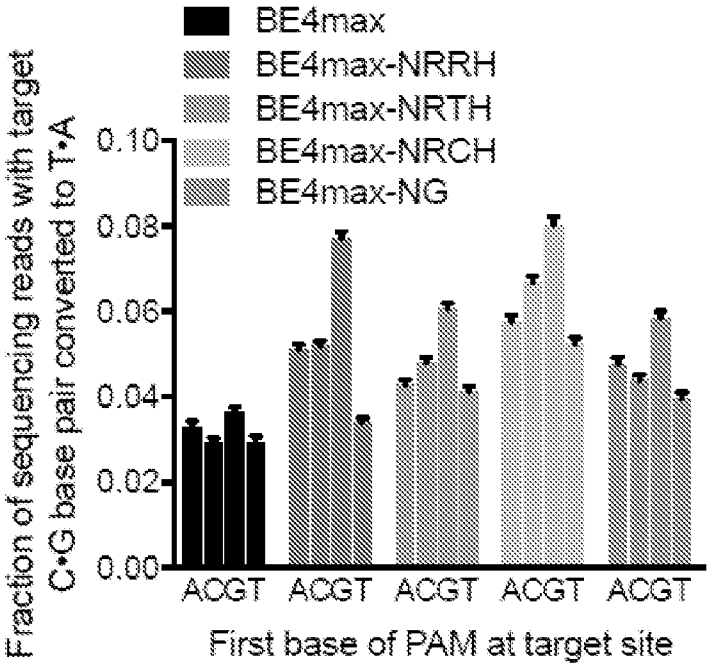


FIG. 47C

D

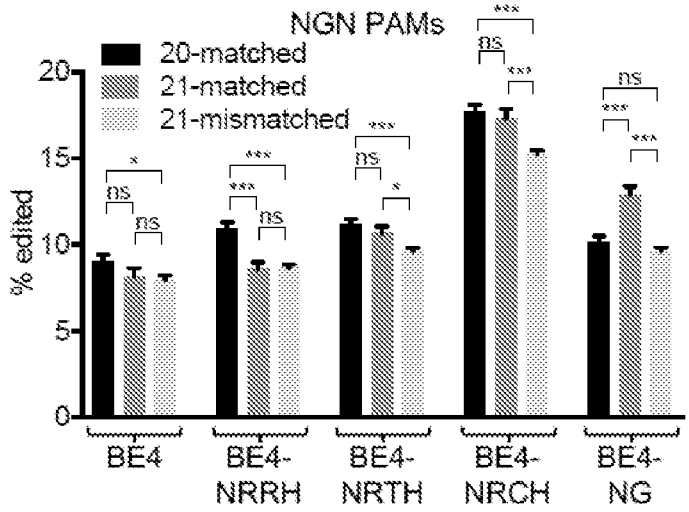


FIG. 47D

E

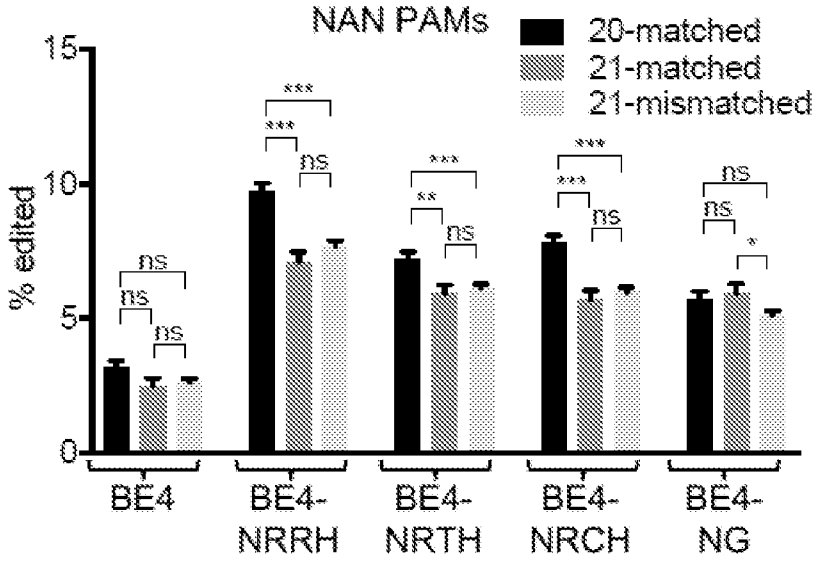


FIG. 47E

F

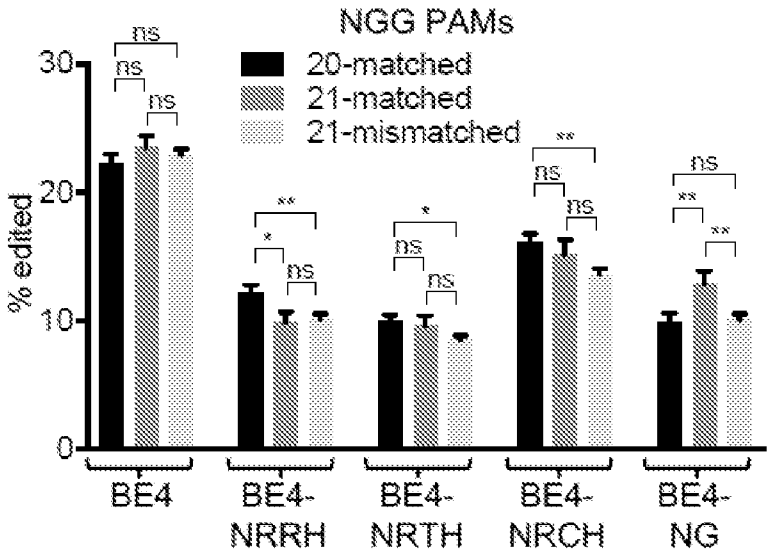


FIG. 47F

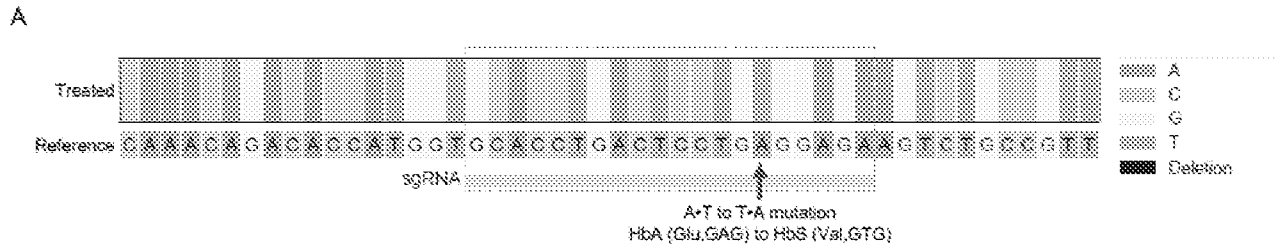


FIG. 48A

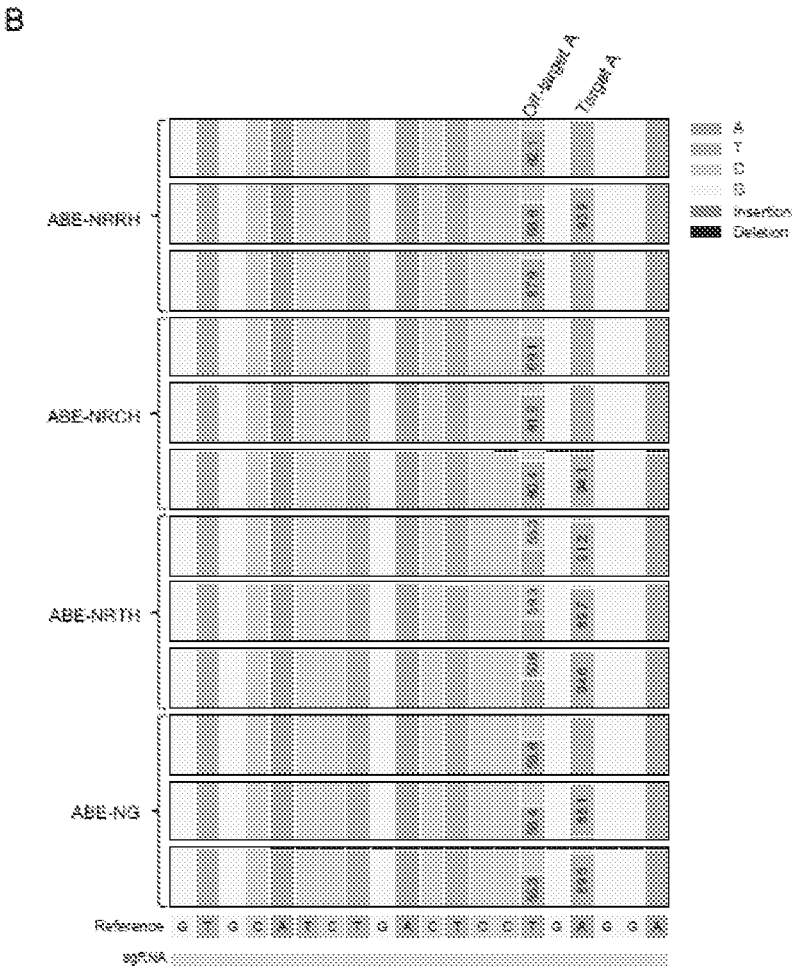


FIG. 48B

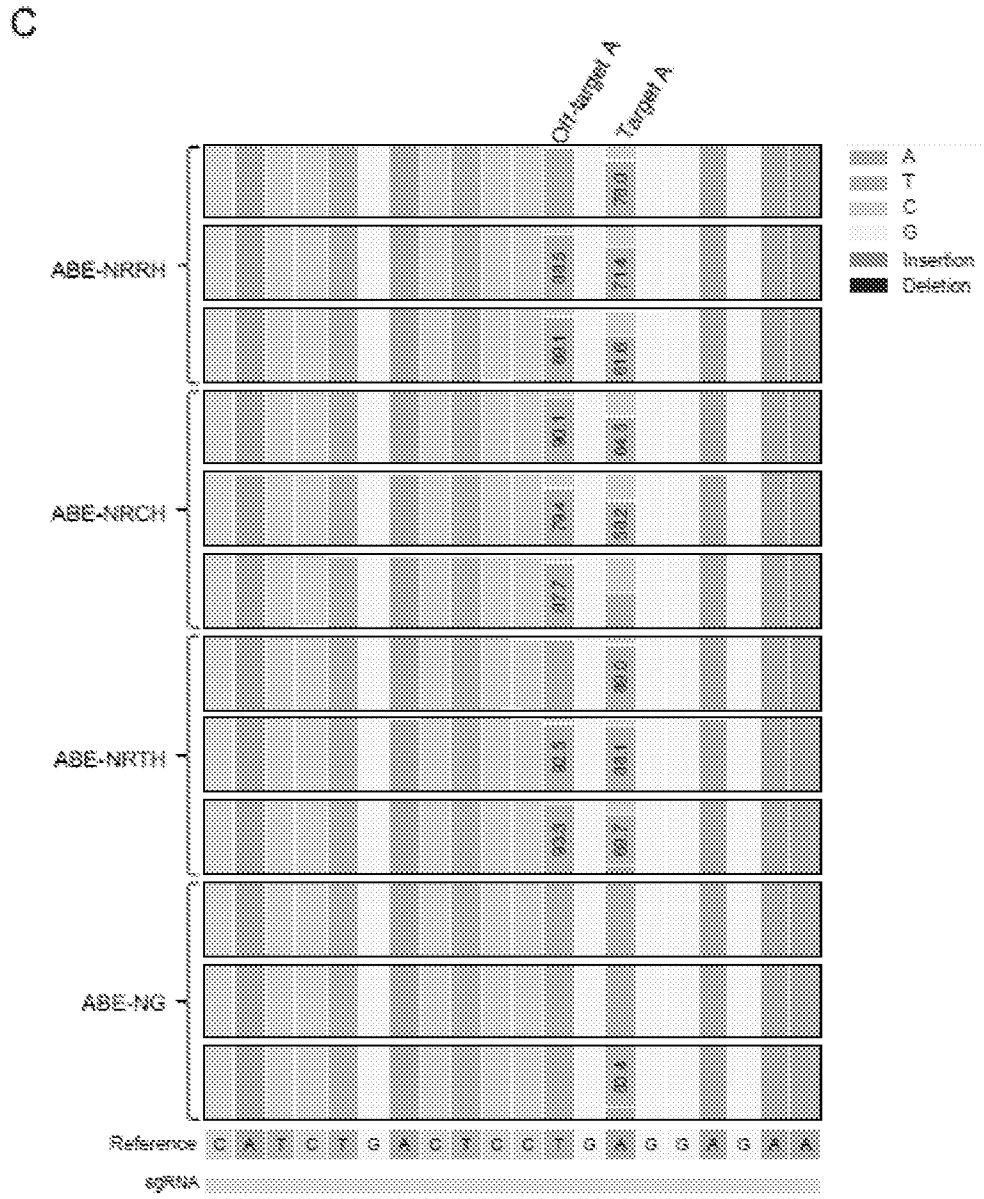

C
Off-target A
Target A
A
T
C
G
Insertion
Deletion
ABE-NRRH
ABE-NRCH
ABE-NRTH
ABE-NG
Reference
sgRNA

FIG. 48C

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 19/47996

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12N 9/00, C12N 9/22, C12N 9/24, C12N 15/62 (2019)

CPC - C12N 15/102, C12N 9/22, C07K 2319/00, C12N 2319/095

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
See Search History document

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- A	US 2017/0121693 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 4 May 2017 (04.05.2017) para [0015], [0153], SEQ ID NO: 2, 10	1-2, 4-5 --- 3, 6
A	US 2016/0340662 A1 (THE BROAD INSTITUTE, INC. et al.) 24 November 2016 (24.11.2016) para [0185]; SEQ ID NO: 44	3, 6
A	US 2018/0073012 A1 (PRESIDENT AND FELLOWS OF HARVARD COLLEGE) 15 March 2018 (15.03.2018) para [0280], [0287]	152, (155-158)/152
A	CN107177625-A (INSTITUTE OF PLANT PROTECTION, CHINESE ACADEMY OF AGRICULTURAL SCIENCES) 19 September 2017 (19.09.2017) pg 44-45 SEQ ID NO 29	152, (155-158)/152
A	NISHIMASU et al., Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA, Cell. 27 February 2014, Vol 156, No 5, pp 935 -949. Especially pg 937, col 2, para 1; pg 937, col 1, para 4; pg 939, col 1, para 3; Figure 3	3, 6

☐ 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
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 January 2020	28 JAN 2020
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Lee Young Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 19/47996

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

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

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

2. ☐ Claims Nos.:
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:

3. ☒ Claims Nos.: 7-16, 23-32, 39-151, 159
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

----------Please see Supplemental Sheet----------

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.:
1-6, 152, (155-158)(in part) limited to X10T, SEQ ID NOs: 122, 27, and 62

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.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/US 19/47996

Continuation of Box No. III Observations where unity of invention is lacking:

Groups I+: Claims 1-6, 17-22, 33-38, 152-158, drawn to a Cas9 protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of a Cas9 of SEQ ID NO: 2, wherein the amino acid sequence of the Cas9 protein comprises at least one to at least nine mutations in selected amino acid residue(s); a Streptococcus pyogenes Cas9 (SpCas9) having a non-canonical PAM activity, and a fusion protein comprising said SpCas9. The Cas9 and fusion proteins will be searched to the extent that the Cas9 variant encompasses amino acid mutation on position 10 (X10T) (relative to SEQ ID NO: 2); the SpCas9 having a non-canonical PAM activity encompasses SEQ ID NO: 122; the fusion protein encompasses cytidine deaminase of SEQ ID NO: 27 or adenosine deaminase of SEQ ID NO: 62. It is believed that claims 1-6, 152, (155-158)(in part) encompass this first named invention, and thus these claims will be searched without fee to the extent that they encompass mutations of X10T (relative to SEQ ID NO: 2), SEQ ID NOs: 122, 27, and 62. Additional mutation(s) and amino acid sequence(s) of Cas9, cytidine deaminase and adenosine deaminase will be searched upon the payment of additional fees. Applicants must specify the claims that encompass any additionally elected mutation(s) and amino acid sequence(s) of Cas9, cytidine deaminase and adenosine deaminase. Applicants must further indicate, if applicable, the claims which encompass the first named invention, if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched. An exemplary election would be a mutation at amino acid position 472 (X472I); a SpCas9 having a non-canonical PAM activity encompassing SEQ ID NO: 123; the fusion protein encompassing cytidine deaminase of SEQ ID NO: 28 and adenosine deaminase of SEQ ID NO: 63 (Claims 17-22, 153, (155-158)(in part)).

The inventions listed as Groups I+ do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

No technical features are shared between the amino acid sequences of Groups I+ and, accordingly, these groups lack unity a priori.

Additionally, even if Groups I+ were considered to share the technical features of including: a Cas9 variant, these shared technical features are previously disclosed by US 2016/0340662 A1 to The Broad Institute, Inc. et al. (hereinafter "Broad Institute").

Broad Institute teaches (instant claim 1) a Cas9 protein comprising an amino acid sequence that is at least 80% identical to the amino acid sequence of a Cas9 as provided by any one of SEQ ID NO: 2 (para [0185]; SEQ ID NO: 44 encoding SpCas9-EGFP-NLS, amino acids 1-1368 exhibit 98.5% identity to claimed SEQ ID NO: 2), wherein the amino acid sequence of the Cas9 protein comprises at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, or at least nine mutations of the amino acid sequence provided in SEQ ID NO: 2 (para [0098] - "As a further example, two or more catalytic domains of Cas9 (RuvC I, RuvC II, and RuvC III) may be mutated to produce a mutated Cas9 substantially lacking all DNA cleavage activity. In some embodiments, a D10A mutation is combined with one or more of H840A, N854A, or N863A mutations to produce a Cas9 enzyme substantially lacking all DNA cleavage activity.").

Some inventions of Group I+ share the technical features of a SpCas9, wherein the SpCas9 has a non-canonical PAM specificity, which is disclosed by US 2018/0073012 A1 to President and Fellows of Harvard College (hereinafter "Harvard '012") (para [0287] - "In some embodiments, the Cas9 domain is a Cas9 domain from Streptococcus pyogenes (SpCas9)."; para [0280] - "Accordingly, in some embodiments, any of the fusion proteins provided herein may contain a Cas9 domain that is capable of binding a nucleotide sequence that does not contain a canonical (e.g., NGG) PAM sequence. Cas9 domains that bind to non-canonical PAM sequences have been described in the art and would be apparent to the skilled artisan. For example, Cas9 domains that bind non-canonical PAM sequences have been described... the entire contents of each are hereby incorporated by reference.").

Some inventions of Group I+ share the technical features of a fusion protein comprising a SpCas9 and an adenosine deaminase, which is disclosed by Harvard '012 (para [0006] - "Other aspects of the disclosure provide fusion proteins comprising a Cas9 domain and an adenosine deaminase domain, for example, an engineered deaminase domain capable of deaminating an adenosine in DNA."; para [0287] - "In some embodiments, the Cas9 domain is a Cas9 domain from Streptococcus pyogenes (SpCas9).").

Some inventions of Group I+ share the technical features of a fusion protein comprising a SpCas9 and a cytidine deaminase, which is disclosed by WO 2017/070632 A2 to President and Fellows of Harvard College (hereinafter "Harvard '632") (para [0013] - "Some aspects of this disclosure provide fusion proteins comprising a Cas9 protein as provided herein that is fused to a second protein (e.g., an enzymatic domain such as a cytidine deaminase domain), thus forming a fusion protein."; para [00183] - "In some embodiments, the Cas9 domain is a Cas9 domain from Streptococcus pyogenes (SpCas9).").

As said technical features were known in the art at the time of the invention, these cannot be considered special technical features that would otherwise unify the groups.

Groups I+ therefore lack unity under PCT Rule 13 because they do not share a same or corresponding special technical feature.

Item 4 (continued):
Claims 7-16, 23-32, 39-151, 159 are improper multiple dependent claims because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).