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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
01 February 2024 (01.02.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/026465 A1

(51) International Patent Classification:
C07K 19/00 (2006.01) *C07K 14/195* (2006.01)
C12N 1/20 (2006.01) *C07K 14/435* (2006.01)
C12N 15/62 (2006.01) *G01N 33/50* (2006.01)
C12N 9/14 (2006.01) *G01N 33/52* (2006.01)
C12N 9/22 (2006.01)

(21) International Application Number:
PCT/US2023/071227

(22) International Filing Date:
28 July 2023 (28.07.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/393,403 29 July 2022 (29.07.2022) US

(71) Applicants: **THE BROAD INSTITUTE, INC.** [US/US]; 415 Main Street, Cambridge, Massachusetts 02142 (US). **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US).

(72) Inventors: **ZHANG, Feng**; c/o 415 Main Street, Cambridge, Massachusetts 02142 (US). **GAO, Alex**; c/o 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US). **WILKINSON, Max**; c/o 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US). **STRECKER, Jonathan**; c/o 415 Main Street, Cambridge, Massachusetts 02142 (US).

(74) Agent: **MILLER, Carin R.** et al.; Johnson, Marcou, Issacs & Nix, LLC, Customer No. 95280, PO Box 691, Hoschton, Georgia 30548 (US).

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

(54) **Title:** PROGRAMMABLE PATTERN RECOGNITION COMPOSITIONS

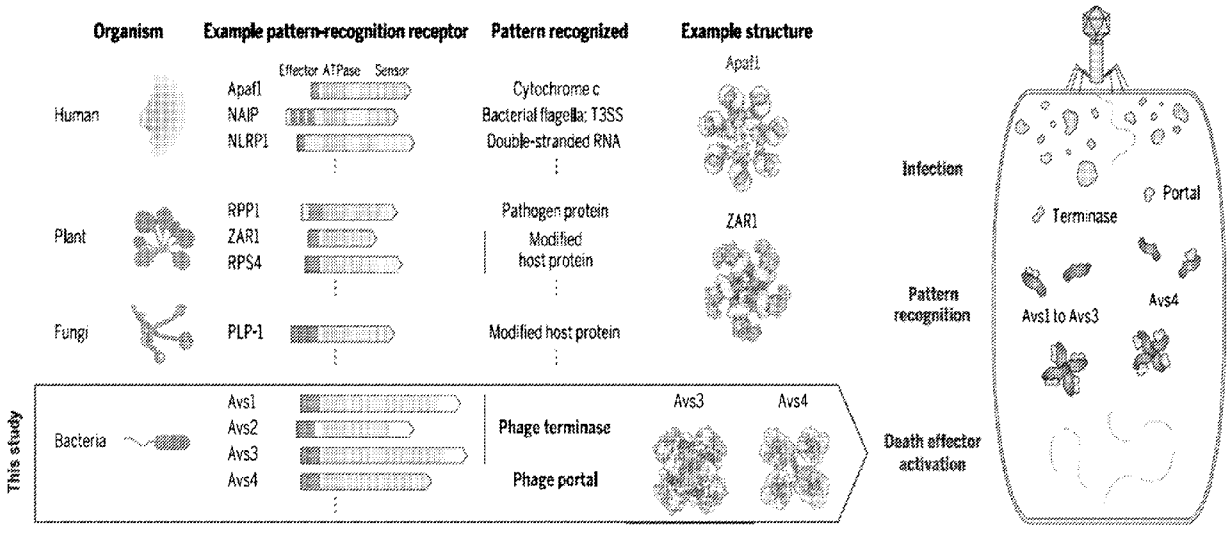


FIG. 26

(57) **Abstract:** Described in several example embodiments herein are engineered programmable pattern recognition compositions and uses thereof. In certain example embodiments, the engineered protein contains an NTPase of a Signal Transduction ATPases with Numerous-associated Domains (STAND) superfamily (STAND NTPase), comprising a pathogen-associated molecular pattern (PAMP) recognition activity, wherein the STAND NTPase and the PAMP recognition activity are derived from the same or different prokaryotes.

[Continued on next page]

WO 2024/026465 A1

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

Published:

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

PROGRAMMABLE PATTERN RECOGNITION COMPOSITIONS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of and priority to U.S. Provisional Patent Application No. 63/393,403, filed on July 29, 2022, the contents of which is incorporated by reference herein in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under Grant No. HL141201 and HG009761-05 awarded by the National Institutes of Health. The government has certain rights in the invention.

SEQUENCE LISTING

[0003] This application contains a sequence listing filed in electronic form as an xml file entitled "BROD-5585WP_ST26.xml", created on July 28, 2023, and having a size of 104,799 bytes. The content of the sequence listing is incorporated herein in its entirety.

TECHNICAL FIELD

[0004] The subject matter disclosed herein is generally directed to prokaryotic innate immunity via pattern recognition of conserved viral proteins.

BACKGROUND

[0005] All organisms have evolved specialized immune proteins, including pattern recognition receptors consisting of nucleotide-binding oligomerization domain-like receptors (NLRs) of the STAND superfamily ubiquitous in eukaryotes. NLRs recognize conserved pathogen-associated molecular patterns, leading to activation of an effector domain and an inflammatory or apoptotic response. The roles of NLRs in eukaryotic immunity are well established, but it is unknown whether prokaryotes use similar defense mechanisms.

[0006] Citation or identification of any document in this application is not an admission that such a document is available as prior art to the present invention.

SUMMARY

[0007] Described in certain example embodiments herein are engineered proteins comprising an effector domain, an effector activation domain, and a recognition domain, wherein binding of a target polypeptide to the recognition domain leads to activation of the effector domain via the effector activation domain, and wherein at least one of the effector, effector activation, and recognition domains is derived from a STAND NTPase protein.

[0008] In certain example embodiments, the STAND NTPase protein is an antiviral STAND (Avs). In certain example embodiments, the Avs is an Avs1, Avs2, Avs3, or Avs4.

[0009] In certain example embodiments, the effector domain is an endonuclease, a protease, a nucleosidase, hydrolase, or caspase-like domain.

[0010] In certain example embodiments, the effector activation domain is an NTPase.

[0011] In certain example embodiments, the recognition domain is engineered to recognize a target polypeptide other than a target polypeptide of a wild-type STAND NTPase protein. In certain example embodiments, the recognition domain comprises tetratricopeptide repeat (TPR) domains.

[0012] In certain example embodiments, wherein a microbe comprises the target polypeptide, optionally wherein the microbe is part of a microbiome.

[0013] In certain example embodiments, the target polypeptide is a phage polypeptide.

[0014] Described in certain example embodiments herein are oligomers comprising at least two of the engineered proteins of the present invention. In some embodimets, the oligomer is a tetramer, tirmer, or dimer.

[0015] Described in certain example embodiments herein are detection compositions comprising (a) an engineered protein of any one of the preceding paragraphs and as described in greater detail elsewhere herein; (b) a detection construct, wherein binding of a target polypeptide to the recognition domain activates the effector domain and mediates effector domain modification of the detection construct resulting in generation of a detectable signal.

[0016] Described in certain example embodiments herein are polynucleotide(s) encoding the engineered protein of any one of the preceding paragraphs and as described in greater detail elsewhere herein.

[0017] Described in certain example embodiments herein are polynucleotide(s) encoding component (a), component (b), or both of the detection composition.

[0018] Described in certain example embodiments herein are vectors and vector systems comprising a polynucleotide encoding an engineered protein described herein and/or a detection composition described herein.

[0019] Described in certain example embodiments herein are cells or cell populations comprising an engineered protein of the present invention described herein, a detection composition of the present invention described herein, a polynucleotide encoding an engineered protein of the present invention and/or a detection composition of the present invention, a vector or vector system of the present invention, or any combination thereof.

[0020] Described in certain example embodiments herein are formulation comprising an engineered protein of the present invention described herein, a detection composition of the present invention described herein, a polynucleotide encoding an engineered protein of the present invention and/or a detection composition of the present invention, a vector or vector system of the present invention, a cell or cell population of the present invention, or any combination thereof, and optionally a pharmaceutically acceptable carrier.

[0021] Described in certain example embodiments herein are methods of modifying a target molecule and/or cell comprising delivering an engineered protein of the present invention described herein, a detection composition of the present invention described herein, a polynucleotide encoding an engineered protein of the present invention and/or a detection composition of the present invention, a vector or vector system of the present invention, or any combination thereof to the target molecule and/or cell, wherein the target molecule and/or cell is or comprises a target polypeptide; and activating an effector domain of the engineered protein by allowing binding of the target polypeptide to the recognition domain thereby activating the effector domain via the effector activation domain, wherein effector domain activity modifies the target molecule and/or cell. In certain example embodiments, delivering comprises *in vitro*, *ex vivo*, or *in vivo* delivery.

[0022] Described in certain example embodiments herein are methods of detecting a target molecule and/or cell, the method comprising combining a detection composition of the present invention or a formulation thereof and a sample or component thereof; and activating an effector domain of the engineered protein via binding of a target polypeptide in the sample to the recognition domain thereby mediating effector domain modification of the detection construct and generation of a detectable signal. In certain example embodiments, the method is performed in whole or in part *in vitro*, *ex vivo*, or *in vivo*.

[0023] Described in certain example embodiments herein are methods of modifying a microbiome structure comprising introducing an engineered protein of any one of the engineered proteins of the present invention capable of recognizing a target polypeptide of one or more microbes in a microbiome into a microbiome, wherein activation of the effector domain via binding of a target polypeptide of one or more microbes in the microbiome to the recognition domain results in modification of the one or more microbes thereby modifying the microbiome structure.

[0024] Described in certain example embodiments herein are methods of engineered phage-resistant bacteria comprising expressing an engineered protein of the present invention capable of recognizing a phage polypeptide in a bacterium or bacteria population.

[0025] Described in certain example embodiments herein are methods of cargo delivery comprising delivering, to a cell, (a) an engineered protein of the present invention; (b) a cargo, (c) a detection composition or (d) any combination thereof, wherein the engineered protein comprises the cargo or wherein the cargo comprises the target polypeptide, wherein the cell optionally comprises the target polypeptide, and wherein activation of the effector domain by binding of the target polypeptide to the recognition domain results in delivery of the cargo.

[0026] These and other aspects, objects, features, and advantages of the example embodiments will become apparent to those having ordinary skill in the art upon consideration of the following detailed description of example embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

[0027] An understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention may be utilized, and the accompanying drawings of which:

[0028] **FIG. 1A-1B** – Example of avs genes (**FIG. 1A**) present in defense islands and (**FIG. 1B**) clustered with other avs genes. Other defense genes are highlighted in gray.

[0029] **FIG. 2** – Heterologous reconstitution of Avs anti-phage activity in E. coli. Plaque assay spots correspond to 10-fold dilutions of phages T7, PhiV-1, P1, Lambda, T4, T5, and ZL19 on E. coli containing Avs-expressing plasmids.

[0030] **FIG. 3A-3F** – Prokaryotic STAND NTPases recognize phage terminase and portal proteins. (**FIG. 3A**) Maximum likelihood tree of the ATPase domain of selected NLR-like

STAND NTPases in four model organisms across kingdoms of life. (**FIG. 3B**) Domain architectures of representative NLR-like genes in **FIG. 3A**. LRR, leucine-rich repeat; TPR, tetratricopeptide repeat; WD40, WD40 repeat; ankyrin, ankyrin repeat; BIR, baculoviral inhibitor of apoptosis repeat; PYD, pyrin domain; FIIND, function to find domain; CARD, caspase activation and recruitment domain; RX-CC, potato virus X resistance protein coiled-coil domain; PLP, patatin-like phospholipase; TIR, toll/interleukin-1 receptor homology domain. (**FIG. 3C**) Schematic of genetic screening approach to identify phage-encoded activators of Avs proteins that induce cell death. (**FIG. 3D**) Genetic screen results for phage-encoded activators. (**FIG. 3E**) Quantification of the phage DNA band intensity in a Southern blot of DNA isolated from phage-infected E. coli. (**FIG. 3F**) Photographs of E. coli co-transformation assays with Avs genes and phage activators identified in **FIG. 3D**.

[0031] **FIG. 4A-4B** – Four distinct clades of Avs proteins. (**FIG. 4A**) Phylogenetic tree of the ATPase domain of selected Avs proteins and other related ATPases identified by PSIBLAST. (**FIG. 4B**) UPGMA dendrogram including the ATPase domains in **FIG. 4A** and profiles of additional ATPases in pfam.

[0032] **FIG. 5A-5B** – Related to **FIG. 3A-3F**. (**FIG. 5A**) Schematic of PhiV-1 fragment screen (**FIG. 1C**). (**FIG. 5B**) Read coverage of PhiV-1 fragment screen, without normalizing to the empty vector control.

[0033] **FIG. 6A-6E** – Related to **FIG. 3A-3F**. (**FIG. 6A**) Schematic for PhiV-1 mutant construction via plasmid homology donors in E. coli. A trans complementation plasmid encoding gp8 or gp19 was maintained in the cells to support phage growth. (**FIG. 6B**) Plaque assay validation of PhiV-1 knockout phages across different complementation plasmids. Spots correspond to 10-fold phage dilutions from right to left. (**FIG. 6C**) Southern blot analysis of phage-infected E. coli cell lysates using a PhiV-1 specific probe. WT, Δgp8, and Δgp19 were used at an MOI of 1, and Δgp19CTD was used at an MOI of 0.25. END-seq analysis (74) of DNA double-strand breaks (DSBs) in (**FIG. 6D**) NotI-digested PhiV-1 DNA as a positive control and (**FIG. 6E**) E. coli cultures infected with wild type or mutant PhiV-1. Input DNA was normalized across samples.

[0034] **FIG. 7A-7D** – Avs proteins are pattern-recognition receptors for the terminase and portal of diverse tailed phages. (**FIG. 7A**) Schematic of plasmid depletion assay. (**FIG. 7B**) Heatmaps of plasmid depletion for the terminase and portal proteins of representative phages spanning nine major tailed phage families. The native Avs promoter was retained for all

homologs except for those outside of the Enterobacteriaceae family (EpAvs1 and CcAvs4).Terminases and portals were induced with 0.002% arabinose. Horizontal black bars indicate groups of terminase proteins with at least 20% pairwise sequence identity. Asterisks indicate prophages. S. flava, Sphingopyxis flava R11H; D. archaeon, Desulfurococcales archaeon ex4484_217_2; E. coli-1, Escherichia coli NCTC9020; E. coli-2, Escherichia coli M885. (**FIG. 7C**) Pairwise amino acid sequence identity between the core folds of the terminases and portals in (**FIG. 7B**), excluding non-conserved regions. (**FIG. 7D**) Activity of four Avs proteins against the human herpesvirus 8 (HHV-8) terminase and portal.

[0035] **FIG. 8** – Related to **FIG. 3A-3F**. Photographs of E. coli co-transformation assays with Avs1-2 and activators from phage PhiV-1. The left spot on each image corresponds to a 10-fold dilution of the right spot.

[0036] **FIG. 9A-9B** – Robustness of the terminase and portal plasmid depletion assay in **FIG. 7A-7D**. (**FIG. 9A**) Specificity of Avs target recognition with avs genes expressed under the control of a lac promoter and weak induction of terminases and portals (0.002% arabinose). (**FIG. 9B**) Specificity of Avs target recognition with native avs promoters and strong induction of terminases and portals (0.2% arabinose). Terminases and portals were expressed under the control of a pBAD promoter. Gray boxes indicate pairwise combinations not assessed due to the toxicity of terminase overexpression.

[0037] **FIG. 10A-10C** – Avs1, Avs2, and Avs3 contain a structurally conserved C-terminal domain essential for defense activity. (**FIG. 10A**) Structures predicted by AlphaFold2 of the C-terminal domains (CTDs) of the seven Avs1-3 homologs investigated in this study. The N- and C-termini are colored blue and red, respectively and represented in greyscale. (**FIG. 10B**) Heatmap of Dali Z-scores of pairwise comparisons between the Avs1-3 CTDs in (A) (smallest Z-score = 7). Z-scores above 2 indicate significant structural similarity (Holm, Methods Mol Biol 2112, 29-42 (2020)). (**FIG. 10C**) Effect of CTD deletion on EcAvs2 defense activity against T7 and PhiV-1. Spots correspond to 10-fold dilutions from right to left.

[0038] **FIG. 11** – Structures of portal proteins predicted by AlphaFold2. The core portal fold is shown in gray. The clip, crown, and other insertions are colored blue, red, and orange, respectively and as represented in greyscale. Asterisks indicate prophages.

[0039] **FIG. 12** – Structures of the N-terminal ATPase domains of large terminases predicted by AlphaFold2. The core ATPase fold is shown in gray.

[0040] **FIG. 13** – Structures of the C-terminal nuclease domains of large terminases predicted by AlphaFold2. The core nuclease fold is shown in gray.

[0041] **FIG. 14A-14H** – SeAvs3 and EcAvs4 are phage-activated DNA endonucleases. (**FIG. 14A**) Domain architecture of SeAvs3 and EcAvs4. (**FIG. 14B**) (SEQ ID NO: 1-6) Alignment of Avs D-QxK nuclease motifs with characterized Cap4 and Mrr representatives. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. (**FIG. 14C-14E**) Agarose gel analysis of SeAvs3 nuclease activity in vitro with a linear dsDNA substrate [(**FIG. 14C**) and (**FIG. 14D**)] and cofactor requirements (**FIG. 14E**). (**FIG. 14F-14H**) Agarose gel analysis of EcAvs4 nuclease activity in vitro with a linear dsDNA substrate (**FIG. 14F-14G**] and cofactor requirements (**FIG. 14H**).

[0042] **FIG. 15A-15B** – Requirements for Avs3 and Avs4 defense activity. Effects of (**FIG. 15A**) Avs3 small ORF deletion and (**FIG. 15B**) Avs3-4 nuclease and ATPase Walker A/B mutations on activity against T7 and PhiV-1.

[0043] **FIG. 16A-16C** – In vitro reconstitution of Avs activity. (**FIG. 16A**) Coomassie stained SDS-PAGE gel of purified Avs proteins and phage triggers. (**FIG. 16B, 16C**) Agarose gel analysis of SeAvs3 nucleic acid substrate specificity. Related to **FIG. 14A-14H**.

[0044] **FIG. 17A-17L** – Bacterial two-hybrid analysis of EcAvs4-portal interactions. (**FIG. 17A**) Schematic of a bacterial two-hybrid system for detecting protein-protein interactions. (**FIG. 17B**) Two-hybrid analysis of pairwise interactions of EcAvs4 and PhiV-1 proteins grown on S-gal indicator plates. (**FIG. 17C**) Interactions between EcAvs4 and the portal and terminase genes from eight phages. (**FIG. 17D**) Schematic of EcAvs4 protein domains. (**FIG. 17E**) Two-hybrid analysis of EcAvs4 mutations and truncations. (**FIG. 17F**) Two-hybrid analysis of PhiV-1 portal deletions. (**FIG. 17G**) Effect of T7 portal deletions on the activation of Avs4 as assessed by plasmid depletion. Arrows represent lac promoters. (**FIG. 17H**) Locations of mutations in the T7 portal (PDB: 6R21) generated by error-prone PCR that abolish activation of Avs4 (Cuervo et al., Nat. Commun. 10, 3746 (2019). (**FIG. 17I**) Schematic of tandem affinity purification of the SeAvs3-terminase complex.. (**FIG. 17J**) Size exclusion chromatography of SeAvs3, PhiV-1 terminase, and the SeAvs3-terminase complex. (**FIG. 17K**) Coomassie-stained SDS-PAGE protein gel of the SeAvs3-terminase complex. (**FIG. 17L**) Effect of terminase domain deletions on the activation of Avs1, Avs2, and Avs3.

The structure of the T4 terminase (Sun et al., Cell. 135, 1251-1262 (2008) is shown as an example.

[0045] **FIG. 18A-18C** – Identification of single amino acid substitutions in the T7 portal protein that abrogate Avs4 activation. (**FIG. 18A**) (SEQ ID NO: 7) Design of a translation-reinitiation reporter system used to facilitate screening of Avs4 mutants. (**FIG. 18B**) Validation of reporter performance via mNeonGreen fluorescence from E. coli colonies. Scale bar: 1 cm. (**FIG. 18C**) Activity and location of the 29 identified portal mutants that abrogate Avs4 activation.

[0046] **FIG. 19A-19G** – Related to **FIG. 17A-17L**. (**FIG. 19A**) Two-hybrid analysis of pairwise interactions between SeAvs3 components and PhiV-1 triggers grown on S-gal indicator plates and (**FIG. 19B**) pairwise interactions between SeAvs3 and the portal and terminase genes from eight phages. (**FIG. 19C**) Schematic for Avs co-purification strategy. (**FIG. 19D**) SDS-PAGE analysis of SeAvs3 and EcAvs4 affinity purification in the presence of gp8 portal or gp19 terminase. Highlighted bands were excised and analyzed by mass spectrometry. (**FIG. 19E**) Total and unique mapped peptides from mass spectrometry analysis of gp19 and gp8 gel bands. (**FIG. 19F**) Size exclusion chromatography of protein standards (a: thyroglubulin, 670 kDa, b: γ-globulin, 158 kDa, c: ovalbumin, 44 Kda, d: myoglobin, 17 kDa, e: vitamin B12, 1.35 kDa). (**FIG. 19G**) Calibration curve of the Superose 6 Increase column.

[0047] **FIG. 20A-20F** – Taxonomic distribution and domain architectures of Avs families. (**FIG. 20A**) Distribution of avs genes across phyla. The values above the bars indicate the number and percentage of genomes containing each gene. PVC, Planctomycetota, Verrucomicrobiota, and Chlamydiota. The values above the bars indicate the number and percentage of genomes containing each gene. (**FIG. 20B**) Number of bacterial and archaeal phyla (minimum 100 sequenced isolates) with at least one detected instance of an avs gene. (**FIG. 20C**) Kernel density plots of the length distribution of Avs proteins, excluding the N-terminal domain. The red lines, as represented in greyscale, indicate medians. ****$p < 0.0001$ (Mann-Whitney). (**FIG. 20D-20E**) Maximum likelihood tree of representatives of the ATPase + C-terminal domain of (**FIG. 20D**) Avs2 terminase sensors (n = 1,255) and (**FIG. 20E**) Avs4 portal sensors (n = 1,089) clustered at 95% sequence identity. See **FIG. 21B-21C** for the trees for Avs1 and Avs3. Stars on the outer ring indicate homologs investigated experimentally in this study. HTH, helix-turn-helix; MBL, metallo-b-lactamase; REase, restriction endo-nuclease. (**FIG. 20F**) Phage plaque assays showing antiphage defense activity of a chimeric

Avs4 with transmembrane N-terminal helices from Sulfurospirillum sp. replacing the Mrr-like nuclease domain of EcAvs4. The X indicates a nuclease domain mutation.

[0048] **FIG. 21A-21C** – Related to **FIG. 20A-20F**. (**FIG. 21A**) Taxonomic distribution of Avs families, stratified by bacterial and archaeal phylum. The bar graphs show the number of genomes available for analysis. Maximum likelihood phylogenetic trees of the (**FIG. 21B**) Avs1 (n = 843) and (**FIG. 21C**) Avs3 (n = 630) families. Inner, middle, and outer rings indicate taxonomy, N-terminal effectors domains, and locus architecture, respectively.

[0049] **FIG. 22A-22B** – Examples of Avs proteins implicated in protein-protein signaling. (**FIG. 22A**) Predicted caspase recruitment by cyanobacterial Avs2 homologs via an N-terminal EAD10 protein recruitment domain that is also shared by proteins encoded in the vicinity. The tree was constructed from a multiple sequence alignment of the caspase. Protein accession numbers refer to the STAND NTPase. (**FIG. 22B**) An Avs3 homolog within a genomic locus from Sulfurovum sp. enriched in TIR domains related to those mediating second messenger signaling (Ofir et al. Nature 600, 116-120 (2021)).

[0050] **FIG. 23A-23C** – Related to **FIG. 20A-20F**. (**FIG. 23A**) (SEQ ID NO: 8-9) Amino acid sequence surrounding the EcAvs4 chimera break point. (**FIG. 23B**) Chimera activity against phage T7. (**FIG. 23C**) Plasmid depletion assay for the target recognition specificity of the EcAvs4 chimera in comparison with EcAvs4.

[0051] **FIG. 24A-24E** – Phage-encoded genes inhibit Avs activity. (**FIG. 24A**) Schematic of a pooled screen in E. coli for phage early genes that rescue Avs-mediated toxicity. CmR, chloramphenicol resistance gene. (**FIG. 24B**) Deep sequencing readout of anti-defense candidate genes co-expressed with SeAvs3, EcAvs4, or KpAvs4. (**FIG. 24C**) A hypervariable early gene locus within a closely related set of wastewater-isolated Autographiviridae phages contains abundant anti-defense genes. The tree was constructed from a concatenated alignment of conserved proteins present in all ten phages. Greyscale represents groups of proteins clustered at 40% sequence identity at 70% coverage. (**FIG. 24D**) Agarose gel analysis showing in vitro reconstitution of anti-SeAvs3 activity by three antidefense candidates. (**FIG. 24E**) Schematic of the mechanism of Avs proteins as antiphage pattern-recognition receptors.

[0052] **FIG. 25A-25B** – Related to **FIG. 24A-24E**. Antidefense genes inhibit Avs activity in bacterial cells. Plaque assays against (**FIG. 25A**) phage ZL19 and (**FIG. 25B**) phage T7 with E. coli strain C containing both an Avs plasmid and an antidefense plasmid. Antidefense genes

were expressed under the control of a J23105 promoter. Spots correspond to 10-fold dilutions from right to left.

[0053] **FIG. 26** - Mechanism and structures of NLR-like defense proteins in prokaryotes. (Left) Comparison of the domain architectures of 11 representative NLR-like pattern-recognition receptors across four kingdoms of life. Selected structures of activated complexes are shown as examples. T3SS, type 3 secretion system. (Right) Defense mechanism of Avs proteins in bacteria and archaea (this study). Target binding triggers the formation of Avs tetramers, which activates an N-terminal effector that disrupts the viral life cycle.

[0054] **FIG. 27A-27N** - Cryo-EM structures of SeAvs3 and EcAvs4 in complex with their cognate triggers. (**FIG. 27A-27B**) Structure of the SeAvs3-terminase complex. (**FIG. 27C-27D**) Structure of the EcAvs4-portal complex. (**FIG. 27E-27F**) ATP molecule in the STAND ATPase active site of EcAvs4 and SeAvs3. The cryo-EM density is shown as a transparent surface. (**FIG. 27G**) SeAvs3 Cap4-like nuclease effector domain. (**FIG. 27H-27I**) Active sites for the inward- and outward-facing protomers of the SeAvs3 Cap4-like nuclease. (**FIG. 27J**) Equivalent view of the active site of HindIII bound to target DNA with two divalent metal ions [Protein Data Bank (PDB) ID 3A4K]. (**FIG. 27K**) Electrostatic surface potential for the SeAvs3 Cap4-like nuclease and the EcAvs4 Mrr-like nuclease. Active sites are indicated by purple circles. Ideal B-form DNA is modeled on both surfaces based on the crystal structure of Hind III bound to its target (PDB ID 3A4K). (**FIG. 27L**) EcAvs4 Mrr-like nuclease effector domain. (**FIG. 27M-27N**) Active sites for the inward- and outward-facing protomers of the EcAvs4 Mrr-like nuclease.

[0055] **FIG. 28A-28I** - Structural basis for viral-fold recognition by SeAvs3 and EcAvs4. (**FIG. 28A**) The interface between SeAvs3 and the PhiV-1 terminase. An SeAvs3 surface view is shown in transparency. SeAvs3 is colored from the N to C terminus according to the key. (**FIG. 28B**) AlphaFold or crystal structures of different terminases modeled into SeAvs3. The ATPase and nuclease domains were individually aligned to the PhiV-1 terminase domains. (**FIG. 28C-28D**) Recognition of the PhiV-1 terminase ATPase and nuclease active sites by the SeAvs3 TPR domain. (**FIG. 28E**) Sequence logos for terminase ATPase Walker A motifs and terminase nuclease active sites. A total of 11,000 terminase sequences were clustered at 30% sequence identity, and motifs were extracted from clusters containing terminases targeted or not targeted by SeAvs3 according to **FIG. 7B** (see also **FIG. 34**). (**FIG. 28F**) Plasmid depletion assay for SeAvs3 coexpressed in E. coli with a terminase ATPase or nuclease domain harboring

active-site mutations. (**FIG. 28G**) The interface between EcAvs4 and the PhiV-1 portal. An EcAvs4 surface view is shown in transparency. EcAvs4 is colored from the N to C terminus according to the key. (**FIG. 28H**) b-sheet augmentation between EcAvs4 and the portal clip domain. (**FIG. 28I**) Comparison of the EcAvs4-bound state of the PhiV-1 portal, the cryo-EM structure of the highly homologous T7 portal in its native virion, and AlphaFold models of diverse portals. A top view of the assembled dodecamer of the T7 portal is also shown.

[0056] **FIG. 29A-29F** - Imaging Avs proteins by electron microscopy. (**FIG. 29A**) Example cryo-EM micrograph of assembled SeAvs3-gp19 complex. (**FIG. 29B**) Representative 2D class averages of SeAvs3-gp19 from 128,500 automatically picked particles. (**FIG. 29C**) 2D averages from cryo-EM imaging of SeAvs3 alone. One class is shown magnified with the structure of SeAvs3 residues 655 – 2087 superimposed, based on the structure of the SeAvs3-gp19 complex. This dataset did not allow high resolution structure determination, potentially due to inherent flexibility in apo-SeAvs3. (**FIG. 29D**) Example cryo-EM micrograph of purified EcAvs4-gp8 complex. (**FIG. 29E**) Representative 2D class averages of the tetrameric and octameric species of EcAvs4-gp8 from 444,626 automatically picked particles. Also shown are 2D averages from a small screening cryo-EM dataset from the same sample diluted 2-fold, showing only the tetrameric species. (**FIG. 29F**) Avs samples imaged by negative-stain electron microscopy using an FEI Tecnai 12 microscope operated at 120 keV. Samples were applied to continuous carbon and stained using 2% uranyl formate. Avs3 and Avs4 do not assemble into tetramers in the absence of their cognate ligands.

[0057] **FIG. 30A-30B** - Cryo-EM data processing scheme. Flowchart outlining the data processing for (**FIG. 30A**) the SeAvs3-gp19 complex and (**FIG. 30B**) the EcAvs4-gp8 complex. Final maps deposited to the EMDB are highlighted.

[0058] **FIG. 31A-31C** - Cryo-EM data statistics. (**FIG. 31A-31B**) Orientation distributions for reconstructions of the SeAvs3-gp19 terminase complex and EcAvs4-gp8 complex. The range of the x-axis, corresponding to the RELION metadata parameter 'rlnAngleRot,' is set according to the symmetry of the reconstruction. (**FIG. 31C**) Gold-standard Fourier-Shell Correlation curves.

[0059] **FIG. 32A-32C** - Cryo-EM map quality and map-to-model fitting. (**FIG. 32A**) Cryo-EM densities colored by local resolution as calculated within RELION. The overall maps are shown filtered by local resolution, while the focus-refined maps are shown auto-sharpened. Sharpened maps are also shown just around the phage ligands. (**FIG. 32B**) Map-to-model

Fourier-Shell Correlation as calculated in PHENIX, softly masking each map around the fitted model. (**FIG. 32C**) Example cryo-EM densities for different parts of the structures.

[0060] **FIG. 33A-33B** - Comparison of activated STAND structures. (**FIG. 33A**) STAND oligomers from different domains of life (29, 30, 47, 76–79), shaded by function. The ROQ1 resistosome structure is a composite by imposing C4 symmetry on PDB 7JLU and merging it with PDB 7JLV and 7JLX (Martin et al., Science 370, eabd9993 (2020)). The NAIP inflammasome structure is an alignment of the C11 symmetric NLRC4 oligomer (with four subunits hidden) (Zhang et al., Science 350, 404-409 (2015)) with the NAIP-NLRC4-flagellin filament structure (Tenthorey et al., Science 358, 888-893 (2017)). (**FIG. 33B**) Two adjacent STAND ATPase domains from these structures, aligned on the nucleotide-binding domain of one ATPase (blue, as represented in greyscale), showing different relative positions of the adjacent ATPase. NBD; nucleotide-binding domain. HD1; helical domain 1. WHD; winged-helix domain.

[0061] **FIG. 34** - Related to **FIG. 28A-28I**. Weblogos of the Walker A motifs of phage terminases. Each motif represents a cluster of terminases that contain at least one representative that was tested experimentally in this study. Terminase sequences (Esterman et al., Virus Evol. 7, veab015 (2021)) were supplemented with the 24 terminases in this study and clustered at 30% sequence identity. Clusters containing terminases that do not activate SeAvs3 are shown in red. The UPGMA tree was built using a procedure described previously (Makarova et al., Nat. Rev. Microbiol. 18, 67-83 (2020)).

[0062] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE EXAMPLE EMBODIMENTS

General Definitions

[0063] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2nd edition (1989) (Sambrook, Fritsch, and Maniatis); Molecular Cloning: A Laboratory Manual, 4th edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F.M. Ausubel et al. eds.); the series Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach (1995)

(M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.): Antibodies, A Laboratory Manual (1988) (Harlow and Lane, eds.): Antibodies A Laboratory Manual, 2nd edition 2013 (E.A. Greenfield ed.); Animal Cell Culture (1987) (R.I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlett, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (2011).

[0064] As used herein, the singular forms "a", "an", and "the" include both singular and plural referents unless the context clearly dictates otherwise.

[0065] The term "optional" or "optionally" means that the subsequent described event, circumstance or substituent may or may not occur, and that the description includes instances where the event or circumstance occurs and instances where it does not.

[0066] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0067] The terms "about" or "approximately" as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/-10% or less, +/-5% or less, +/-1% or less, and +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier "about" or "approximately" refers is itself also specifically, and preferably, disclosed.

[0068] As used herein, a "biological sample" may contain whole cells and/or live cells and/or cell debris. The biological sample may contain (or be derived from) a "bodily fluid". The present invention encompasses embodiments wherein the bodily fluid is selected from amniotic fluid, aqueous humour, vitreous humour, bile, blood serum, breast milk, cerebrospinal fluid, cerumen (earwax), chyle, chyme, endolymph, perilymph, exudates, feces, female ejaculate, gastric acid, gastric juice, lymph, mucus (including nasal drainage and phlegm), pericardial fluid, peritoneal fluid, pleural fluid, pus, rheum, saliva, sebum (skin oil), semen,

sputum, synovial fluid, sweat, tears, urine, vaginal secretion, vomit and mixtures of one or more thereof. Biological samples include cell cultures, bodily fluids, cell cultures from bodily fluids. Bodily fluids may be obtained from a mammal organism, for example by puncture, or other collecting or sampling procedures.

[0069] The terms "subject," "individual," and "patient" are used interchangeably herein to refer to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, simians, humans, farm animals, sport animals, and pets. Tissues, cells and their progeny of a biological entity obtained in vivo or cultured in vitro are also encompassed.

[0070] Various embodiments are described hereinafter. It should be noted that the specific embodiments are not intended as an exhaustive description or as a limitation to the broader aspects discussed herein. One aspect described in conjunction with a particular embodiment is not necessarily limited to that embodiment and can be practiced with any other embodiment(s). Reference throughout this specification to "one embodiment", "an embodiment," "an example embodiment," means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases "in one embodiment," "in an embodiment," or "an example embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some, but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

[0071] All publications, published patent documents, and patent applications cited herein are hereby incorporated by reference to the same extent as though each individual publication, published patent document, or patent application was specifically and individually indicated as being incorporated by reference.

OVERVIEW

[0072] Bacteria and archaea have evolved numerous defense mechanisms against viral infections involving a wide range of strategies and enzymatic activities (Makarova et al. J.

Bacteriol. 193, 6039–6056 (2011); Makarova et al., Annu. Rev. Microbiol. 71, 233–261 (2017); Doron et al., Science. 359 (2018), doi:10.1126/science.aar4120; and Gao et al., Science. 369, 1077–1084 (2020)). Defense systems are activated by viral nucleic acids, in the case of restriction-modification and CRISPR-Cas systems; or by different types of infection-induced cellular stress, including DNA double-strand breaks (Klainman et al., Nucleic Acids Res. 42, 328–339 (2014)), inhibition of host transcription (Guegler et al., Mol. Cell. 81, 2361–2373.e9 (2021)), cytosolic nucleotide depletion (Cheng et al., Nucleic Acids Res. 49, 5216–5229 (2021)), and the disruption of translation elongation factor EF-Tu (Bingham et al., J. Biol. Chem. 275, 23219–23226 (2000)) or RecBCD repair nuclease (Millman et al., Y. Oppenheimer-Shaanan, R. Sorek, Bacterial Retrons Function In Anti-Phage Defense. Cell. 183, 1551–1561.e12 (2020)). Alternatively, some systems constitutively synthesize small molecules that interfere with phage replication (Kronheim et al., Nature. 564 (2018) 283-286; Bernheim et al., Nature. 589, 120-124 (2021)). However, for numerous defense systems, the mechanisms of activation remain uncharacterized, and it appears likely that distinct modes of activation exist within the diverse repertoire of recently discovered systems (Makarova et al. J. Bacteriol. 193, 6039–6056 (2011); Doron et al., Science. 359 (2018), doi:10.1126/science.aar4120; Gao et al., Science. 369, 1077–1084 (2020); Makarova et al., Nucleic Acids Res. 41, 4360-4377 (2013)). NLRs are among the key proteins involved in immunity, cell signaling, and particularly programmed cell death in eukaryotes (Koonin et al., Cell Death Differ. 9, 394-404 (2002); Leipe et al., J. Mol. Biol. 343, 1-28 (2004); Zhao et al., Nature. 477, 596-600, (2011); Kofoed et al., Nature. 477. 592-595 (2011); Caruso et al., Immunity. 41, 898-908 (2014); Jones et al., Science. 354 (2016), doi:10.112/science.aaf6395; Heller et al. Proc. Natl. Acad. Scie. U.S.A. 115, E2292-E2301 (2018); Bauernfried et la., Science. 371 (2021), doi:10.1126/science.abd0811). In animal and plant innate immunity, these proteins function by recognizing pathogen-associated molecular patterns (PAMPs). Animal NLRs consist of a central STAND NTPase domain, a C-terminal region containing leucine-rich repeats (LRRs) or WD40 repeats, and in many cases, an N-terminal pyrin domain or caspase activation and recruitment domain (CARD). Similarly, plant NLRs contain the STAND domain, a C-terminal LRR array and often an N-terminal TIR (Toll/interleukin-1 receptor) domain. The diverse eukaryotic NLRs recognize many different PAMPs; for example, animal NOD1 and NOD2 proteins recognize peptidoglycan fragments from the bacterial cell wall (Caruso et al., Immunity. 41, 898-908 (2014)), NLRP1 binds viral dsRNA

(Bauernfried et la., Science. 371 (2021), doi:10.1126/science.abd0811), and NAIP detects bacterial flagellin and type 3 secretion systems (Zhao et al., Nature. 477, 596-600, (2011); Kofoed et al., Nature. 477. 592-595 (2011)). In all of these cases, recognition of the PAMP leads to oligomerization of the NLR and recruitment of effector proteins. Bacteria and archaea, especially those with complex signaling systems, also encode a diverse repertoire of STAND NTPases that are predicted to be involved in signal transduction and possibly in programmed cell death (Koonin et al., Cell Death Differ. 9, 394-404 (2002); Leipe et al., J. Mol. Biol. 343, 1-28 (2004)). However, the functions of these proteins are largely unknown, with the exception of several that have been characterized as transcription regulators (Danot et al., Proc. Natl. Acad. Sci. U. S. A. 98, 435–440 (2001); Horinouchi et al. Gene. 95, 49–56 (1990); Ye et al., Microbiol. Mol. Biol. Rev. 84 (2020), doi:10.1128/MMBR.00061-19).

[0073] Applicant demonstrates herein that antiviral STAND (Avs) homologs in bacteria and archaea are pattern recognition receptors that detect conserved viral proteins and activate diverse N-terminal effectors, including DNA endonucleases. This work further reveals remarkable similarity between the defense strategies of prokaryotes and eukaryotes and extends the paradigm of pattern recognition of pathogen-specific proteins across all domains of life. Embodiments disclosed herein provide programmable pattern recognition proteins that are capable of recognizing and binding a molecular pattern. The programmable pattern recognition proteins can have one or more effector domains that can be activated upon pattern recognition. In this way, the programmable pattern recognition proteins can be engineered to specifically recognize a target pattern (i.e., programmed), which can lead to effector activity at or in proximity to the recognized target pattern. Combining different pattern recognition capabilities with different effector functions can provide, without limitation, a modular system with a myriad of utilities such as molecular pattern-based in vitro diagnostics, cargo delivery, therapeutic applications, and microbiome structure engineering. Other embodiments, applications, and uses are described herein and will be appreciated in view of the present exemplary embodiments and working examples herein.

PROGRAMMABLE PATTERN RECOGNITION PROTEINS

[0074] Described in several example embodiments herein are engineered programmable pattern recognition proteins. The programmable pattern recognition proteins comprise an effector domain, an effector activation domain, and a pattern recognition domain, wherein binding of the recognition domain to a target molecule leads to activation of the effector

domain, and wherein at least one of the effector domains, effector activation domain, or pattern recognition domain is derived from a Signal Transduction ATPases with Numerous-associated Domains (STAND) protein. In one example embodiment, the engineered protein comprises a STAND NTPase. In one example embodiment the STAND NTPase functions as the effector activation domain and further comprises an effector domain and a pattern recognition domain derived from the same STAND protein or from an ortholog or homolog thereof. The effector domain may also be a non-STAND effector domain.

[0075] Generally, and without being bound by theory, upon pattern recognition by the engineered protein, the engineered protein is activated. Activation can include activating the STAND NTPase and/or other effector domains of the engineered protein. In some embodiments, the activity of the engineered protein includes nuclease and/or protease activity. Thus, when activated in response to pattern recognition, such as a PAMP or other molecular pattern associated with a target cell or molecule (e.g., a target polypeptide), the engineered protein can have effector function (e.g., nuclease, protease, etc. activity) at the target molecule and/or cell. In some embodiments, such effector function can lead to cell death or cell or molecule modification. Other functions and activities will be appreciated in view of the description herein.

[0076] In some embodiments, the engineered protein has a central STAND NTPase that is flanked by an N-terminal region and/or a C-terminal region. In some embodiments, the N-terminal region has one or more effector domains. In some embodiments, the C-terminal domain comprises one or more structural and/or interaction motifs.

[0077] In some embodiments, the STAND NTPase, the N-terminal region, and/or the C-terminal region can be engineered such that the engineered protein recognizes a specific molecular pattern, has a specific desired effector function in addition to any effector function of the STAND NTPase, and/or has specific interaction capabilities beyond molecular pattern recognition and/or interaction. Without being bound by theory, the protein compositions of the present invention provide the ability to have a modular and programmable composition in which molecular pattern recognition, effector functionality and effector activation can be configured so as to target a particular cell or molecule comprising or otherwise associated with a target molecular pattern and provide a desired effector action at the targeted cell or molecule.

Effector Domains

[0078] In some embodiments, the engineered protein composition of the present invention comprises one or more effector domains. In some embodiments one or more effector domains are derived from a STAND protein. In some embodiments one or more effector domains are derived from a STAND NTPase protein. In some embodiments one or more effector domains are derived from a prokaryotic STAND protein. In some embodiments, one or more effector domains are derived from a prokaryotic STAND NTPase. STAND proteins and STAND NTPAse proteins are discussed and described in greater detail elsewhere herein.

[0079] In some embodiments, one or more effector domains are not derived from a STAND protein and/or STAND NTPase.

[0080] In some embodiments, the N-terminal region, the C-terminal region, or both the N- and the C-terminal regions of the engineered protein comprises the one or more effector domains. In some embodiments, one or more of the effector domains are contained between the N-terminal region and the C-terminal region of the engineered protein.

[0081] In some embodiments, the one or more effector domains are independently selected from a nuclease, a nickase, a protease or peptidase, nucleosidase, a helicase, a methylase, an acetylase, a demethylase, a deacetylase, a transcriptase, a hydrolase, a phosphatase, a phosphorylase, a caspase or caspase like domain, a glycosylase, a lipase, a transferase, any combination thereof, and/or the like.

Exemplary Effector Domains

[0082] Exemplary nucleases include, without limitation, Cas proteins and systems (see e.g., Koonin and Makarova et al., Origins and evolution of CRISPR-Cas systems. Phil. Trans. R. Soc. B374201800872018008 7 and Makarova et al., CRISPR J. 2018 Oct 1; 1(5): 325–336). In some embodiments, the Cas is a Cas having collateral nuclease activity (e.g., a Cas12 or a Cas13). In some embodiments, the Cas is a Cas nickase or dead Cas. In some embodiments, the nuclease is a single stranded DNA (ssDNA) nuclease. In some embodiments, the nuclease is a dsDNA nuclease. In some embodiments, the nuclease is an exonuclease. In some embodiments, the nuclease is an endonuclease. In some embodiments, the nuclease is a circular DNA nuclease. In some embodiments, the nuclease is a linear nuclease. In some embodiments, the nuclease is an RNA nuclease. In some embodiments, one or more effector domains comprise one or more PD-DExK-family nuclease domains. In some embodiments, the nuclease activity is organism and phage independent.

[0083] Exemplary proteases include without limitation, aspartic, glutamic, and metalloproteases, cysteine, serine, and threonine proteases. In some embodiments, the protease/peptidase comprises a TPR and/or CHAT domain. In some embodiments, the protease comprises or is a caspase or caspase like protein or functional domain thereof. In some embodiments, the protease is a bacterial protease or functional domain thereof (See e.g., Culp and Wright. J. Antibiotics. 70:366-377 (2017)). In some embodiments, the protease is a eukaryotic protease or a functional domain thereof (see e.g., Quesada et al. Nuc. Acid. Res. 2009 37:D239-D243).

[0084] In some embodiments, an effector domain comprises nuclease, protease, nucleosidase, sirtunins (SIR2), Toll/interleukin-1 receptor homology (TIR), cytidine monophosphate (CMP) hydrolase and/or caspase-like enzyme activities. In some embodiments, the effector domain comprises dsDNA nuclease activity. In some embodiments, the effector domain comprises circular DNA and/or linear DNA nuclease activity. In some embodiments, the effector is SIR2, TIR, or a CMP hydrolase.

[0085] In some embodiments, the one or more effector domains comprise one or more D-QxK and/or one or more E-Q-QxK catalytic motifs.

Effector Domain Targets

[0086] The target of an effector domain of the engineered programmable pattern recognition protein composition of the present invention can be any target comprising, be fused to, linked to, tethered to, coupled to, or otherwise integrated or associated with a target polypeptide and/or target molecular pattern, optionally a PAMP, that is recognized by the engineered protein composition of the present invention. In some embodiments, the target is a cell. In some embodiments the target is a polypeptide or peptide. In some embodiments, the target is a nucleic acid (e.g., DNA or RNA). In some embodiments the target is a double stranded (ds) nucleic acid, such as dsRNA or dsDNA. In some embodiments the target is a circular DNA.

[0087] In some embodiments, an effector domain acts on the same molecule that contains or is fused to, linked to, tethered to, coupled to, or otherwise integrated or associated with the target polypeptide, target molecule, and/or target molecular pattern recognized and/or bound by the engineered protein composition of the present invention. In some embodiments, an effector domain acts on a molecule that does not contain or is not fused to, linked to, tethered to, coupled to, or otherwise integrated or associated with the target polypeptide, target

molecule, and/or target molecular pattern recognized and/or bound by the engineered protein composition of the present invention. For example, the recognition/binding activity may be used to target the engineered protein composition of the present invention to a specific cell but that the effector function may be carried out on a component of that cell, such as a protein or nucleic acid within the targeted cell not directly containing the target polypeptide, target molecule, and/or target molecular pattern. In another example, the recognition activity may be used to target the engineered protein composition of the present invention to a specific region in an organism, on a device or substrate, such as a region on a microfluidic chip, lateral flow device, or region within an organism. In another example, the effector domains (s) of the engineered composition then may take effect on any substrate molecule (with or without a target polypeptide, target molecule, and/or target molecular pattern) that is within effective proximity of the engineered protein composition of the present invention.

Effector Activation Domains

[0088] The engineered proteins can contain an effector activation domain. Without being bound by theory, the effector activation domain can interact with the recognition domain, target molecule, and/or the effector domain such that the effector domain is activated. In some embodiments, the effector activation domain is or is derived from a STAND protein. In some embodiments, the effector activation is or is derived from a STAND NTPase protein. In some embodiments, the effector activation is or is derived from a prokaryotic STAND protein. In some embodiments, effector activation is or is derived from a prokaryotic STAND NTPase. STAND proteins and STAND NTPAse proteins are discussed and described in greater detail elsewhere herein. In some embodiments, the effector activation domain

[0089] In some embodiments, the N-terminal region, the C-terminal region, or both the N- and the C-terminal regions of the engineered protein comprises the effector activation domain or component thereof. In some embodiments, an effector activation domain is contained between the N-terminal region and the C-terminal region of the engineered protein.

Pattern Recognition Domains

[0090] The engineered proteins contain a pattern recognition domain, which is also referred to herein as a “recognition domain”. In some embodiments, the recognition domain is capable of recognizing and/or binding a target polypeptide, such as once comprising a specific molecular pattern. Exemplary molecular patterns include 2-D and 3D structures. A non-limiting example of a molecular pattern are pathogen-associated molecular patterns, which are

described in further detail below. In some embodiments, the recognition domain contains one or more tetratricopeptide repeat (TPR) domains.

[0091] In some embodiments, the recognition domain or portion thereof is in the N terminal region, C-terminal region, or both of the engineered protein of the present invention. In some embodiments, the recognition domain is contained between the N-terminal region and the C-terminal region of the engineered protein.

[0092] In some embodiments, the recognition domain recognizes a native target polypeptide and/or molecular pattern of wild-type STAND protein. In some embodiments, the recognition domain recognizes a native target polypeptide and/or molecular pattern of wild-type prokaryotic STAND protein. In some embodiments, the recognition domain recognizes a native target polypeptide and/or molecular pattern of wild-type STAND NTPase protein. In some embodiments, the recognition domain recognizes a native target polypeptide and/or molecular pattern of wild-type prokaryotic STAND NTPAse protein. In some embodiments, the recognition domain targets a PAMP recognized by a wild-type STAND protein. In some embodiments, the recognition domain targets a PAMP recognized by a wild-type STAND NTPase protein, optionally a prokaryotic wild-type STAND protein or STAND NTPase protein.

[0093] In some embodiments, the recognition domain is engineered to recognize a target polypeptide and/or molecular pattern other than a native target polypeptide or molecular pattern of a wild-type STAND protein or STAND NTPase protein. In other words, the recognition domain can be engineered to recognize a target polypeptide and/or molecular pattern that is not a native recognition partner (or target) to a wild-type STAND protein or wild-type STAND NTpase protein. In some embodiments, the recognition domain recognizes a target polypeptide and/or molecular pattern that is not a native recognition partner (or target) to a wild-type STAND protein or STAND NTPase. In some embodiments, the recognizes a PAMP that is not a native PAMP for a wild-type STAND protein or STAND NTPase. In some embodiments, the recognizes a PAMP that is not a native PAMP for a wild-type prokaryotic STAND protein or STAND NTPase.

[0094] In some embodiments, the recognition domain is derived from a STAND protein. In some embodiments, recognition domain is derived from a STAND NTPase protein. In some embodiments, the recognition domain is derived from a prokaryotic STAND protein. In some embodiments, the recognition domain is derived from a prokaryotic STAND NTPase. STAND

proteins and STAND NTPAse proteins are discussed and described in greater detail elsewhere herein.

Pathogen-associated Molecular Pattern (PAMP)

[0095] PAMPs are known in the art as molecular motifs that form structural "patterns" whose structure is recognized by receptors and proteins. The term originated from the observation that classes of microbes, particularly pathogenic microbes, contained structural motifs that were recognized by cell receptors that stimulated the immune response. Although the term originated from the study of host-pathogen interaction, it will be appreciated that in the context of the present invention PAMPs are not limited to those relating to pathogenic cells or molecules. As previously mentioned, the engineered protein compositions have molecular pattern recognition activity. In some embodiments, the engineered protein compositions of the present invention have PAMP recognition activity. In other words, in some embodiments the engineered proteins of the present invention can recognize PAMPs. Without being bound by theory, by engineering the protein composition to recognize particular PAMPs, the targets of the protein can be specified. Further, by incorporating different PAMPs target molecule specificity of protein can be engineered to recognize different target molecules. Thus, in some embodiments, the PAMPs recognized by the engineered proteins of the present invention may be native to a target cell or molecule or may be exogenous to the target cell or molecule. Where the PAMPs are exogenous to a target cell or molecule, the PAMPs may be fused to, linked to, tethered to, coupled to, or otherwise integrated or associated with the target cell or molecule.

[0096] In some embodiments, the PAMPs are proteins, peptides, sugars or other carbohydrates, lipopolysaccharides, peptidoglycans, nucleic acids (particularly double stranded variants), and/or the like. It will be appreciated that although PAMPs are traditionally thought of as being associated with pathogens, that PAMPs may also be found or associated with non-pathogenic organisms or cells.

[0097] In some embodiments the PAMP recognized by the engineered protein of the present invention is a large terminase subunit. In some embodiments, the PAMP recognized by the engineered protein present invention is a large terminase subunit of a virus or phage. In some embodiments, the PAMP recognized by the engineered protein of the present invention is gp19 or a structural homologue thereof. In some embodiments, the PAMP recognized by the engineered protein of the present invention is a portal protein. In some embodiments, the portal protein is a viral or a phage portal protein. In some embodiments, the PAMP recognized by the

engineered protein of the present invention is a gp8 portal protein or a structural homologue thereof. Exemplary terminase and portal proteins are shown in **Table 1**. In some embodiments, the PAMP recognized by the engineered protein is or comprises an ATPase domain or portion thereof or a 3-D structural feature thereof and/or a nuclease domain or a portion thereof or 3-D structural feature thereof of a large terminase subunit. In some embodiments, the PAMP recognized by the engineered protein is or comprises an ATPase domain or portion thereof, or a 3-D structural feature thereof and/or a nuclease domain or a portion thereof or 3-D structural feature thereof of a gp19 protein or a structural homologue thereof. In some embodiments, the PAMP recognized by the engineered protein is or comprises an portal protein or portion thereof, or a 3-D structural feature thereof. In some embodiments, the PAMP recognized by the engineered protein is or comprises a gp8 portal protein or structural homologue thereof, a portion thereof or a 3-D structural feature thereof. .

Table 1. Phage terminases and portals tested.

#	Phage	Family	Gene	Accession	Amino Acids	Codon	Note
1	T3	Autographiviridae	Terminase	NP_523347.1	586	Native	
1	T3	Autographiviridae	Portal	YP_009792958.1	535	Native	
2	T7	Autographiviridae	Terminase	NP_042010.1	586	E. coli	
2	T7	Autographiviridae	Portal	NP_041995.1	536	E. coli	
3	PhiV-1	Autographiviridae	Terminase	QLF85798.1	586	E. coli	
3	PhiV-1	Autographiviridae	Portal	QLF85783.1	535	E. coli	
4	Pisces	Autographiviridae	Terminase	QEG09602.1	587	E. coli	
4	Pisces	Autographiviridae	Portal	QEG09587.1	522	E. coli	
5	RsoP1EGY	Autographiviridae	Terminase	YP_009799710.1	561	E. coli	Ralstonia phage RsoP1EGY
5	RsoP1EGY	Autographiviridae	Portal	YP_009799742.1	519	E. coli	Ralstonia phage RsoP1EGY
6	Uncultured	Autographiviridae	Terminase	CAB4140543.1	551	E. coli	Uncultured Caudovirales phage
6	Uncultured	Autographiviridae	Portal	CAB4140338.1	518	E. coli	Uncultured Caudovirales phage
7	S. flava*	Autographiviridae	Terminase	WP_079638097.1	589	E. coli	Sphingopyxis flava R11H (prophage)
7	S. flava*	Autographiviridae	Portal	WP_079638083.1	520	E. coli	Sphingopyxis flava R11H (prophage)
8	D. archaeon*	Autographiviridae	Terminase	OYT57761.1	559	E. coli	Desulfurococcales archaeon ex4484_217_2 (prophage)
8	D. archaeon*	Autographiviridae	Portal	OYT57739.1	533	E. coli	Desulfurococcales archaeon ex4484_217_2 (prophage)
9	SP6	Autographiviridae	Terminase	NP_853601.1	631	E. coli	
9	SP6	Autographiviridae	Portal	NP_853590.1	515	E. coli	
10	ECBP5	Autographiviridae	Terminase	YP_009146419.1	616	E. coli	
10	ECBP5	Autographiviridae	Portal	YP_009146408.1	530	E. coli	
11	PGT2	Autographiviridae	Terminase	YP_009795474.1	602	E. coli	

11	PGT2	Autographiviridae	Portal	YP_009795464.1	517	E. coli	
12	Minorna	Autographiviridae	Terminase	YP_009820994.1	618	E. coli	
12	Minorna	Autographiviridae	Portal	YP_009821007.1	531	E. coli	
13	Lidtsur	Autographiviridae	Terminase	YP_009821647.1	580	E. coli	
13	Lidtsur	Autographiviridae	Portal	YP_009821637.1	521	E. coli	
14	E. coli-1*	Unknown	Terminase	WP_001619041.1	536	E. coli	E. coli NCTC9020 (prophage)
14	E. coli-1*	Unknown	Portal	WP_059328803.1	541	E. coli	E. coli NCTC9020 (prophage)
15	N4	Schitoviridae	Terminase	YP_950546.1	530	E. coli	
15	N4	Schitoviridae	Portal	YP_950537.1	763	E. coli	
16	ZL19	Drexlerviridae	Terminase	YP_009789200.1	521	Native	
16	ZL19	Drexlerviridae	Portal		422	Native	
17	ST32	Chaseviridae	Terminase	YP_009790711.1	672	E. coli	
17	ST32	Chaseviridae	Portal	YP_009790713.1	438	E. coli	
18	E. coli-2*	Podoviridae	Terminase	WP_000132532.1	489	E. coli	E. coli M885 (prophage)
18	E. coli-2*	Podoviridae	Portal	WP_000852419.1	559	E. coli	E. coli M885 (prophage)
19	P1	Myoviridae	Terminase	WP_000124150.1	494	Native	
19	P1	Myoviridae	Portal		569	Native	
20	Lambda	Siphoviridae	Terminase	NP_040581.1	641	Native	
20	Lambda	Siphoviridae	Portal		512	Native	Mature version following proteolytic cleavage
21	P22	Podoviridae	Terminase	WP_015975189.1	499	E. coli	
21	P22	Podoviridae	Portal	WP_015975195.1	725	E. coli	
22	T5	Demerecviridae	Terminase	YP_006983.1	438	Native	
22	T5	Demerecviridae	Portal		394	Native	Mature version following proteolytic cleavage
23	CBA120	Ackermannviridae	Terminase	YP_004957856.1	736	E. coli	
23	CBA120	Ackermannviridae	Portal	YP_004957852.1	560	E. coli	
24	T4	Myoviridae	Terminase	NP_049776.1	610	Native	
24	T4	Myoviridae	Portal	WP_015969330.1	524	Native	
See also Data S3 of Gao et al., Science 377, eabm4096 (2022), which is incorporated herein by reference as if expressed in its entirety herein.							

STAND NTPases

[0098] As discussed elsewhere herein, the engineered protein composition comprises a STAND protein or component thereof. In some embodiments, the STAND protein or component thereof is a STAND NTPase. In some embodiments, the engineered protein comprises components derived from an Avs (anti-viral STAND) or a homolog thereof. In some embodiments, the STAND NTPase is an Avs NTPase. In some embodiments, the Avs comprises Avs1-4 protein families as shown in **Fig. 1A**. In some embodiments, the Avs is an Avs1, Avs2, Avs3, or Avs4 protein or a protein from an Avs1, Avs2, Avs3, Avs4 protein family as sown in **Fig. 1A**. In some embodiments, the Avs protein or homologs thereof further comprise an N-terminal effector domain, and a PAMP recognition region comprising a central core region and a C-terminal tetratricopeptide repeat (TPR) domain. In some embodiments the

PAMP recognition region comprising the central core region is or comprises a STAND NTPase.

[0099] In some embodiments, the engineered protein comprises a protein, a STAND Protein, STAND NTPase, Avs or homologue thereof that is 80-100 percent identical to any a protein, a STAND Protein, STAND NTPase, Avs or homologue thereof of any one or more of **Tables 2, 3, 4, 5, 6, and 7**. In some embodiments, the engineered protein comprises a protein, a STAND Protein, STAND NTPase, Avs or homologue thereof that is 80% to/or 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100% identical to a protein, a STAND Protein, STAND NTPase, Avs or homologue thereof of any one or more of **Tables 2, 3, 4, 5, 6, and 7** or is a homolog thereof. In some embodiments, the protein, STAND Protein, STAND NTPase, Avs, or homologue thereof is from an organism having a genome as in **Data S8** of Gao et al. "Prokaryotic innate immunity via pattern recognition of conserved viral proteins," Science, 377, eabm4096 (2022), which is incorporated by reference as if expressed in its entirety herein.

[0100] In some embodiments, the Avs1-4 recognize PAMPs in phage proteins such as gp19, a large terminase subunit, and gp8, a portal protein. Other target PAMPs are described in greater detail elsewhere herein. In some embodiments, Avs1-3 recognize PAMPs in gp19, and Avs4, recognize PAMPs in gp8. Other target PAMPs are described in greater detail elsewhere herein.

Table 2. STAND ATPases tested.

#	Name	Category	Accession	Amino Acids	Organism	NTD	# ORFs	Promoter	Codon	Active in portal/te-rminase screen	Target
1	SeAvs1	Avs1	ECC9552481.1	1909	Salmonella enterica NCTC10718	Cap4	1	Native	Native	+	Terminase
2	EcAvs1	Avs1	WP_060615938.1	2016	Escherichia coli NCTC9036	Cap4	1	Native	Native	+	Terminase
3	EpAvs1	Avs1	WP_048696970.1	1939	Erwinia piriflorinigrans CFBP5888	MBL + serine protease	3	bla	Native	+	Terminase
4	EcAvs2	Avs2	WP_063118745.1	1484	Escherichia coli NCTC9087	Unknown	1	Native	Native	+	Terminase (ATPase domain)
5	Ec2Avs2	Avs2	WP_001017806.1	1375	Escherichia coli NCTC10650	Mrr	1	Native	Native	+	Terminase (ATPase

											domain)
6	SeAvs3	Avs3	WP_126523998.1	2092	Salmonella enterica NCTC13175	Cap4	2	Native	Native	+	Terminase
7	KvAvs3	Avs3	WP_139964370.1	2137	Klebsiella variicola PF-1	Cap4	1	Native	Native	+	Terminase
8	AcAvs3	Avs3	WP_148571801.1	2085	Arcobacter cryaerophilus M830A	TIR	1	lac	E. coli	–	
9	EcAvs4	Avs4	WP_044068927.1	1587	Escherichia coli NCTC11132	Mrr	1	Native	Native	+	Portal
10	Ec2Avs4	Avs4	EEW5978513.1	1571	Escherichia coli 382_h	Mrr	1	Native	Native	+	Portal
11	KpAvs4	Avs4	WP_087775949.1	1700	Klebsiella pneumoniae UNMC_8078	Cap4	1	Native	Native	+	Portal
12	CcAvs4	Avs4	WP_083892326.1	1757	Coralloccus coralloides DSM2259	Nucleosidase	1	lac	E. coli	+	Portal
13	Ec3Avs4	Avs4	WP_054622834.1	1727	Escherichia coli NCTC11132	Cap4	1	Native	Native	+	Portal
14	SsAvs4	Avs4	WP_181361350.1	1568	Streptomyces sp. A244	TIR	1	lac	E. coli	–	
15		Avs4	HEA06652.1	1802	Archaeon isolate HyVt-335	Metallophosphoesterase	1	lac	E. coli	–	
16		Avs5	WP_001515187.1	769	Escherichia coli NCTC13384	SIR2	1	Native	Native	–	
17		Other	WP_023893333.1	1618	Salmonella enterica NCTC9912	Cap4	2	Native	Native	–	
18		Other	WP_079981830.1	1523	Salmonella enterica NCTC8785	Cap4	1	Native	Native	–	
19		Other	WP_000963924.1	1283	Escherichia coli NCTC9091	Mrr	1	Native	Native	–	
20		Other	WP_001543765.1	932	Escherichia coli NCTC11132	Cap4	1	Native	Native	–	
See also Data S2 of Gao et al., Science 377, eabm4096 (2022), which is incorporated herein by reference as if expressed in its entirety herein.											

Table 3. Avs1 homologs and domain architectures.

#	Accession	Name	Organism	Phylum	*A*	*B*	*C*	*D*	*E*	*F*	*G*

1	WP_116657571.1		Pseudomonas sp. NDM	Proteobacteria	1	1	Cap4_split1	1	WP_116657570.1	480	1701
2	WP_169904304.1		Pseudomonas	Proteobacteria	2	1	Cap4_split1	1	WP_169904303.1	496	1717
3	WP_032654422.1		Pseudomonas syringae	Proteobacteria	3	2	Cap4_split1	1	WP_032654423.1	480	1701
4	WP_110392720.1		Pseudomonas sp. NP28-5	Proteobacteria	3	2	Cap4_split1	0			
5	WP_005756314.1		Pseudomonas amygdali	Proteobacteria	4	2	Cap4_split1	1	WP_005756312.1	480	1701
6	WP_054088810.1		Pseudomonas syringae group genomosp. 3	Proteobacteria	4	2	Cap4_split1	0			
7	WP_074805577.1		Pseudomonas syringae	Proteobacteria	5	3	Cap4_split1	1	WP_046463948.1	480	1701
8	WP_122686209.1		Pseudomonas viridiflava	Proteobacteria	5	3	Cap4_split1	0			
9	WP_053486411.1		Pseudomonas coronafaciens	Proteobacteria	5	3	Cap4_split1	0			
10	WP_103310294.1		unclassified Pseudomonas	Proteobacteria	6	1	Cap4_split1	1	WP_103310295.1	479	1700
11	WP_122431875.1		Pseudomonas viridiflava	Proteobacteria	7	2	Cap4_split1	1	WP_122431874.1	480	1701
12	WP_191892790.1		Pseudomonas coronafaciens	Proteobacteria	7	2	Cap4_split1	0			
13	WP_115284356.1		Pseudomonas	Proteobacteria	8	1	Cap4_split1	1	WP_115284355.1	479	1700
14	WP_136492832.1		Pseudomonas atacamensis	Proteobacteria	9	1	Cap4_split1	1	WP_136492833.1	479	1700
15	WP_203321224.1		Pseudomonas aeruginosa	Proteobacteria	10	1	Cap4_split1	1	WP_203321223.1	479	1700
16	WP_043050031.1		Pseudomonas fluorescens group	Proteobacteria	11	2	Cap4_split1	1	WP_043050030.1	479	1700
17	WP_045791674.1		Pseudomonas simiae	Proteobacteria	11	2	Cap4_split1	0			
18	WP_177111197.1		Pseudomonas reactans	Proteobacteria	12	1	Cap4_split1	1	WP_177111196.1	496	1717
19	WP_015275432.1		Pseudomonas	Proteobacteria	13	6	Cap4_split1	1	WP_015275433.1	480	1701
20	MBH3549593.1		Pseudomonas aeruginosa	Proteobacteria	13	6	Cap4_split1	0			
21	MBG4810009.1		Pseudomonas aeruginosa	Proteobacteria	13	6	Cap4_split1	0			
22	WP_150607813.1		Pseudomonas oleovorans	Proteobacteria	13	6	Cap4_split1	0			
23	HAR05290.1		Pseudomonas sp.	Proteobacteria	13	6	Cap4_split1	0			
24	WP_200632549.1		Pseudomonas stutzeri	Proteobacteria	13	6	Cap4_split1	0			
25	WP_078458415.1		Pseudomonas	Proteobacteria	14	6	Cap4_split1	1	WP_078458416.1	481	1702
26	WP_044872471.1		Pseudomonas sp. LFM046	Proteobacteria	14	6	Cap4_split1	0			
27	WP_043157822.1		Pseudomonas aeruginosa	Proteobacteria	14	6	Cap4_split1	0			
28	WP_203305947.1		Pseudomonas aeruginosa	Proteobacteria	14	6	Cap4_split1	0			
29	WP_033992020.1		Pseudomonas aeruginosa	Proteobacteria	14	6	Cap4_split1	0			
30	WP_128716627.1		Pseudomonas aeruginosa	Proteobacteria	14	6	Cap4_split1	0			
31	MAC13301.1		Alcanivorax sp.	Proteobacteria	15	1	Cap4_split1	1	MAC13302.1	478	1702
32	CAE68177079.1		Paraburkholderia aspalathi	Proteobacteria	16	1	Cap4	1			1700
33	WP_176314510.1		Burkholderia vietnamiensis	Proteobacteria	17	2	Cap4_split1	1	WP_208456383.1*	480	1701
34	WP_183708088.1		Paraburkholderia tropica	Proteobacteria	17	2	Cap4_split1	0			

35	WP_175832586.1		Burkholderia cenocepacia	Proteobacteria	18	1	Cap4_split1	1	WP_208450643.1*	480	1699
36	WP_213781613.1		Caballeronia sp. dw_276	Proteobacteria	19	1	Cap4_split1	1	WP_213781614.1*	480	1701
37	WP_212682614.1		Undibacterium sp. BYS107W	Proteobacteria	20	1	Cap4_split1	1	WP_212682613.1*	480	1701
38	OEZ62099.1		Duganella sp. HH105	Proteobacteria	21	1	Cap4	1			1701
39	WP_059412799.1		Cupriavidus basilensis	Proteobacteria	22	1	Cap4_split1	1	WP_197090381.1*	480	1701
40	WP_195795245.1		Kinneretia sp. DAIF2	Proteobacteria	23	1	Cap4_split1	1	WP_213086214.1*	480	1702
41	WP_011514896.1		Cupriavidus	Proteobacteria	24	6	Cap4_split1	1	WP_029309942.1	480	1702
42	CAI30193.1		Cupriavidus metallidurans CH34	Proteobacteria	24	6	Cap4_split1	0			
43	WP_104670280.1		Cupriavidus metallidurans	Proteobacteria	24	6	Cap4_split1	0			
44	CUJ62901.1		Achromobacter xylosoxidans	Proteobacteria	24	6	Cap4_split1	0			
45	WP_054449387.1		Achromobacter xylosoxidans	Proteobacteria	24	6	Cap4_split1	0			
46	WP_169272799.1		Achromobacter sp. Bel	Proteobacteria	24	6	Cap4_split1	0			
47	WP_179632436.1		Sphaerotilus montanus	Proteobacteria	25	1	Cap4_split1	1	WP_179632435.1	480	1701
48	WP_174251182.1		Rhodoferax sp. BAB1	Proteobacteria	26	1	Cap4_split1	1	WP_174251183.1	480	1702
49	MBN8769362.1		Stenotrophomonas sp.	Proteobacteria	27	1	Cap4_split1	1	MBN8793399.1*	480	1701
50	WP_111588971.1		Aeromonas salmonicida	Proteobacteria	28	2	Cap4_split1	1	WP_111589010.1	480	1695
51	WP_059112482.1		Aeromonas salmonicida	Proteobacteria	28	2	Cap4_split1	0			
52	WP_202014778.1		Aeromonas caviae	Proteobacteria	29	1	Cap4_split1	1	WP_202014780.1*	480	1695
53	WP_103251931.1		Aeromonas veronii	Proteobacteria	30	1	Cap4_split1	1	WP_103251938.1	480	1701
54	WP_133843332.1		Erwinia rhapontici	Proteobacteria	31	1	Cap4_split1	1	WP_133843511.1	480	1701
55	WP_004883354.1		Klebsiella pneumoniae	Proteobacteria	32	13	Cap4_split1	1	WP_004883356.1	480	1701
56	WP_174510509.1		Klebsiella oxytoca	Proteobacteria	32	13	Cap4_split1	0			
57	WP_117252345.1		Klebsiella	Proteobacteria	32	13	Cap4_split1	0			
58	WP_097416920.1		Escherichia coli	Proteobacteria	32	13	Cap4_split1	0			
59	WP_050870084.1		Yersinia enterocolitica	Proteobacteria	32	13	Cap4_split1	0	WP_050870083.1	480	
60	WP_033145356.1		Enterobacter	Proteobacteria	32	13	Cap4_split1	0	WP_033145357.1	480	
61	WP_170898484.1		Escherichia coli	Proteobacteria	32	13	Cap4_split1	0			
62	WP_181501804.1		Citrobacter freundii	Proteobacteria	32	13	Cap4_split1	0			
63	WP_050917582.1		Yersinia enterocolitica	Proteobacteria	32	13	Cap4_split1	0			
64	WP_050148877.1		Yersinia	Proteobacteria	32	13	Cap4_split1	0			
65	OWF69923.1		Yersinia frederiksenii	Proteobacteria	32	13	Cap4_split1	0			
66	WP_050078135.1		Yersinia intermedia	Proteobacteria	32	13	Cap4_split1	0			
67	WP_050915785.1		Yersinia enterocolitica	Proteobacteria	32	13	Cap4_split1	0			
68	WP_050154071.1		Salmonella enterica	Proteobacteria	33	2	Cap4_split1	1	WP_050154073.1	480	1701

69	EGF2641 708.1		Escherichia coli	Proteobacteria	33	2	Cap4_spl it1	0			
70	WP_1549 29802.1		Klebsiella grimontii	Proteobacteria	34	4	Cap4_spl it1	1	WP_154929 906.1	48 0	17 01
71	WP_1816 83925.1		Citrobacter sp. RHB25-C09	Proteobacteria	34	4	Cap4_spl it1	0			
72	ECD9477 175.1		Salmonella enterica subsp. houtenae	Proteobacteria	34	4	Cap4_spl it1	0			
73	EAQ0528 334.1		Salmonella enterica	Proteobacteria	34	4	Cap4_spl it1	0			
74	WP_0931 79900.1		Variovorax sp. YR266	Proteobacteria	35	1	Cap4_spl it1	1	WP_093179 841.1	48 0	17 00
75	WP_0932 43534.1		Variovorax sp. EL159	Proteobacteria	36	1	Cap4_spl it1	1	WP_093243 537.1	48 0	17 05
76	WP_2053 67580.1		Variovorax sp. PDNC026	Proteobacteria	37	1	Cap4_spl it1	1	WP_205367 915.1*	48 0	17 03
77	MBS4018 228.1		Dechloromonas sp.	Proteobacteria	38	1	Cap4_spl it1	1	MBS40182 27.1	48 0	17 22
78	WP_1244 15356.1		Pseudomonas sp. R4-35-07	Proteobacteria	39	3	Cap4_spl it1	1	WP_124415 357.1	48 0	17 25
79	WP_0032 16565.1		Pseudomonas fluorescens	Proteobacteria	39	3	Cap4_spl it1	0			
80	WP_1693 55762.1		Pseudomonas sp. ADAK18	Proteobacteria	39	3	Cap4_spl it1	0			
81	WP_1118 77001.1		Acidovorax anthurii	Proteobacteria	40	1	Cap4_spl it1	1	WP_211322 407.1*	48 0	17 22
82	WP_0411 68053.1		Pseudomonas putida	Proteobacteria	41	11	Cap4_spl it1	1	WP_041167 757.1	48 0	17 25
83	WP_1628 83232.1		Pseudomonas cichorii	Proteobacteria	41	11	Cap4_spl it1	0			
84	WP_2006 24679.1		Pseudomonas sp. MF6747	Proteobacteria	41	11	Cap4_spl it1	0			
85	OUM308 80.1		Pseudomonas sp. 1239	Proteobacteria	41	11	Cap4_spl it1	0			
86	WP_1225 23538.1		Pseudomonas viridiflava	Proteobacteria	41	11	Cap4_spl it1	0			
87	WP_1681 86830.1		Pseudomonas sp. 1239	Proteobacteria	41	11	Cap4_spl it1	0			
88	WP_1204 48543.1		Pseudomonas fluorescens	Proteobacteria	41	11	Cap4_spl it1	0			
89	WP_1506 35179.1		Pseudomonas fluorescens	Proteobacteria	41	11	Cap4_spl it1	0			
90	WP_1867 23347.1		Pseudomonas sp. RW1P2	Proteobacteria	41	11	Cap4_spl it1	0			
91	BAN5484 8.1		Pseudomonas putida NBRC 14164	Proteobacteria	41	11	Cap4_spl it1	0			
92	WP_1002 63782.1		Pseudomonas sp. WCS365	Proteobacteria	41	11	Cap4_spl it1	0			
93	WP_2102 24966.1		Pseudomonas aeruginosa	Proteobacteria	42	2	Cap4_spl it1	1	WP_210224 965.1	48 0	17 23
94	WP_1241 27276.1		Pseudomonas aeruginosa	Proteobacteria	42	2	Cap4_spl it1	0			
95	VVN9879 1.1		Pseudomonas fluorescens	Proteobacteria	43	1	Cap4	1			17 19
96	WP_1147 31527.1		Pseudomonas fluorescens	Proteobacteria	44	1	Cap4_spl it1	1	WP_114731 528.1	48 0	17 13
97	WP_1243 26117.1		Pseudomonas chlororaphis	Proteobacteria	45	2	Cap4_spl it1	1	WP_124303 545.1	48 0	17 19
98	WP_1243 03547.1		Pseudomonas chlororaphis	Proteobacteria	45	2	Cap4_spl it1	0			
99	WP_0274 85560.1		Rhodanobacter sp. OR87	Proteobacteria	46	1	Cap4_spl it1	1	WP_196806 989.1*	48 0	17 01
100	WP_1840 36709.1		Chitinivorax tropicus	Proteobacteria	47	1	Cap4_spl it1	1	WP_184036 706.1	48 0	17 01
101	WP_0604 56796.1		Serratia marcescens	Proteobacteria	48	5	Cap4_spl it1	1	WP_072269 983.1	48 1	17 03
102	WP_1299 38599.1		Serratia proteamaculans	Proteobacteria	48	5	Cap4_spl it1	0			

103	WP_073529324.1		Serratia marcescens	Proteobacteria	48	5	Cap4_split1	0			
104	WP_085115530.1		Serratia proteamaculans	Proteobacteria	48	5	Cap4_split1	0			
105	WP_057533514.1		Yersinia pseudotuberculosis	Proteobacteria	48	5	Cap4_split1	0			
106	MBN5205556.1		Serratia marcescens	Proteobacteria	49	1	Cap4_split1	1	MBN5205555.1	481	1703
107	MBG271121.1		Proteus mirabilis	Proteobacteria	50	1	Cap4_split1	1	MBG2711120.1	476	1698
108	WP_193401433.1		unclassified Pectobacterium	Proteobacteria	51	1	Cap4_split1	1	WP_193401432.1	480	1702
109	WP_001287788.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	1	WP_000999054.1	482	1705
110	WP_126508643.1		Enterobacter hormaechei	Proteobacteria	52	38	Cap4_split1	0			
111	HBA8817055.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
112	WP_097443991.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
113	WP_141419883.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
114	WP_105455445.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
115	WP_116548318.1		Klebsiella aerogenes	Proteobacteria	52	38	Cap4_split1	0			
116	WP_045413315.1		Klebsiella aerogenes	Proteobacteria	52	38	Cap4_split1	0			
117	HAU4369245.1		Citrobacter amalonaticus	Proteobacteria	52	38	Cap4_split1	0			
118	EGI7185241.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
119	WP_046886077.1		Enterobacter cloacae	Proteobacteria	52	38	Cap4_split1	0			
120	MBS8909098.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
121	WP_121827608.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
122	HBB0231624.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
123	EGS1809788.1		Enterobacter cloacae	Proteobacteria	52	38	Cap4_split1	0			
124	EFK2887574.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
125	WP_089540885.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
126	WP_032177813.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
127	WP_110249092.1		Citrobacter freundii	Proteobacteria	52	38	Cap4_split1	0			
128	HAU9189547.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
129	EFC9633343.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
130	WP_097724975.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
131	EFE7703410.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
132	EFN7232166.1		Escherichia coli O26:H32	Proteobacteria	52	38	Cap4_split1	0			
133	HAY3962231.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
134	WP_001567850.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
135	EFJ9660999.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			
136	EFB2128891.1		Escherichia coli	Proteobacteria	52	38	Cap4_split1	0			

137	WP_1171 10702.1		Klebsiella variicola	Proteobacteria	52	38	Cap4_spl it1	0			
138	WP_0721 22142.1		Klebsiella variicola	Proteobacteria	52	38	Cap4_spl it1	0			
139	WP_1171 47673.1		Klebsiella variicola	Proteobacteria	52	38	Cap4_spl it1	0			
140	WP_2106 97501.1		Klebsiella variicola	Proteobacteria	52	38	Cap4_spl it1	0			
141	EBQ9598 946.1		Salmonella enterica subsp. enterica serovar Carmel	Proteobacteria	52	38	Cap4_spl it1	0			
142	EAZ0395 408.1		Salmonella enterica	Proteobacteria	52	38	Cap4_spl it1	0			
143	WP_1608 63016.1		Enterobacter hormaechei	Proteobacteria	52	38	Cap4_spl it1	0			
144	EEJ24472 89.1		Salmonella enterica subsp. enterica	Proteobacteria	52	38	Cap4_spl it1	0			
145	EEH9717 241.1		Salmonella enterica subsp. enterica	Proteobacteria	52	38	Cap4_spl it1	0			
146	ECK2142 237.1		Salmonella enterica subsp. enterica serovar Enteritidis	Proteobacteria	52	38	Cap4_spl it1	0			
147	WP_0395 03612.1		Pectobacterium brasiliense	Proteobacteria	53	4	Cap4_spl it1	1	WP_039503 614.1	48 1	17 04
148	WP_1136 25649.1		Pectobacterium peruviense	Proteobacteria	53	4	Cap4_spl it1	0			
149	WP_1224 57866.1		Pantoea agglomerans	Proteobacteria	53	4	Cap4_spl it1	0			
150	WP_1912 33258.1		Citrobacter freundii	Proteobacteria	53	4	Cap4_spl it1	0			
151	WP_0696 70427.1		Shewanella colwelliana	Proteobacteria	54	1	Cap4_spl it1	1	WP_069670 357.1	48 0	17 02
152	WP_1985 95778.1		Vibrio breoganii	Proteobacteria	55	2	Cap4_spl it1	1	WP_102547 329.1	48 0	17 03
153	PML6118 9.1		Vibrio breoganii	Proteobacteria	55	2	Cap4_spl it1	0			
154	WP_1116 09133.1		Marinomonas arctica	Proteobacteria	56	1	Cap4_spl it1	1	WP_111609 134.1	48 0	16 95
155	WP_1316 80797.1		Providencia rettgeri	Proteobacteria	57	4	Cap4_spl it1	1	WP_131680 796.1	48 0	16 96
156	MBG601 7503.1		Proteus mirabilis	Proteobacteria	57	4	Cap4_spl it1	0			
157	WP_0674 23683.1		Morganella psychrotolerans	Proteobacteria	57	4	Cap4_spl it1	0			
158	WP_2108 35482.1		Providencia rettgeri	Proteobacteria	57	4	Cap4_spl it1	0			
159	HAS6043 641.1		Vibrio vulnificus	Proteobacteria	58	1	Cap4_spl it1	1	HAS604364 2.1	48 0	17 00
160	WP_1717 31032.1		Vibrio sp. 070316B	Proteobacteria	59	1	Cap4_spl it1	1	WP_171731 031.1	48 0	17 00
161	WP_1685 23419.1		Vibrio aestuarianus	Proteobacteria	60	5	Cap4_spl it1	1	WP_168523 417.1	48 0	17 00
162	WP_2118 67640.1		Pseudoalteromonas sp. M8	Proteobacteria	60	5	Cap4_spl it1	0			
163	WP_1511 82891.1		Photobacterium damselae	Proteobacteria	60	5	Cap4_spl it1	0			
164	WP_1382 61402.1		Photobacterium damselae	Proteobacteria	60	5	Cap4_spl it1	0			
165	WP_1687 86834.1		Vibrio aestuarianus	Proteobacteria	60	5	Cap4_spl it1	0			
166	MBQ073 2401.1		Oleispira antarctica	Proteobacteria	61	1	Cap4_spl it1	1	MBQ07324 02.1	48 2	17 07
167	WP_0870 86685.1		Caballeronia catudaia	Proteobacteria	62	2	Cap4_spl it1	1	--	48 2	16 95
168	WP_1161 46723.1		Paraburkholderia sp. BL27I4N3	Proteobacteria	62	2	Cap4_spl it1	0			

169	WP_157814774.1		Janthinobacterium sp. 64	Proteobacteria	63	1	Cap4	1			1690
170	WP_189347147.1		Undibacterium macrobrachii	Proteobacteria	64	1	Cap4	1			1693
171	WP_080418657.1		Burkholderia ubonensis	Proteobacteria	65	2	Cap4_split1	1	WP_157646956.1	486	1702
172	KVO08517.1		Burkholderia ubonensis	Proteobacteria	65	2	Cap4_split1	0			
173	WP_185735829.1		Burkholderia cepacia	Proteobacteria	66	2	Cap4_split1	1	WP_124909626.1	481	1697
174	RQZ76976.1		Burkholderia cepacia	Proteobacteria	66	2	Cap4_split1	0			
175	WP_166894642.1		unclassified Burkholderia	Proteobacteria	67	1	Cap4_split1	1	WP_166894644.1	481	1697
176	WP_059448035.1		Burkholderia territorii	Proteobacteria	68	2	Cap4_split1	1	WP_155623943.1	481	1697
177	WP_176327889.1		Burkholderia vietnamiensis	Proteobacteria	68	2	Cap4_split1	0			
178	WP_176322019.1		Burkholderia vietnamiensis	Proteobacteria	69	1	Cap4_split1	1	WP_176322018.1*	481	1697
179	WP_012470189.1		Geobacter lovleyi	Proteobacteria	70	1	Cap4	1			1694
180	WP_188251732.1		Pseudomonas sp. PSB1	Proteobacteria	71	1	Cap4_split1	1	WP_188251733.1	488	1691
181	WP_019459208.1		Acinetobacter sp. GG2	Proteobacteria	72	1	Cap4	1			1654
182	HAV57686679.1		Acinetobacter baumannii	Proteobacteria	73	1	Cap4	1			1654
183	WP_078202157.1		Acinetobacter baumannii	Proteobacteria	74	2	Cap4	1			1654
184	HAV55860358.1		Acinetobacter baumannii	Proteobacteria	74	2	Cap4	0			
185	WP_017481124.1		Acinetobacter pittii	Proteobacteria	75	1	Cap4	1			1654
186	WP_195772999.1		Acinetobacter pollinis	Proteobacteria	76	1	Cap4	1			1654
187	WP_191017497.1		Acinetobacter seifertii	Proteobacteria	77	1	Cap4	1			1671
188	WP_159414855.1		Acinetobacter gyllenbergii	Proteobacteria	78	2	Cap4	1			1667
189	GGH30620.1		Acinetobacter courvalinii	Proteobacteria	78	2	Cap4	0			
190	WP_055415338.1		Acinetobacter soli	Proteobacteria	79	1	Cap4	1			1676
191	WP_151959007.1		Acinetobacter junii	Proteobacteria	80	5	Cap4	1			1676
192	WP_180034312.1		Acinetobacter sp. YH12233	Proteobacteria	80	5	Cap4	0			
193	MBP9785987.1		Acinetobacter sp.	Proteobacteria	80	5	Cap4	0			
194	WP_136700480.1		Acinetobacter junii	Proteobacteria	80	5	Cap4	0			
195	WP_075696496.1		Acinetobacter junii	Proteobacteria	80	5	Cap4	0			
196	WP_034688549.1		Acinetobacter	Proteobacteria	81	4	Cap4	1			1677
197	WP_034671215.1		Acinetobacter sp. 1461402	Proteobacteria	81	4	Cap4	0			
198	WP_034680054.1		Acinetobacter sp. 983759	Proteobacteria	81	4	Cap4	0			
199	WP_034675232.1		Acinetobacter sp. 230853	Proteobacteria	81	4	Cap4	0			
200	ECC9552481.1	SeAvs1	Salmonella enterica subsp. salamae	Proteobacteria	82	6	Cap4	1			1656
201	EAT2182975.1		Salmonella enterica	Proteobacteria	82	6	Cap4	0			
202	WP_200028719.1		Citrobacter koseri	Proteobacteria	82	6	Cap4	0			

203	WP_200109620.1		Klebsiella aerogenes	Proteobacteria	82	6	Cap4	0			
204	EDW2793192.1		Salmonella enterica	Proteobacteria	82	6	Cap4	0			
205	MBB1201215.1		Enterobacteriaceae bacterium 89	Proteobacteria	82	6	Cap4	0			
206	WP_095123820.1		Klebsiella pneumoniae	Proteobacteria	83	4	Cap4	1			1655
207	WP_164474406.1		Enterobacter hormaechei	Proteobacteria	83	4	Cap4	0			
208	WP_200035944.1		Citrobacter sedlakii	Proteobacteria	83	4	Cap4	0			
209	WP_205920381.1		Enterobacter hormaechei	Proteobacteria	83	4	Cap4	0			
210	WP_059743746.1		Shewanella frigidimarina	Proteobacteria	84	1	Cap4	1			1655
211	WP_165491340.1		Pseudoalteromonas carrageenovora	Proteobacteria	85	4	Cap4	1			1655
212	SOU43283.1		Pseudoalteromonas carrageenovora IAM 12662	Proteobacteria	85	4	Cap4	0			
213	WP_182748923.1		Pseudoalteromonas sp. SR41-1	Proteobacteria	85	4	Cap4	0			
214	WP_158657451.1		Pseudoalteromonas carrageenovora	Proteobacteria	85	4	Cap4	0			
215	WP_065591722.1		unclassified Aliivibrio	Proteobacteria	86	1	Cap4	1			1655
216	WP_196584888.1		Aliivibrio fischeri	Proteobacteria	87	2	Cap4	1			1655
217	MUK65110.1		Aliivibrio fischeri	Proteobacteria	87	2	Cap4	0			
218	MBP6467958.1		Fusobacteriaceae bacterium	Fusobacteria	88	1	Cap4	1			1666
219	PZP15067.1		Aliarcobacter butzleri	Proteobacteria	89	1	Cap4	1			1665
220	WP_152056673.1		Aliarcobacter butzleri	Proteobacteria	90	1	Cap4	1			1673
221	WP_099333681.1		Malaciobacter canalis	Proteobacteria	91	1	Cap4	1			1669
222	WP_076085314.1		Poseidonibacter parvus	Proteobacteria	92	1	Cap4	1			1668
223	MBU1369472.1		Bacteroidetes bacterium	Bacteroidetes	93	1	Cap4	1			1675
224	WP_024016747.1		Alteromonas mediterranea	Proteobacteria	94	1	Cap4	1			1677
225	WP_108699709.1		Nitrosomonas sp. Nm141	Proteobacteria	95	1	Cap4	1			1667
226	WP_009727411.1		Methylophaga lonarensis	Proteobacteria	96	1	Cap4	1			1669
227	WP_124012527.1		unclassified Shewanella	Proteobacteria	97	1	Cap4	1			1669
228	WP_202720845.1		Shewanella schlegeliana	Proteobacteria	98	1	Cap4	1			1668
229	WP_036253828.1		Methylobacter sp. BBA5.1	Proteobacteria	99	1	Cap4	1			1671
230	PCJ26211.1		SAR86 cluster bacterium	Proteobacteria	100	1	Cap4	1			1667
231	WP_079803454.1		Salmonella enterica	Proteobacteria	101	10	Cap4	1			1669
232	WP_197751539.1		Serratia ureilytica	Proteobacteria	101	10	Cap4	0			
233	WP_197792407.1		Serratia ureilytica	Proteobacteria	101	10	Cap4	0			
234	WP_201884797.1		Serratia ureilytica	Proteobacteria	101	10	Cap4	0			

235	EGL9225644.1		Salmonella enterica	Proteobacteria	101	10	Cap4	0			
236	WP_072209110.1		Enterobacterales	Proteobacteria	101	10	Cap4	0			
237	HAU5975501.1		Serratia marcescens	Proteobacteria	101	10	Cap4	0			
238	WP_200145124.1		Klebsiella aerogenes	Proteobacteria	101	10	Cap4	0			
239	WP_175067630.1		Salmonella sp. SG126	Proteobacteria	101	10	Cap4	0			
240	HAX9723969.1		Serratia marcescens	Proteobacteria	101	10	Cap4	0			
241	WP_157360376.1		Aliagarivorans marinus	Proteobacteria	102	1	Cap4	1			1675
242	WP_128452452.1		Escherichia coli	Proteobacteria	103	25	Cap4	1			1669
243	WP_198080631.1		Klebsiella pneumoniae	Proteobacteria	103	25	Cap4	0			
244	WP_200557355.1		Klebsiella pneumoniae	Proteobacteria	103	25	Cap4	0			
245	WP_164943452.1		Klebsiella	Proteobacteria	103	25	Cap4	0			
246	WP_117049154.1		Klebsiella pneumoniae	Proteobacteria	103	25	Cap4	0			
247	WP_211652685.1		Klebsiella pneumoniae	Proteobacteria	103	25	Cap4	0			
248	WP_089578328.1		Escherichia coli	Proteobacteria	103	25	Cap4	0			
249	MBJ9284378.1		Citrobacter freundii	Proteobacteria	103	25	Cap4	0			
250	HAW4047280.1		Escherichia coli	Proteobacteria	103	25	Cap4	0			
251	STF06378.1		Escherichia coli	Proteobacteria	103	25	Cap4	0			
252	HAW0910320.1		Escherichia coli	Proteobacteria	103	25	Cap4	0			
253	EHG5997258.1		Escherichia fergusonii	Proteobacteria	103	25	Cap4	0			
254	WP_112959226.1		Escherichia coli	Proteobacteria	103	25	Cap4	0			
255	WP_187797463.1		Escherichia coli	Proteobacteria	103	25	Cap4	0			
256	EFN4982839.1		Escherichia coli	Proteobacteria	103	25	Cap4	0			
257	EAT5156679.1		Salmonella enterica	Proteobacteria	103	25	Cap4	0			
258	HAV8949656.1		Escherichia coli	Proteobacteria	103	25	Cap4	0			
259	WP_101806208.1		Escherichia coli	Proteobacteria	103	25	Cap4	0			
260	WP_116155437.1		Pectobacterium aquaticum	Proteobacteria	103	25	Cap4	0			
261	WP_166864003.1		Enterobacter sp. Cy-643	Proteobacteria	103	25	Cap4	0			
262	WP_201482922.1		Escherichia coli	Proteobacteria	103	25	Cap4	0			
263	WP_215261980.1		Escherichia coli	Proteobacteria	103	25	Cap4	0			
264	WP_070596068.1		Klebsiella pneumoniae	Proteobacteria	103	25	Cap4	0			
265	EFC6276956.1		Escherichia coli	Proteobacteria	103	25	Cap4	0			
266	WP_195807605.1		Klebsiella pneumoniae	Proteobacteria	103	25	Cap4	0			
267	WP_071704155.1		Yersinia ruckeri	Proteobacteria	104	1	Cap4	1			1670
268	WP_049582950.1		Photorhabdus luminescens	Proteobacteria	105	2	Cap4	1			1669
269	WP_112877439.1		unclassified Photorhabdus	Proteobacteria	105	2	Cap4	0			

270	TNH05577.1		Pasteurellaceae bacterium Phil31	Proteobacteria	106	1	Cap4	1			1669
271	WP_166267319.1		Providencia sp. M-27	Proteobacteria	107	1	Cap4	1			1669
272	WP_161156133.1		Vibrio sp. CAIM 722	Proteobacteria	108	1	Cap4	1			1669
273	CAD6365970.1		Shewanella hafniensis	Proteobacteria	109	1	Cap4	1			1671
274	WP_213898093.1		Vibrio alginolyticus	Proteobacteria	110	2	Cap4	1			1671
275	WP_104028830.1		Vibrio jasicida	Proteobacteria	110	2	Cap4	0			
276	WP_142626215.1		Vibrio cholerae	Proteobacteria	111	5	Cap4	1			1669
277	WP_164561230.1		Providencia rettgeri	Proteobacteria	111	5	Cap4	0			
278	WP_050498135.1		Gammaproteobacteria	Proteobacteria	111	5	Cap4	0			
279	WP_096865192.1		Providencia rettgeri	Proteobacteria	111	5	Cap4	0			
280	EGR1125489.1		Vibrio cholerae	Proteobacteria	111	5	Cap4	0			
281	WP_071815637.1		Alteromonas sp. V450	Proteobacteria	112	1	Cap4	1			1670
282	WP_144374815.1		Shewanella xiamenensis	Proteobacteria	113	1	Cap4	1			1668
283	WP_191867214.1		Paraglaciecola oceanifecundans	Proteobacteria	114	5	Cap4	1			1671
284	WP_108716793.1		Proteus mirabilis	Proteobacteria	114	5	Cap4	0			
285	MBG2851235.1		Proteus mirabilis	Proteobacteria	114	5	Cap4	0			
286	MBE4271102.1		Vibrio parahaemolyticus	Proteobacteria	114	5	Cap4	0			
287	WP_125891661.1		Providencia rettgeri	Proteobacteria	114	5	Cap4	0			
288	WP_099380019.1		Janthinobacterium sp. BJB426	Proteobacteria	115	1	Cap4	1			1676
289	WP_082591364.1		Duganella sp. Root198D2	Proteobacteria	116	2	Cap4	1			1672
290	KRB97285.1		Duganella sp. Root198D2	Proteobacteria	116	2	Cap4	0			
291	OUR59858.1		Colwellia sp. 39_35_sub15_T18	Proteobacteria	117	1	Cap4	1			1659
292	WP_101164442.1		Pseudoalteromonas sp. 78C3	Proteobacteria	118	1	Cap4	1			1654
293	WP_143026624.1		Ferrimonas sediminum	Proteobacteria	119	1	Cap4	1			1693
294	WP_210455656.1		Vibrio crassostreae	Proteobacteria	120	1	Cap4	1			1695
295	WP_161673629.1		Vibrio toranzoniae	Proteobacteria	121	1	Cap4	1			1695
296	WP_065597301.1		Aliivibrio fischeri	Proteobacteria	122	1	Cap4	1			1696
297	WP_213899712.1		Vibrio alginolyticus	Proteobacteria	123	5	Cap4	1			1695
298	EGQ8154626.1		Vibrio alginolyticus	Proteobacteria	123	5	Cap4	0			
299	WP_100267553.1		Vibrio fluvialis	Proteobacteria	123	5	Cap4	0			
300	WP_032080178.1		Vibrio fluvialis	Proteobacteria	123	5	Cap4	0			
301	WP_206980880.1		Vibrio sp. Vb2880	Proteobacteria	123	5	Cap4	0			
302	WP_050914027.1		Vibrio harveyi	Proteobacteria	124	2	Cap4	1			1695

303	WP_047879560.1		Photobacterium aquae	Proteobacteria	124	2	Cap4	0			
304	WP_103193784.1		Vibrio vulnificus	Proteobacteria	125	1	Cap4	1			1695
305	WP_065332285.1		Epibacterium mobile	Proteobacteria	126	1	Cap4	1			1647
306	WP_156127892.1		Pseudooceanicola atlanticus	Proteobacteria	127	1	Cap4	1			1647
307	WP_063305201.1		Pseudovibrio sp. W64	Proteobacteria	128	1	Cap4	1			1640
308	WP_093523993.1		Pseudovibrio ascidiaceicola	Proteobacteria	129	1	Cap4	1			1640
309	WP_169352948.1		Pseudomonas sp. ADAK18	Proteobacteria	130	1	Cap4	1			1650
310	WP_110816619.1		Psychrobacter sp. YP14	Proteobacteria	131	1	Cap4	1			1653
311	WP_102083150.1		Psychrobacter sp. Choline-02u-9	Proteobacteria	132	1	Cap4	1			1653
312	EDY81180.1		Verrucomicrobiae bacterium DG1235	Verrucomicrobia	133	1	Cap4	1			1649
313	WP_197893697.1		Pseudomonas monteilii	Proteobacteria	134	1	Cap4	1			1651
314	WP_203304029.1		Pseudomonas weihenstephanensis	Proteobacteria	135	1	Cap4	1			1651
315	MBD368542.1		Kangiella sp.	Proteobacteria	136	1	Cap4	1			1656
316	WP_139093589.1		Vibrio	Proteobacteria	137	3	Cap4	1			1654
317	WP_140166007.1		Vibrio parahaemolyticus	Proteobacteria	137	3	Cap4	0			
318	TOL35066.1		Vibrio parahaemolyticus	Proteobacteria	137	3	Cap4	0			
319	WP_080156448.1		Photobacterium piscicola	Proteobacteria	138	1	Cap4	1			1653
320	WP_045067429.1		Photobacterium kishitanii	Proteobacteria	139	1	Cap4	1			1654
321	WP_193291430.1		Vibrio parahaemolyticus	Proteobacteria	140	5	Cap4	1			1655
322	WP_152118590.1		Proteus mirabilis	Proteobacteria	140	5	Cap4	0			
323	WP_170355903.1		Pasteurella multocida	Proteobacteria	140	5	Cap4	0			
324	AVR65437.1		Shewanella algae	Proteobacteria	140	5	Cap4	0			
325	WP_193304452.1		Vibrio parahaemolyticus	Proteobacteria	140	5	Cap4	0			
326	SDJ57962.1		Halomonas gudaonensis	Proteobacteria	141	2	Cap4	1			1655
327	WP_143005088.1		Halomonas gudaonensis	Proteobacteria	141	2	Cap4	0			
328	WP_101837294.1		Klebsiella variicola	Proteobacteria	142	1	Cap4	1			1655
329	SNY72170.1		Pantoea sp. GL120224-02	Proteobacteria	143	2	Cap4	1			1655
330	WP_141399028.1		Pantoea sp. GL120224-02	Proteobacteria	143	2	Cap4	0			
331	WP_123828920.1		Klebsiella	Proteobacteria	144	5	Cap4	1			1655
332	WP_063411438.1		Enterobacter cloacae	Proteobacteria	144	5	Cap4	0			
333	WP_148769445.1		Enterobacter sp. C6	Proteobacteria	144	5	Cap4	0			
334	WP_181485256.1		unclassified Klebsiella	Proteobacteria	144	5	Cap4	0			

335	EFL31090107.1		Escherichia coli	Proteobacteria	144	5	Cap4	0			
336	WP_167324154.1		Halomonas zhanjiangensis	Proteobacteria	145	1	Cap4	1			1655
337	WP_147269638.1		Pseudidiomarina maritima	Proteobacteria	146	4	Cap4	1			1655
338	WP_034776900.1		Pseudidiomarina salinarum	Proteobacteria	146	4	Cap4	0			
339	WP_110076096.1		Pseudidiomarina tainanensis	Proteobacteria	146	4	Cap4	0			
340	WP_146778636.1		Pseudidiomarina tainanensis	Proteobacteria	146	4	Cap4	0			
341	MAP255517.1		Methylophaga sp.	Proteobacteria	147	1	Cap4	1			1655
342	WP_211595492.1		Halomonas boliviensis	Proteobacteria	148	1	Cap4	1			1655
343	KAE8438454.1		Halomonas piezotolerans	Proteobacteria	149	2	Cap4	1			1655
344	WP_168444419.1		Halomonas piezotolerans	Proteobacteria	149	2	Cap4	0			
345	WP_211877150.1		Roseomonas hellenica	Proteobacteria	150	1	Cap4_split1	1	WP_211877149.1	483	1691
346	WP_198370146.1		Roseomonas sp. OP-27	Proteobacteria	151	1	Cap4_split1	1	WP_198370145.1*	481	1688
347	ABM57702.1		Verminephrobacter eiseniae EF01-2	Proteobacteria	152	2	Cap4_split1	1	ABM57703.1	488	1699
348	WP_157048452.1		Verminephrobacter eiseniae	Proteobacteria	152	2	Cap4_split1	0			
349	WP_209605364.1		Herbaspirillum sp. LeCh32-8	Proteobacteria	153	1	Cap4_split1	1	WP_209605362.1	477	1682
350	WP_120311019.1		Caballeronia udeis	Proteobacteria	154	1	Cap4_split1	1	WP_120311020.1	482	1686
351	ASL481110.1		Burkholderia sp. AD24	Proteobacteria	155	1	Cap4_split1	1	ASL48111.1	482	1686
352	WP_214571231.1		unclassified Dyella	Proteobacteria	156	1	Cap4_split1	1	WP_214571232.1*	486	1687
353	WP_122793411.1		Burkholderia	Proteobacteria	157	1	Cap4_split1	1	WP_147485751.1	488	1694
354	MBL8470762.1		Rhodocyclaceae bacterium	Proteobacteria	158	1	Cap4_split1	1	MBL8470761.1	483	1686
355	MBC7160932.1		Immundisolibacter sp.	Proteobacteria	159	1	Cap4_split1	1	MBC7160931.1	489	1691
356	MBK7532300.1		Piscinibacter sp.	Proteobacteria	160	1	Cap4_split1	1	MBK7532301.1*	493	1695
357	WP_057405870.1		Pseudomonas syringae	Proteobacteria	161	5	Cap4_split1	1	WP_057405871.1	477	1684
358	WP_122436418.1		Pseudomonas viridiflava	Proteobacteria	161	5	Cap4_split1	0			
359	KPX66221.1		Pseudomonas syringae pv. lapsa	Proteobacteria	161	5	Cap4_split1	0			
360	WP_213630441.1		Pseudomonas syringae	Proteobacteria	161	5	Cap4_split1	0			
361	WP_084307412.1		Pseudomonas flavescens	Proteobacteria	161	5	Cap4_split1	0	WP_139207472.1	474	
362	WP_198799704.1		Pseudomonas sp. MF6787	Proteobacteria	162	1	Cap4_split1	1	WP_198799705.1	475	1682
363	WP_209966054.1		unclassified Pseudomonas	Proteobacteria	163	1	Cap4_split1	1	WP_209966056.1	481	1679
364	WP_126570276.1		Pseudomonas sp. C 49-2	Proteobacteria	164	1	Cap4_split1	1	WP_126570277.1	482	1694
365	OJW68038.1		Aeromonas sp. 62-46	Proteobacteria	165	1	Cap4_split1	1	--	476	1678
366	WP_194759458.1		Aeromonas hydrophila	Proteobacteria	166	1	Cap4_split1	1	WP_194759457.1	476	1678
367	WP_096253271.1		Pseudomonas aeruginosa	Proteobacteria	167	12	Cap4_split1	1	WP_124119684.1	476	1677
368	WP_033959272.1		Pseudomonas aeruginosa	Proteobacteria	167	12	Cap4_split1	0			

369	WP_181016898.1		Pseudomonas aeruginosa	Proteobacteria	167	12	Cap4_split1	0			
370	SUD28664.1		Pseudomonas mendocina	Proteobacteria	167	12	Cap4_split1	0			
371	WP_161791170.1		Pseudomonas aeruginosa	Proteobacteria	167	12	Cap4_split1	0			
372	WP_181880615.1		Pseudomonas mendocina	Proteobacteria	167	12	Cap4_split1	0			
373	KSF69295.1		Pseudomonas aeruginosa	Proteobacteria	167	12	Cap4_split1	0			
374	WP_160326970.1		Pseudomonas aeruginosa	Proteobacteria	167	12	Cap4_split1	0			
375	MBH8746740.1		Pseudomonas aeruginosa	Proteobacteria	167	12	Cap4_split1	0			
376	WP_109223942.1		Pseudomonas aeruginosa	Proteobacteria	167	12	Cap4_split1	0			
377	WP_162597750.1		Pseudomonas aeruginosa	Proteobacteria	167	12	Cap4_split1	0			
378	WP_194932708.1		Pseudomonas aeruginosa	Proteobacteria	167	12	Cap4_split1	0			
379	WP_186949339.1		Undibacterium sp. CY18W	Proteobacteria	168	1	Cap4_split1	1	WP_186949340.1	481	1682
380	MBD8723055.1		Oxalobacteraceae sp. CFBP 13708	Proteobacteria	169	1	Cap4_split1	1	MBD8723054.1*	476	1676
381	WP_199578885.1		Bosea vaviloviae	Proteobacteria	170	2	Cap4_split1	1	WP_158516217.1	466	1671
382	AOO815633.1		Bosea vaviloviae	Proteobacteria	170	2	Cap4_split1	0			
383	MYD45833.1		Gammaproteobacteria bacterium	Proteobacteria	171	1	Cap4_split1	1	MYD45834.1	481	1676
384	WP_200167881.1		Brevundimonas diminuta	Proteobacteria	172	1	Cap4_split1	1	WP_200167880.1	483	1687
385	WP_021697216.1		Brevundimonas	Proteobacteria	173	1	Cap4_split1	1	MBG7613713.1*	465	1664
386	WP_100197588.1		Pseudorhodobacter sp. MZDSW-24AT	Proteobacteria	174	1	Cap4_split1	1	WP_157792023.1	465	1665
387	PWR03370.1		Rhodobacteraceae bacterium TG-679	Proteobacteria	175	1	Cap4_split1	1	--	465	1665
388	WP_076484770.1		Rhodobacter aestuarii	Proteobacteria	176	1	Cap4_split1	1	WP_076484769.1	465	1665
389	WP_193445563.1		Rhizobium indigoferae	Proteobacteria	177	1	Cap4_split1	1	WP_193445564.1	481	1688
390	WP_200299685.1		Bradyrhizobium diazoefficiens	Proteobacteria	178	1	Cap4_split1	1	WP_200299686.1	487	1695
391	WP_193369333.1		Pelagibius sp. NBU2595	Proteobacteria	179	1	Cap4_split1	1	WP_193369332.1*	485	1690
392	MYB27409.1		Candidatus Dadabacteria bacterium	Unclassified	180	1	Cap4_split1	1	MYB27410.1	486	1697
393	MBU2678194.1		Gammaproteobacteria bacterium	Proteobacteria	181	1	Cap4_split1	1	MBU2678193.1*	485	1691
394	WP_023712771.1		unclassified Mesorhizobium	Proteobacteria	182	1	Cap4_split1	1	ESX47353.1	490	1685
395	WP_062587300.1		Rhizobium sp. Leaf311	Proteobacteria	183	1	Cap4_split1	1	WP_156462039.1	485	1681
396	WP_181293582.1		Rhizobium sp. WYCCWR 11146	Proteobacteria	184	1	Cap4_split1	1	WP_181293581.1	485	1681
397	WP_077067114.1		Pseudorhizobium marinum	Proteobacteria	185	1	Cap4_split1	1	WP_156034888.1	478	1681
398	WP_192182868.1		Mesorhizobium amorphae	Proteobacteria	186	1	Cap4_split1	1	WP_192182869.1	475	1678
399	WP_018096623.1		Sinorhizobium meliloti	Proteobacteria	187	2	Cap4_split1	1	WP_152513689.1	479	1683

400	WP_128203966.1		Sinorhizobium meliloti	Proteobacteria	187	2	Cap4_split1	0			
401	WP_168290154.1		Rhizobium leguminosarum	Proteobacteria	188	3	Cap4_split1	1	WP_168290156.1	481	1684
402	NKM97974.1		Rhizobium leguminosarum bv. viciae	Proteobacteria	188	3	Cap4_split1	0			
403	WP_133866992.1		Azorhizobium sp. AG788	Proteobacteria	188	3	Cap4_split1	0			
404	WP_140630235.1		unclassified Mesorhizobium	Proteobacteria	189	1	Cap4_split1	1	WP_140630237.1	489	1689
405	WP_110787752.1		Rhodopseudomonas palustris	Proteobacteria	190	2	Cap4	1			1690
406	PZA10199.1		Rhodopseudomonas palustris	Proteobacteria	190	2	Cap4	0			
407	WP_192439374.1		Ensifer sp. ENS03	Proteobacteria	191	3	Cap4_split1	1	WP_192439375.1	489	1695
408	OWZ90519.1		Sinorhizobium sp. LM21	Proteobacteria	191	3	Cap4_split1	0			
409	WP_181432630.1		Ensifer adhaerens	Proteobacteria	191	3	Cap4_split1	0			
410	RWK54005.1		Mesorhizobium sp.	Proteobacteria	192	1	Cap4_split1	1	RWK54006.1*	492	1698
411	TIW23419.1		Mesorhizobium sp.	Proteobacteria	193	1	Cap4_split1	1	TIW23420.1*	488	1693
412	OYR26806.1		Brucella lupini	Proteobacteria	194	4	Cap4_split1	1	OYR26805.1	488	1693
413	WP_143851087.1		Brucella/Ochrobactrum group	Proteobacteria	194	4	Cap4_split1	0			
414	RCI78364.1		Brucella anthropi	Proteobacteria	194	4	Cap4_split1	0			
415	WP_147251702.1		Brucella anthropi	Proteobacteria	194	4	Cap4_split1	0			
416	QWP75899.1		Lysobacter sp. K5869	Proteobacteria	195	1	Cap4	1			1681
417	WP_094001825.1		Stenotrophomonas sp. CC120222-04	Proteobacteria	196	1	Cap4	1			1688
418	MBE0574374.1		Desulfuromonadales bacterium	Proteobacteria	197	1	Cap4_split1	1	MBE0574373.1	486	1690
419	MBI3246072.1		Deltaproteobacteria bacterium	Proteobacteria	198	1	Cap4_split1	1	MBI3246073.1*	485	1690
420	WP_191948002.1		Pseudomonas fluorescens	Proteobacteria	199	1	Cap4_split1	1	WP_191948003.1	484	1690
421	WP_159813976.1		Pseudomonas sp. 18058	Proteobacteria	200	1	Cap4_split1	1	WP_159813978.1	484	1691
422	KUK47893.1		Actinobacteria bacterium 66_15	Actinobacteria	201	2	Cap4	1			1728
423	HAL29492.1		Coriobacteriia bacterium	Actinobacteria	201	2	Cap4	0			
424	WP_141258264.1		Myxococcus	Proteobacteria	202	2	Cap4	1			1733
425	WP_163973350.1		Myxococcus sp. CA018	Proteobacteria	202	2	Cap4	0			
426	HIF26250.1		Micavibrio sp.	Proteobacteria	203	1	Cap4	1			1694
427	MYB42507.1		Chloroflexi bacterium	Chloroflexi	204	1	Cap4_split1	1	MYB42508.1	486	1699
428	WP_138084940.1		Phragmitibacter flavus	Verrucomicrobia	205	1	Cap4	1			1741
429	MXZ72702.1		Acidobacteria bacterium	Acidobacteria	206	1	Cap4_split1	1	MXZ72703.1*	513	1744
430	WP_201365809.1		Dictyobacter sp. SOSP1-9	Chloroflexi	207	2	Cap4_split1	1	WP_201365810.1	507	1744
431	GHO88201.1		Dictyobacter sp. SOSP1-9	Chloroflexi	207	2	Cap4_split1	0			
432	WP_094015648.1		Paenibacillus rigui	Firmicutes	208	1	Cap4_split1	1	WP_094015647.1	483	1690

433	PKO04447.1		Chloroflexi bacterium HGW-Chloroflexi-3	Chloroflexi	209	1	Cap4_split1	1	PKO04448.1*	519	1750
434	TDA67676.1		Chloroflexi bacterium	Chloroflexi	210	1	Cap4_split1	1	TDA67677.1*	519	1750
435	WP_015333064.1		Fibrella aestuarina	Bacteroidetes	211	1	Cap4	1			1724
436	OFZ70213.1		Bdellovibrionales bacterium RIFOXYD1_FULL_44_7	Proteobacteria	212	1	Cap4	1			1704
437	QQR58261.1		Candidatus Melainabacteria bacterium	Candidatus Melainabacteria	213	1	Cap4	1			1703
438	PIP84473.1		Elusimicrobia bacterium CG22_combo_CG10-13_8_21_14_all_63_91	Elusimicrobia	214	1	Cap4	1			1714
439	PKM98254.1		Elusimicrobia bacterium HGW-Elusimicrobia-3	Elusimicrobia	215	1	Cap4	1			1713
440	WP_094397539.1		Thermoanaerobacterium thermosaccharolyticum	Firmicutes	216	3	Cap4	1			1722
441	WP_209455027.1		Thermoanaerobacterium butyriciformans	Firmicutes	216	3	Cap4	0			
442	WP_014063424.1		Thermoanaerobacter wiegelii	Firmicutes	216	3	Cap4	0			
443	WP_046940816.1		Bacillus cereus	Firmicutes	217	2	Cap4	1			1715
444	EDZ56979.1		Bacillus cereus H3081.97	Firmicutes	217	2	Cap4	0			
445	MBI2484824.1		Candidatus Uhrbacteria bacterium	Unclassified	218	1	Cap4	1			1717
446	OGF27145.1		Candidatus Falkowbacteria bacterium RIFOXYB2_FULL_34_18	Unclassified	219	1	Cap4	1			1705
447	KKR93645.1		Candidatus Roizmanbacteria bacterium GW2011_GWA1_41_13	Unclassified	220	1	Cap4	1			1702
448	NLT24896.1		Syntrophorhabdus sp.	Proteobacteria	221	1	Cap4	1			1709
449	WP_117389706.1		Mucilaginibacter sp. MYSH2	Bacteroidetes	222	1	Cap4	1			1707
450	WP_158798018.1		Pedobacter sp. L105	Bacteroidetes	223	1	Cap4	1			1707
451	WP_175635986.1		Pedobacter sp. Marseille-Q2390	Bacteroidetes	224	1	Cap4	1			1711
452	WP_184492448.1		Algoriphagus iocasae	Bacteroidetes	225	1	Cap4	1			1707
453	WP_165446374.1		Chryseobacterium indologenes	Bacteroidetes	226	2	Cap4	1			1714
454	VFA41361.1		Chryseobacterium indologenes	Bacteroidetes	226	2	Cap4	0			
455	WP_123881555.1		Chryseobacterium carnipullorum	Bacteroidetes	227	2	Cap4	1			1714

456	AZA50683.1		Chryseobacterium carnipullorum	Bacteroidetes	227	2	Cap4	0			
457	RYZ55990.1		Sphingobacteriales bacterium	Bacteroidetes	228	1	Cap4	1			1702
458	WP_112375741.1		Sphingobacterium multivorum	Bacteroidetes	229	1	Cap4	1			1705
459	WP_158638970.1		Sphingobacterium siyangense	Bacteroidetes	230	2	Cap4	1			1705
460	TWI15509.1		Sphingobacterium siyangense	Bacteroidetes	230	2	Cap4	0			
461	WP_104384141.1		Sphingobacterium sp. HMA12	Bacteroidetes	231	1	Cap4	1			1705
462	WP_048501883.1		Chryseobacterium sp. BLS98	Bacteroidetes	232	3	Cap4	1			1704
463	WP_202089732.1		Chryseobacterium sp. L7	Bacteroidetes	232	3	Cap4	0			
464	WP_123318684.1		Chryseobacterium nakagawai	Bacteroidetes	232	3	Cap4	0			
465	WP_121966561.1		Myroides sp. N17-2	Bacteroidetes	233	1	Cap4	1			1697
466	WP_073335416.1		Chryseobacterium carnipullorum	Bacteroidetes	234	2	Cap4	1			1702
467	WP_068941783.1		Chryseobacterium timonianum	Bacteroidetes	234	2	Cap4	0			
468	WP_124802160.1		Epilithonimonas vandammei	Bacteroidetes	235	1	Cap4	1			1697
469	WP_124757888.1		Kaistella daneshvariae	Bacteroidetes	236	1	Cap4	1			1693
470	WP_188618526.1		Cloacibacterium rupense	Bacteroidetes	237	1	Cap4	1			1693
471	WP_087711271.1		Chryseobacterium mucoviscidosis	Bacteroidetes	238	1	Cap4	1			1697
472	WP_099750132.1		Deinococcus sp. UR1	Deinococcus-Thermus	239	1	Cap4	1			1691
473	PHR55999.1		Arcobacter sp.	Proteobacteria	240	1	Cap4_split3	1	PHR55998.1	283	1600
474	WP_109076644.1		Aliarcobacter skirrowii	Proteobacteria	241	1	Cap4_split3	1	WP_109066718.1	295	1610
475	NQY22865.1		Campylobacteraceae bacterium	Proteobacteria	242	1	Cap4_split3	1	NQY22866.1	285	1621
476	HIP12364.1		Arcobacter sp.	Proteobacteria	243	1	Cap4_split3	1	HIP12365.1	290	1597
477	WP_110454257.1		Gilliamella apicola	Proteobacteria	244	1	Cap4_split3	1	WP_110454258.1	286	1582
478	WP_195186899.1		Pasteurella multocida	Proteobacteria	245	2	Cap4_split3	1	WP_151249361.1	293	1580
479	WP_195188799.1		Pasteurella multocida	Proteobacteria	245	2	Cap4_split3	0			
480	WP_035689165.1		Avibacterium paragallinarum	Proteobacteria	246	3	Cap4_split3	1	WP_051128134.1	279	1567
481	WP_039170362.1		Gallibacterium anatis	Proteobacteria	246	3	Cap4_split3	0			
482	WP_065232651.1		Gallibacterium anatis	Proteobacteria	246	3	Cap4_split3	0			
483	WP_188156072.1		Mannheimia bovis	Proteobacteria	247	1	Cap4_split3	1	WP_188156071.1	279	1573
484	WP_194621199.1		Psychrobacter sp. N25K4-3-2	Proteobacteria	248	1	Cap4_split3	1	WP_194621198.1	287	1588
485	WP_123774338.1		Acinetobacter sp. BIGb0102	Proteobacteria	249	1	Cap4_split3	1	WP_123774339.1	288	1586
486	WP_181845096.1		Pectobacterium aroidearum	Proteobacteria	250	1	Cap4_split3	1	WP_181845097.1	289	1589
487	EAX37258841.1		Salmonella enterica	Proteobacteria	251	4	Cap4_split3	1	EAX3725840.1	286	1586

488	SUG26152.1		Salmonella enterica subsp. arizonae	Proteobacteria	251	4	Cap4_split3	0			
489	WP_045308351.1		Enterobacter hormaechei	Proteobacteria	251	4	Cap4_split3	0			
490	EEI0564516.1		Salmonella enterica	Proteobacteria	251	4	Cap4_split3	0			
491	WP_124967285.1		Kosakonia sp. CCTCC M2018092	Proteobacteria	252	1	Cap4_split3	1	WP_124967286.1	286	1584
492	WP_211750866.1		Klebsiella variicola	Proteobacteria	253	1	Cap4_split3	1	WP_211750865.1	286	1584
493	WP_072055721.1		Aliivibrio fischeri	Proteobacteria	254	1	Cap4_split3	1	WP_072055722.1	287	1584
494	WP_065609470.1		Aliivibrio logei	Proteobacteria	255	1	Cap4_split3	1	WP_065609472.1	287	1586
495	WP_182259590.1		Colwellia sp. BRX8-4	Proteobacteria	256	1	Cap4_split3	1	WP_182259591.1	287	1587
496	WP_064435980.1		Pseudoalteromonas neustonica	Proteobacteria	257	1	Cap4_split3	1	WP_064435981.1	287	1586
497	WP_201288012.1		Psychromonas sp. L1A2	Proteobacteria	258	1	Cap4_split3	1	WP_160060759.1	287	1586
498	WP_143696632.1		Vibrio vulnificus	Proteobacteria	259	3	Cap4_split3	1	WP_080532675.1	287	1584
499	WP_130252539.1		Vibrio vulnificus	Proteobacteria	259	3	Cap4_split3	0			
500	HAS8489727.1		Vibrio vulnificus	Proteobacteria	259	3	Cap4_split3	0			
501	HAV1325330.1		Vibrio parahaemolyticus	Proteobacteria	260	1	Cap4_split3	1	HAV1325329.1	289	1581
502	PTP08174.1		Vibrio splendidus	Proteobacteria	261	2	Cap4_split3	1	PTP08175.1	289	1579
503	WP_108175602.1		Vibrio splendidus	Proteobacteria	261	2	Cap4_split3	0			
504	WP_000056155.1		Vibrio	Proteobacteria	262	5	Cap4_split3	1	WP_001087594.1	287	1583
505	WP_105398697.1		Photorhabdus luminescens	Proteobacteria	262	5	Cap4_split3	0			
506	EKG91523.1		Vibrio cholerae HE-16	Proteobacteria	262	5	Cap4_split3	0			
507	WP_109717844.1		Pantoea allii	Proteobacteria	262	5	Cap4_split3	0			
508	PWK95325.1		Pantoea allii	Proteobacteria	262	5	Cap4_split3	0			
509	WP_063367277.1		Pseudoalteromonas luteoviolacea	Proteobacteria	263	4	Cap4_split3	1	WP_063367278.1	287	1584
510	OEF30231.1		Vibrio rumoiensis 1S-45	Proteobacteria	263	4	Cap4_split3	0			
511	WP_141705947.1		Vibrio rumoiensis	Proteobacteria	263	4	Cap4_split3	0			
512	MAD04285.1		Pseudoalteromonas sp.	Proteobacteria	263	4	Cap4_split3	0	MAD04284.1	287	
513	WP_055018160.1		Pseudoalteromonas sp. P1-25	Proteobacteria	264	1	Cap4_split3	1	WP_055018159.1	287	1584
514	WP_181718163.1		Pseudoalteromonas	Proteobacteria	265	2	Cap4_split3	1	WP_181718162.1	287	1584
515	QID24536.1		Pseudoalteromonas translucida TAC125	Proteobacteria	265	2	Cap4_split3	0			
516	MBO4712479.1		Fibrobacter sp.	Fibrobacteres	266	1	Cap4	1			1559
517	MBR4006569.1		Fibrobacter sp.	Fibrobacteres	267	1	Cap4	1			1563
518	WP_072827710.1		Fibrobacter sp. UWB12	Fibrobacteres	268	1	Cap4	1			1564
519	MBR4789688.1		Treponema sp.	Spirochaetes	269	1	Cap4	1			1568

520	MBP3742546.1		Treponema sp.	Spirochaetes	270	1	Cap4	1			1569
521	WP_150093345.1		Adhaeribacter rhizoryzae	Bacteroidetes	271	1	Cap4	1			1588
522	EKD26246.1		uncultured bacterium	Unclassified	272	1	Cap4	1			1584
523	WP_076373822.1		Mucilaginibacter lappiensis	Bacteroidetes	273	1	Cap4	1			1603
524	WP_012457616.1		Porphyromonas gingivalis	Bacteroidetes	274	2	Cap4	1			1602
525	WP_077110668.1		Porphyromonas gingivalis	Bacteroidetes	274	2	Cap4	0			
526	WP_146131399.1		Mongoliibacter ruber	Bacteroidetes	275	1	Cap4	1			1604
527	WP_034891622.1		Gillisia sp. Hel_I_29	Bacteroidetes	276	1	Cap4	1			1606
528	MBL7767504.1		Flavipsychrobacter sp.	Bacteroidetes	277	1	Cap4	1			1612
529	OHC84250.1		Sphingobacteriia bacterium RIFOXYC2_FULL_35_18	Bacteroidetes	278	2	Cap4	1			1607
530	MBA4259371.1		Chitinophaga sp.	Bacteroidetes	278	2	Cap4	0			
531	WP_024981731.1		Flavobacterium succinicans	Bacteroidetes	279	1	Cap4	1			1605
532	WP_101471535.1		Flavobacterium lindanitolerans	Bacteroidetes	280	1	Cap4	1			1607
533	WP_011586309.1		Cytophaga hutchinsonii	Bacteroidetes	281	1	Cap4	1			1612
534	WP_155084577.1		Cyanobacterium aponinum	Cyanobacteria	282	1	Cap4	1			1607
535	WP_029274551.1		Pedobacter borealis	Bacteroidetes	283	1	Cap4	1			1614
536	WP_208232817.1		Gelidibacter sp. DF109	Bacteroidetes	284	1	Cap4	1			1611
537	WP_151147743.1		Candidatus Galacturonibacter soehngenii	Firmicutes	285	1	Cap4	1			1661
538	WP_202769080.1		Clostridium sp. YIM B02565	Firmicutes	286	1	Cap4	1			1662
539	MBE5967941.1		Lachnospiraceae bacterium	Firmicutes	287	1	Cap4	1			1662
540	WP_163951660.1		Paenibacillus sp. SYP-B3998	Firmicutes	288	1	Cap4	1			1656
541	WP_104496793.1		Paenibacillus peoriae	Firmicutes	289	1	Cap4	1			1651
542	WP_166276294.1		Saccharibacillus sp. VR-M41	Firmicutes	290	1	Cap4	1			1647
543	WP_021258677.1		Paenibacillus alvei	Firmicutes	291	1	Cap4	1			1648
544	WP_166246703.1		Bacillus sp. MM2020_4	Firmicutes	292	1	Cap4	1			1653
545	CDB16208.1		Clostridium sp. CAG:221	Firmicutes	293	1	Cap4	1			1651
546	WP_157336201.1		Paenibacillus sp. N10	Firmicutes	294	1	Cap4	1			1652
547	WP_064202770.1		Brevibacillus brevis	Firmicutes	295	1	Cap4	1			1651
548	WP_048367863.1		Bacillus	Firmicutes	296	2	Cap4	1			1653
549	WP_144464710.1		Bacillus velezensis	Firmicutes	296	2	Cap4	0			
550	WP_003485355.1		Clostridium sporogenes	Firmicutes	297	2	Cap4	1			1651
551	WP_085333428.1		Clostridium sporogenes	Firmicutes	297	2	Cap4	0			
552	WP_199670274.1		Bacillus cereus	Firmicutes	298	1	Cap4	1			1623

553	WP_045145142.1		Clostridium butyricum	Firmicutes	299	2	Cap4	1			1624
554	WP_195958207.1		Clostridium butyricum	Firmicutes	299	2	Cap4	0			
555	WP_092560226.1		[Clostridium] populeti	Firmicutes	300	1	Cap4	1			1624
556	CDE27315.1		Catenibacterium sp. CAG:290	Firmicutes	301	1	Cap4	1			1666
557	WP_117640031.1		Blautia obeum	Firmicutes	302	3	Cap4	1			1670
558	WP_173753164.1		Blautia massiliensis	Firmicutes	302	3	Cap4	0			
559	WP_118170693.1		Ruminococcus sp. AM41-10BH	Firmicutes	302	3	Cap4	0			
560	HIT12813.1		Candidatus Scatomonas merdigallinarum	Firmicutes	303	1	Cap4	1			1670
561	WP_172471685.1		[Clostridium] cocleatum	Firmicutes	304	1	Cap4	1			1670
562	CRL41062.1		[Eubacterium] rectale	Firmicutes	305	2	Cap4	1			1670
563	NSI36933.1		[Eubacterium] rectale	Firmicutes	305	2	Cap4	0			
564	WP_008399866.1		Clostridium sp. L2-50	Firmicutes	306	1	Cap4	1			1670
565	MBP3326687.1		Coprococcus sp.	Firmicutes	307	1	Cap4	1			1670
566	WP_118036018.1		Megamonas rupellensis	Firmicutes	308	1	Cap4	1			1668
567	WP_157350905.1		Roseburia intestinalis	Firmicutes	309	4	Cap4	1			1669
568	WP_118210387.1		Roseburia sp. AM23-20	Firmicutes	309	4	Cap4	0			
569	WP_117920625.1		Roseburia sp. OF03-24	Firmicutes	309	4	Cap4	0			
570	WP_118744407.1		Clostridium sp. AM27-31LB	Firmicutes	309	4	Cap4	0			
571	WP_101877123.1		Lachnoclostridium edouardi	Firmicutes	310	1	Cap4	1			1667
572	MBS7032676.1		Clostridium sp.	Firmicutes	311	1	Cap4	1			1672
573	WP_002573793.1		Enterocloster bolteae	Firmicutes	312	2	Cap4	1			1670
574	WP_195638627.1		Enterocloster bolteae	Firmicutes	312	2	Cap4	0			
575	WP_118682909.1		Eisenbergiella sp. OF01-20	Firmicutes	313	1	Cap4	1			1670
576	WP_188616007.1		Lysinibacillus alkalisoli	Firmicutes	314	2	Cap4	1			1682
577	WP_182078375.1		Staphylococcus aureus	Firmicutes	314	2	Cap4	0			
578	WP_062921786.1		Bacillus sp. GZT	Firmicutes	315	1	Cap4	1			1683
579	WP_063671358.1		Priestia megaterium	Firmicutes	316	1	Cap4	1			1689
580	WP_098295918.1		Bacillus cereus	Firmicutes	317	3	Cap4	1			1685
581	WP_098382676.1		Bacillus cereus	Firmicutes	317	3	Cap4	0			
582	WP_137026626.1		Bacillus cereus	Firmicutes	317	3	Cap4	0			
583	WP_151577403.1		Bacillus toyonensis	Firmicutes	318	1	Cap4	1			1681
584	WP_165903300.1		Bacillus sp. OK085	Firmicutes	319	1	Cap4	1			1729
585	MBC7933016.1		Rubrivivax sp.	Proteobacteria	320	1	Cap4	1			1768

586	AVX30043.1		Carboxydocella thermautotrophica	Firmicutes	321	1	Cap4	1			1737
587	HHW40238.1		Syntrophomonadaceae bacterium	Firmicutes	322	3	Cap4	1			1738
588	MQL53963.1		Desulfofundulus thermobenzoicus	Firmicutes	322	3	Cap4	0			
589	WP_194175025.1		Desulfofundulus thermobenzoicus	Firmicutes	322	3	Cap4	0			
590	WP_126390593.1		Pseudanabaena sp. ABRG5-3	Cyanobacteria	323	2	Cap4	1			1735
591	BBC26618.1		Pseudanabaena sp. ABRG5-3	Cyanobacteria	323	2	Cap4	0			
592	WP_193914247.1		Desmonostoc muscorum	Cyanobacteria	324	1	Cap4	1			1743
593	WP_190445555.1		Trichocoleus sp. FACHB-90	Cyanobacteria	325	1	Cap4	1			1747
594	MBL9192879.1		Opitutaceae bacterium	Verrucomicrobia	326	1	Cap4	1			1720
595	WP_102319176.1		Enterovibrio norvegicus	Proteobacteria	327	1	Cap4	1			1748
596	MBP7407614.1		Flavobacteriales bacterium	Bacteroidetes	328	1	Cap4	1			1746
597	ODT04201.1		Gemmatimonadetes bacterium SCN 70-22	Gemmatimonadetes	329	1	Cap4	1			1811
598	WP_075472887.1		Moritella viscosa	Proteobacteria	330	1	Cap4	1			1727
599	WP_105173092.1		Pseudoalteromonas sp. T1lg24	Proteobacteria	331	1	Cap4	1			1745
600	WP_196122811.1		Vibrio sp. NFV-1	Proteobacteria	332	1	Cap4	1			1746
601	WP_131389756.1		Acinetobacter sp. ANC 3791	Proteobacteria	333	1	Cap4	1			1749
602	WP_174492929.1		Acinetobacter sp. Marseille-Q1623	Proteobacteria	334	1	Cap4	1			1743
603	WP_049067819.1		Acinetobacter baumannii	Proteobacteria	335	4	Cap4	1			1743
604	WP_199936627.1		Acinetobacter bereziniae	Proteobacteria	335	4	Cap4	0			
605	WP_104078661.1		Acinetobacter pittii	Proteobacteria	335	4	Cap4	0			
606	WP_068930485.1		Acinetobacter pittii	Proteobacteria	335	4	Cap4	0			
607	WP_023243729.1		Salmonella enterica	Proteobacteria	336	112	Cap4	1			1765
608	EEJ5120819.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
609	EBO5291929.1		Salmonella enterica subsp. enterica serovar Typhimurium	Proteobacteria	336	112	Cap4	0			
610	EBQ8940182.1		Salmonella enterica subsp. enterica serovar Rideau	Proteobacteria	336	112	Cap4	0			
611	ECC3761078.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
612	ECC9025563.1		Salmonella enterica subsp. enterica	Proteobacteria	336	112	Cap4	0			
613	ECK7271965.1		Salmonella enterica subsp. enterica	Proteobacteria	336	112	Cap4	0			

614	MBH527 9457.1		Salmonella enterica	Proteobacteria	33 6	112	Cap4	0			
615	EGB9337 861.1		Salmonella enterica	Proteobacteria	33 6	112	Cap4	0			
616	EBF3225 147.1		Salmonella enterica subsp. enterica serovar Derby	Proteobacteria	33 6	112	Cap4	0			
617	EDW265 7512.1		Salmonella enterica subsp. enterica serovar Havana	Proteobacteria	33 6	112	Cap4	0			
618	WP_0802 17969.1		Salmonella enterica	Proteobacteria	33 6	112	Cap4	0			
619	EBS5001 142.1		Salmonella enterica subsp. enterica serovar Havana	Proteobacteria	33 6	112	Cap4	0			
620	HAV1236 661.1		Salmonella enterica	Proteobacteria	33 6	112	Cap4	0			
621	EBZ6302 717.1		Salmonella enterica subsp. enterica serovar Gombe	Proteobacteria	33 6	112	Cap4	0			
622	WP_2062 26125.1		Salmonella enterica	Proteobacteria	33 6	112	Cap4	0			
623	ECD4882 665.1		Salmonella enterica subsp. enterica serovar Coleypark	Proteobacteria	33 6	112	Cap4	0			
624	EDV7421 364.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	33 6	112	Cap4	0			
625	ECS3046 893.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	33 6	112	Cap4	0			
626	EBN2020 484.1		Salmonella enterica	Proteobacteria	33 6	112	Cap4	0			
627	EAW800 9524.1		Salmonella enterica	Proteobacteria	33 6	112	Cap4	0			
628	ECV9578 157.1		Salmonella enterica	Proteobacteria	33 6	112	Cap4	0			
629	ECB9063 908.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	33 6	112	Cap4	0			
630	EBK5212 658.1		Salmonella enterica	Proteobacteria	33 6	112	Cap4	0			
631	ECQ3958 242.1		Salmonella enterica	Proteobacteria	33 6	112	Cap4	0			
632	EAY0862 313.1		Salmonella enterica	Proteobacteria	33 6	112	Cap4	0			
633	WP_0453 43318.1		Salmonella enterica	Proteobacteria	33 6	112	Cap4	0			
634	EEU8743 754.1		Salmonella enterica	Proteobacteria	33 6	112	Cap4	0			
635	ECS2871 083.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	33 6	112	Cap4	0			
636	EAM1641 665.1		Salmonella enterica	Proteobacteria	33 6	112	Cap4	0			
637	EAO6951 748.1		Salmonella enterica	Proteobacteria	33 6	112	Cap4	0			
638	ECS6954 151.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	33 6	112	Cap4	0			

639	ECX5120138.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
640	EAA0466975.1		Salmonella enterica subsp. enterica	Proteobacteria	336	112	Cap4	0			
641	ECS5219086.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
642	EEH7644862.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
643	EDT2060418.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
644	WP_069541490.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
645	EBZ1331496.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
646	EDE3910639.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
647	EDU1070179.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
648	EBZ0734032.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
649	EBP4153245.1		Salmonella enterica subsp. enterica	Proteobacteria	336	112	Cap4	0			
650	EBX9014100.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
651	EEF4323673.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
652	EEI9066706.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
653	EBF3502884.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
654	EBX1342434.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
655	ECV0820666.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
656	EFR3396725.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
657	WP_039591093.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
658	EBS3575896.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
659	EFB1202646.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
660	EEI9057207.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
661	EEI5842790.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			

662	EDU27082196.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
663	EDC5888721.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
664	EEH6233977.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
665	EDZ7877699.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
666	EEF9854870.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
667	ECD6006804.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
668	EDG2258741.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
669	WP_045791138.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
670	EBJ6116172.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
671	EDQ4526683.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
672	EBG4747258.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
673	ECH0172799.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
674	EAU0243492.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
675	EGE5621347.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
676	EEP7472060.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
677	EDY3291756.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
678	EGB9759844.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
679	WP_214183877.1		Klebsiella quasipneumoniae	Proteobacteria	336	112	Cap4	0			
680	EDN4182227.1		Salmonella enterica subsp. salamae	Proteobacteria	336	112	Cap4	0			
681	WP_109219637.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
682	ECG8596201.1		Salmonella enterica subsp. salamae	Proteobacteria	336	112	Cap4	0			
683	EBH0163743.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
684	EAY9817293.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
685	WP_094445039.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
686	EHM6010433.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
687	EDO1486059.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
688	ECE0218511.1		Salmonella enterica subsp. enterica	Proteobacteria	336	112	Cap4	0			
689	WP_070793948.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			

690	EBP0117635.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
691	EAQ5419848.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
692	EEG7230121.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
693	WP_104446679.1		Citrobacter sp. CFNIH10	Proteobacteria	336	112	Cap4	0			
694	EBQ7316327.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
695	WP_139165059.1		Klebsiella quasipneumoniae	Proteobacteria	336	112	Cap4	0			
696	WP_043000262.1		Citrobacter sp. MGH 55	Proteobacteria	336	112	Cap4	0			
697	WP_141522371.1		Klebsiella	Proteobacteria	336	112	Cap4	0			
698	WP_023290308.1		Klebsiella quasipneumoniae	Proteobacteria	336	112	Cap4	0			
699	EFU3136520.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
700	EAW0645185.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
701	ECZ6304080.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
702	EDV0545055.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
703	EBQ2999646.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
704	EEE2725064.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
705	EAT2901790.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
706	EBO8006798.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
707	WP_210669698.1		Klebsiella oxytoca	Proteobacteria	336	112	Cap4	0			
708	EDV5313650.1		Salmonella enterica subsp. enterica	Proteobacteria	336	112	Cap4	0			
709	EBG5296560.1		Salmonella enterica subsp. enterica	Proteobacteria	336	112	Cap4	0			
710	EBP7401633.1		Salmonella enterica subsp. enterica	Proteobacteria	336	112	Cap4	0			
711	EAO7543034.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
712	ECG8590388.1		Salmonella enterica subsp. salamae	Proteobacteria	336	112	Cap4	0			
713	WP_079774603.1		Salmonella bongori	Proteobacteria	336	112	Cap4	0			
714	EBE8414264.1		Salmonella enterica	Proteobacteria	336	112	Cap4	0			
715	ECF6031258.1		Salmonella enterica subsp. salamae	Proteobacteria	336	112	Cap4	0			
716	ECJ6396979.1		Salmonella enterica subsp. enterica serovar Anatum	Proteobacteria	336	112	Cap4	0			
717	PEN26835.1		Klebsiella sp. MBT K-1	Proteobacteria	336	112	Cap4	0			

718	OFN59454.1		Enterobacter sp. HMSC055A11	Proteobacteria	336	112	Cap4	0			
719	WP_086926337.1		Acidovorax carolinensis	Proteobacteria	337	1	Cap4	1			1767
720	WP_158239149.1		Uliginosibacterium sp. TH139	Proteobacteria	338	1	Cap4	1			1773
721	ASL43707.1		Burkholderia sp. AD24	Proteobacteria	339	1	Cap4	1			1764
722	WP_138679060.1		Candidatus Accumulibacter phosphatis	Proteobacteria	340	2	Cap4	1			1779
723	TMQ74799.1		Candidatus Accumulibacter phosphatis	Proteobacteria	340	2	Cap4	0			
724	TVT73162.1		Denitromonas halophilus	Proteobacteria	341	1	Cap4	1			1763
725	RDE49642.1		Candidatus Accumulibacter phosphatis	Proteobacteria	342	1	Cap4	1			1764
726	MBS0368619.1		Proteobacteria bacterium	Proteobacteria	343	1	Cap4	1			1759
727	WP_157511114.1		Frateuria sp. Soil773	Proteobacteria	344	2	Cap4	1			1761
728	KRE89023.1		Frateuria sp. Soil773	Proteobacteria	344	2	Cap4	0			
729	WP_177025330.1		Pseudomonas yamanorum	Proteobacteria	345	1	Cap4	1			1752
730	WP_175384923.1		Pseudomonas corrugata	Proteobacteria	346	1	Cap4	1			1750
731	WP_166572030.1		Pseudomonas sp. R5(2019)	Proteobacteria	347	2	Cap4	1			1750
732	WP_198792111.1		Pseudomonas sp. MF6784	Proteobacteria	347	2	Cap4	0			
733	WP_143520667.1		Pseudomonas syringae	Proteobacteria	348	3	Cap4	1			1750
734	PBP29238.1		Pseudomonas syringae	Proteobacteria	348	3	Cap4	0			
735	WP_164558430.1		Pseudomonas viridiflava	Proteobacteria	348	3	Cap4	0			
736	WP_204126722.1		Pseudomonas sp. RDP1	Proteobacteria	349	3	Cap4	1			1750
737	WP_125916365.1		Pseudomonas fluorescens	Proteobacteria	349	3	Cap4	0			
738	RRW69061.1		Pseudomonas fluorescens	Proteobacteria	349	3	Cap4	0			
739	WP_157811398.1		Pseudomonas qingdaonensis	Proteobacteria	350	1	Cap4	1			1749
740	OJX40197.1		Chloroflexi bacterium 44-23	Chloroflexi	351	1	Cap4	1			1722
741	WP_210839890.1		Micromonospora sp. C31	Actinobacteria	352	1	Cap4	1			1693
742	KJK42864.1		Lentzea aerocolonigenes	Actinobacteria	353	2	Cap4	1			1823
743	WP_045316098.1		Lentzea aerocolonigenes	Actinobacteria	353	2	Cap4	0			
744	MBM4523044.1		Rhodococcus hoagii	Actinobacteria	354	1	Cap4	1			1965
745	WP_141355261.1		Glutamicibacter nicotianae	Actinobacteria	355	1	Cap4	1			1953
746	WP_022920441.1		Ornithinimicrobium pekingense	Actinobacteria	356	2	Cap4	1			1935
747	GGK64076.1		Ornithinimicrobium pekingense	Actinobacteria	356	2	Cap4	0			
748	KQS63925.1		Modestobacter sp. Leaf380	Actinobacteria	357	2	Cap4	1			1850

749	WP_157519631.1		Modestobacter sp. Leaf380	Actinobacteria	357	2	Cap4	0			
750	MYI77479.1		Gammaproteobacteria bacterium	Proteobacteria	358	2	Cap4	1			1845
751	MXZ56454.1		Gammaproteobacteria bacterium	Proteobacteria	358	2	Cap4	0			
752	MXZ78427.1		Acidimicrobiia bacterium	Actinobacteria	359	1	Cap4	1			1871
753	WP_116416929.1		Subtercola boreus	Actinobacteria	360	1	Cap4	1			1762
754	WP_190599854.1		Synechocystis sp. FACHB-383	Cyanobacteria	361	1	Cap4	1			1869
755	EDZ40398.1		Leptospirillum sp. Group II 5-way CG	Nitrospirae	362	1	Cap4	1			1848
756	WP_189734036.1		Streptomyces aureoverticillatus	Actinobacteria	363	2	Cap4	1			1822
757	GGR99589.1		Streptomyces aureoverticillatus	Actinobacteria	363	2	Cap4	0			
758	WP_203902900.1		Virgisporangium aliadipatigenens	Actinobacteria	364	2	Cap4	1			1825
759	GIJ49425.1		Virgisporangium aliadipatigenens	Actinobacteria	364	2	Cap4	0			
760	WP_153264307.1		Nonomuraea nitratireducens	Actinobacteria	365	1	Cap4	1			1816
761	WP_033442285.1		Saccharothrix sp. NRRL B-16314	Actinobacteria	366	1	Cap4	1			1843
762	WP_043718871.1		Kutzneria sp. 744	Actinobacteria	367	1	Cap4	1			1814
763	WP_129773469.1		Streptomyces sp. L-9-10	Actinobacteria	368	2	Cap4	1			1822
764	RYJ19693.1		Streptomyces sp. L-9-10	Actinobacteria	368	2	Cap4	0			
765	WP_142233769.1		Streptomyces asterosporus	Actinobacteria	369	3	Cap4	1			1830
766	QDI72665.1		Streptomyces asterosporus	Actinobacteria	369	3	Cap4	0			
767	WP_198504315.1		Streptomyces alfalfae	Actinobacteria	369	3	Cap4	0			
768	WP_159024928.1		Streptomyces sp. CB01373	Actinobacteria	370	5	Cap4	1			1829
769	PJM93948.1		Streptomyces sp. CB01373	Actinobacteria	370	5	Cap4	0			
770	WP_210962996.1		Streptomyces sp. b94	Actinobacteria	370	5	Cap4	0			
771	WP_181446908.1		Streptomyces sp. NTH33	Actinobacteria	370	5	Cap4	0			
772	WP_053636830.1		Streptomyces sp. XY152	Actinobacteria	370	5	Cap4	0			
773	WP_189556181.1		Streptomyces pilosus	Actinobacteria	371	2	Cap4	1			1829
774	WP_112438358.1		Streptomyces cadmiisoli	Actinobacteria	371	2	Cap4	0			
775	WP_107112946.1		Streptomyces hygroscopicus	Actinobacteria	372	1	Cap4	1			1829
776	WP_167345939.1		Streptomyces iakyrus	Actinobacteria	373	1	Cap4	1			1823
777	WP_184576034.1		Lipingzhangella halophila	Actinobacteria	374	1	Cap4	1			1824
778	WP_174885752.1		Streptomyces samsunensis	Actinobacteria	375	2	Cap4	1			1823

779	WP_187286748.1		Streptomyces malaysiensis	Actinobacteria	375	2	Cap4	0			
780	WP_128846454.1		Streptomyces lavendulae	Actinobacteria	376	2	Cap4	1			1823
781	WP_214300731.1		Streptomyces sp. Tu10	Actinobacteria	376	2	Cap4	0			
782	WP_159401400.1		Streptomyces sp. NBRC 110468	Actinobacteria	377	1	Cap4	1			1823
783	SDS54760.1		Streptomyces sp. TLI_053	Actinobacteria	378	2	Cap4	1			1824
784	WP_159424672.1		Streptomyces sp. TLI_053	Actinobacteria	378	2	Cap4	0			
785	WP_037624146.1		Streptomyces aureus	Actinobacteria	379	1	Cap4	1			1826
786	WP_129857413.1		Streptomyces albidoflavus	Actinobacteria	380	1	Cap4	1			1823
787	SBU96833.1		Streptomyces sp. Ncost-T6T-1	Actinobacteria	381	2	Cap4	1			1822
788	WP_176711138.1		Streptomyces sp. Ncost-T6T-1	Actinobacteria	381	2	Cap4	0			
789	WP_120021296.1		Amycolatopsis panacis	Actinobacteria	382	1	Cap4_split2	1	WP_120021295.1	385	1865
790	WP_100443950.1		Glycomyces xiaoerkulensis	Actinobacteria	383	1	Cap4_split2	1	--	393	1862
791	WP_084457413.1		Nocardia coubleae	Actinobacteria	384	1	Cap4_split2	1	--	372	1834
792	MAS556491.1		Pimelobacter sp.	Actinobacteria	385	1	Cap4_split2	1	MAS55648.1*	380	1844
793	WP_134395822.1		unclassified Cryobacterium	Actinobacteria	386	1	Cap4_split2	1	WP_134395824.1	389	1857
794	WP_182814434.1		Propioniciclava sp. MC1683	Actinobacteria	387	1	Cap4_split2	1	WP_182814435.1	387	1854
795	WP_046379641.1		Rhodococcus erythropolis	Actinobacteria	388	1	Cap4_split2	1	WP_155404297.1*	386	1851
796	WP_184511458.1		Terracoccus luteus	Actinobacteria	389	1	Cap4_split2	1	WP_184511457.1	389	1856
797	WP_121180219.1		Nocardiopsis sp. Huas11	Actinobacteria	390	1	Cap4_split2	1	WP_147453800.1	384	1850
798	WP_110484418.1		Arthrobacter psychrolactophilus	Actinobacteria	391	1	Cap4_split2	1	WP_110484417.1	386	1855
799	SKY32392.1		Mycobacteroides abscessus subsp. massiliense	Actinobacteria	392	2	Cap4_split2	1	SKY32369.1	390	1857
800	WP_165693960.1		Mycobacteroides abscessus	Actinobacteria	392	2	Cap4_split2	0			
801	WP_150407708.1		Nocardia colli	Actinobacteria	393	2	Cap4_split2	1	WP_150407707.1	360	1828
802	WP_195105959.1		Nocardia blacklockiae	Actinobacteria	393	2	Cap4_split2	0			
803	WP_157743201.1		Micromonospora coriariae	Actinobacteria	394	1	Cap4_split2	1	WP_157743202.1	384	1848
804	WP_088961824.1		Micromonospora purpureochromogenes	Actinobacteria	395	1	Cap4_split2	1	WP_157745965.1	388	1854
805	WP_186461703.1		Micromonospora sp. CNZ297	Actinobacteria	396	1	Cap4_split2	1	WP_145824104.1	388	1854
806	WP_091435238.1		Micromonospora yangpuensis	Actinobacteria	397	2	Cap4_split2	1	WP_139135609.1	388	1854
807	WP_117665791.1		Micromonospora sp. MW-13	Actinobacteria	397	2	Cap4_split2	0			
808	WP_133169447.1		Rhodococcus opacus	Actinobacteria	398	2	Cap4_split2	1	WP_105423671.1	377	1830
809	PQP14015.1		Rhodococcus opacus	Actinobacteria	398	2	Cap4_split2	0			

810	WP_191315914.1		Amycolatopsis bullii	Actinobacteria	399	1	Cap4_split2	1	WP_191315913.1	382	1823
811	WP_076162191.1		Amycolatopsis coloradensis	Actinobacteria	400	1	Cap4_split2	1	WP_076162193.1	382	1823
812	WP_158715013.1		Kitasatospora phosalacinea	Actinobacteria	401	1	Cap4_split2	1	WP_158715014.1	389	1880
813	WP_125540962.1		Streptomyces sp. WAC05292	Actinobacteria	402	1	Cap4_split2	1	WP_125540963.1	398	1876
814	WP_093851967.1		Streptomyces pini	Actinobacteria	403	1	Cap4_split2	1	WP_093851966.1	393	1876
815	SCK04840.1		Streptomyces sp. LamerLS-316	Actinobacteria	404	2	Cap4_split2	1	SCK04843.1	394	1877
816	WP_143672598.1		unclassified Streptomyces	Actinobacteria	404	2	Cap4_split2	0			
817	WP_193483258.1		Streptomyces janthinus	Actinobacteria	405	1	Cap4_split2	1	WP_193483259.1	394	1876
818	WP_062032545.1		Streptomyces griseorubiginosus	Actinobacteria	406	3	Cap4_split2	1	WP_062032542.1	394	1877
819	WP_125501196.1		unclassified Streptomyces	Actinobacteria	406	3	Cap4_split2	0			
820	WP_197924810.1		Streptomyces sp. HB-N217	Actinobacteria	406	3	Cap4_split2	0			
821	WP_146057732.1		Streptomyces sp. SM10	Actinobacteria	407	1	Cap4_split2	1	WP_103514986.1	394	1878
822	WP_028815275.1		Streptomyces flavidovirens	Actinobacteria	408	1	Cap4_split2	1	WP_028815274.1	393	1876
823	WP_053644355.1		unclassified Streptomyces	Actinobacteria	409	8	Cap4_split2	1	WP_146610753.1	391	1875
824	KOX41541.1		Streptomyces sp. NRRL F-6492	Actinobacteria	409	8	Cap4_split2	0			
825	KOX25831.1		Streptomyces sp. NRRL F-6491	Actinobacteria	409	8	Cap4_split2	0			
826	WP_146610752.1		unclassified Streptomyces	Actinobacteria	409	8	Cap4_split2	0			
827	WP_181801066.1		Streptomyces sp. W18L9	Actinobacteria	409	8	Cap4_split2	0			
828	WP_050791664.1		Streptomyces pristinaespiralis	Actinobacteria	409	8	Cap4_split2	0			
829	EDY64031.1		Streptomyces pristinaespiralis ATCC 25486	Actinobacteria	409	8	Cap4_split2	0			
830	WP_030327100.1		Streptomyces sp. NRRL B-3229	Actinobacteria	409	8	Cap4_split2	0			
831	WP_191878421.1		Streptomyces filipinensis	Actinobacteria	410	1	Cap4_split2	1	WP_191878422.1	395	1880
832	WP_215103704.1		Streptomyces sp. McG3	Actinobacteria	411	2	Cap4_split2	1	WP_215103702.1	393	1873
833	WP_215072319.1		Streptomyces sp. ISL-36	Actinobacteria	411	2	Cap4_split2	0			
834	WP_143662829.1		unclassified Streptomyces	Actinobacteria	412	3	Cap4_split2	1	WP_143662830.1	397	1882
835	WP_143669109.1		Streptomyces sp. wa1063	Actinobacteria	412	3	Cap4_split2	0			
836	WP_143671003.1		Streptomyces sp. b84	Actinobacteria	412	3	Cap4_split2	0			
837	WP_078646958.1		Streptomyces hygroscopicus	Actinobacteria	413	1	Cap4_split2	1	WP_078646957.1	394	1877
838	WP_214333654.1		Streptomyces sp. Tu102	Actinobacteria	414	1	Cap4_split2	1	WP_214333655.1	396	1878
839	WP_053665805.1		Streptomyces sp. NRRL F-7442	Actinobacteria	415	1	Cap4_split2	1	WP_053665803.1	393	1884
840	WP_141202819.1		Streptomyces griseorubiginosus	Actinobacteria	416	1	Cap4_split2	1	WP_141202820.1	392	1876

841	DAA9954 3.1		Candidatus Gastranaerophil ales bacterium HUM_10	Candidatus Melainabacteria	41 7	1	Cap4	1			17 50
842	HBG4994 3.1		Cyanobacteria bacterium UBA9971	Cyanobacteria	41 8	1	Cap4	1			17 59
843	MXX585 23.1		Rhodothermace ae bacterium	Bacteroidetes	41 9	2	Cap4	1			17 67
844	MYD199 00.1		Rhodothermace ae bacterium	Bacteroidetes	41 9	2	Cap4	0			
845	WP_0690 73807.1		Nostoc sp. KVJ20	Cyanobacteria	42 0	1	Cap4	1			17 82
846	WP_0965 61258.1		Scytonema sp. NIES-4073	Cyanobacteria	42 1	1	Cap4	1			17 82
847	WP_0277 16778.1		Desulfuromona s sp. TF	Proteobacteria	42 2	1	Cap4	1			17 81
848	NAT1019 1.1		ANME-1 cluster archaeon AG-394-G06	Euryarchaeota	42 3	1	Cap4	1			17 80
849	HEI02307 .1		Paenibacillacea e bacterium	Firmicutes	42 4	1	Cap4	1			17 80
850	WP_1294 41787.1		Paenibacillus protaetiae	Firmicutes	42 5	1	Cap4	1			17 75
851	WP_1851 80740.1		Cohnella lubricantis	Firmicutes	42 6	1	Cap4	1			17 75
852	MBT4196 613.1		Gammaproteob acteria bacterium	Proteobacteria	42 7	1	Cap4	1			17 82
853	WP_0485 14451.1		Megasphaera cerevisiae	Firmicutes	42 8	1	Cap4	1			17 81
854	OPX9171 3.1		Pelotomaculum sp. PtaB.Bin013	Firmicutes	42 9	1	Cap4	1			17 82
855	WP_1284 25676.1		Gudongella oleilytica	Firmicutes	43 0	1	Cap4	1			17 82
856	WP_0945 51400.1		Petroclostridiu m xylanilyticum	Firmicutes	43 1	1	Cap4	1			17 82
857	WP_1474 39940.1		Haloarcula sp. Atlit-7R	Euryarchaeota	43 2	1	Cap4_sh ort	1			12 15
858	WP_1888 53868.1		Haloarcula salaria	Euryarchaeota	43 3	1	Cap4_sh ort	1			12 17
859	WP_1572 31301.1		Halostagnicola larsenii	Euryarchaeota	43 4	1	Cap4_sh ort	1			11 97
860	WP_1617 85354.1		Halorubrum halophilum	Euryarchaeota	43 5	1	Cap4_sh ort	1			12 24
861	WP_1358 29852.1		Halorussus sp. HD8-83	Euryarchaeota	43 6	1	Cap4_sh ort	1			12 28
862	WP_1914 51671.1		Halogeometricu m sp. CBA1124	Euryarchaeota	43 7	1	Cap4_sh ort	1			12 27
863	WP_1473 47829.1		Collinsella sp. TM05-38	Actinobacteria	43 8	2	Cap4	1			16 82
864	RGJ6889 0.1		Collinsella sp. TM05-38	Actinobacteria	43 8	2	Cap4	0			
865	WP_0553 10717.1		Collinsella aerofaciens	Actinobacteria	43 9	2	Cap4	1			16 71
866	WP_1180 65450.1		Collinsella sp. AM33-4BH	Actinobacteria	43 9	2	Cap4	0			
867	KAB1635 984.1		Ellagibacter isourolithinifaci ens	Actinobacteria	44 0	2	Cap4	1			16 71
868	WP_2113 61098.1		Ellagibacter isourolithinifaci ens	Actinobacteria	44 0	2	Cap4	0			
869	OKZ7591 8.1		Clostridium sp. 26_22	Firmicutes	44 1	1	Cap4	1			16 71
870	MBD894 1444.1		Clostridiales bacterium	Firmicutes	44 2	1	Cap4	1			16 71

871	WP_182536476.1		Fontibacillus solani	Firmicutes	443	1	Cap4	1			1700
872	WP_089438672.1		Blautia hansenii	Firmicutes	444	1	Cap4	1			1671
873	WP_076235080.1		Clostridium botulinum	Firmicutes	445	1	Cap4	1			1675
874	WP_027638541.1		Clostridium cadaveris	Firmicutes	446	2	Cap4	1			1677
875	WP_074844505.1		Clostridium cadaveris	Firmicutes	446	2	Cap4	0			
876	WP_086582618.1		Lactococcus petauri	Firmicutes	447	1	Cap4	1			1678
877	WP_185521574.1		Listeria booriae	Firmicutes	448	2	Cap4	1			1675
878	EAD0711469.1		Listeria monocytogenes	Firmicutes	448	2	Cap4	0			
879	WP_046326361.1		Listeria seeligeri	Firmicutes	449	1	Cap4	1			1676
880	WP_098205474.1		Bacillus cereus group	Firmicutes	450	7	Cap4	1			1692
881	WP_106102018.1		Bacillus wiedmannii	Firmicutes	450	7	Cap4	0			
882	WP_098357692.1		unclassified Bacillus (in: Bacteria)	Firmicutes	450	7	Cap4	0			
883	WP_106405464.1		Bacillus wiedmannii	Firmicutes	450	7	Cap4	0			
884	WP_097919021.1		Bacillus cereus	Firmicutes	450	7	Cap4	0			
885	WP_000106701.1		Bacillus cereus	Firmicutes	450	7	Cap4	0			
886	WP_137049954.1		Bacillus cereus	Firmicutes	450	7	Cap4	0			
887	WP_119493377.1		Mammaliicoccus sciuri	Firmicutes	451	1	Cap4	1			1684
888	WP_094244789.1		Aeribacillus pallidus	Firmicutes	452	1	Cap4	1			1687
889	WP_101635524.1		Bacillus sp. V5-8f	Firmicutes	453	1	Cap4	1			1684
890	WP_090869186.1		Oceanobacillus limi	Firmicutes	454	1	Cap4	1			1683
891	WP_077595967.1		Oceanobacillus iheyensis	Firmicutes	455	1	Cap4	1			1685
892	WP_095255941.1		Alkalihalobacillus clausii	Firmicutes	456	1	Cap4	1			1686
893	WP_203339610.1		Planomicrobium sp. REN14	Firmicutes	457	1	Cap4	1			1686
894	WP_152393382.1		Paenibacillus guangzhouensis	Firmicutes	458	1	Cap4	1			1689
895	WP_016132786.1		Bacillus cereus	Firmicutes	459	1	Cap4	1			1688
896	WP_144559770.1		Bacillus thuringiensis	Firmicutes	460	1	Cap4	1			1688
897	WP_000106702.1		Bacillus cereus	Firmicutes	461	2	Cap4	1			1689
898	QEF16827.1		Bacillus cereus	Firmicutes	461	2	Cap4	0			
899	WP_152122030.1		Bacillus sp. B1-WWTP-T-0.5-Post-4	Firmicutes	462	1	Cap4	1			1687
900	NTL99137.1		Enterococcus faecium	Firmicutes	463	1	Cap4	1			1621
901	WP_034700198.1		Enterococcus hirae	Firmicutes	464	1	Cap4	1			1621
902	WP_141442202.1		Enterococcus faecalis	Firmicutes	465	1	Cap4	1			1619
903	WP_140451912.1		Priestia megaterium	Firmicutes	466	1	Cap4	1			1631
904	WP_062197145.1		Massilibacterium senegalense	Firmicutes	467	1	Cap4	1			1633

905	WP_098607432.1		Bacillus cereus	Firmicutes	468	1	Cap4	1			1632
906	WP_176549876.1		Bacillus cereus group	Firmicutes	469	2	Cap4	1			1633
907	PHG45976.1		Bacillus wiedmannii	Firmicutes	469	2	Cap4	0			
908	WP_064056217.1		Bacillus wiedmannii	Firmicutes	470	1	Cap4	1			1632
909	HAR58302.1		Acholeplasmatales bacterium	Tenericutes	471	2	Cap4_short	1			1144
910	MBS5852467.1		Staphylococcus sp.	Firmicutes	471	2	Cap4_short	0			
911	MBR5090676.1		Ruminiclostridium sp.	Firmicutes	472	1	Cap4_short	1			1229
912	WP_074963502.1		Ruminococcus albus	Firmicutes	473	1	Cap4_short	1			1231
913	WP_117865428.1		Ruminococcus sp. AF16-50	Firmicutes	474	2	Cap4_short	1			1230
914	MBS6918574.1		Ruminococcus bicirculans	Firmicutes	474	2	Cap4_short	0			
915	MBQ7891033.1		Erysipelotrichaceae bacterium	Firmicutes	475	1	Cap4_short	1			1232
916	MBQ2815562.1		Clostridia bacterium	Firmicutes	476	1	Cap4_short	1			1227
917	MBS6686200.1		[Clostridium] spiroforme	Firmicutes	477	1	Cap4_short	1			1223
918	MBQ8804394.1		Tyzzerella sp.	Firmicutes	478	1	Cap4_short	1			1222
919	WP_050047188.1		Enterococcus cecorum	Firmicutes	479	2	Cap4_short	1			1213
920	WP_087663059.1		Enterococcus cecorum	Firmicutes	479	2	Cap4_short	0			
921	WP_118198077.1		Blautia obeum	Firmicutes	480	14	Cap4_short	1			1213
922	RHE99511.1		[Eubacterium] rectale	Firmicutes	480	14	Cap4_short	0			
923	WP_077327351.1		Anaerostipes hadrus	Firmicutes	480	14	Cap4_short	0			
924	WP_055144652.1		Coprococcus eutactus	Firmicutes	480	14	Cap4_short	0			
925	WP_215718249.1		Anaerostipes hadrus	Firmicutes	480	14	Cap4_short	0			
926	WP_209301506.1		Anaerobutyricum soehngenii	Firmicutes	480	14	Cap4_short	0			
927	WP_186847587.1		Coprococcus sp. NSJ-10	Firmicutes	480	14	Cap4_short	0			
928	WP_215633524.1		Coprococcus catus	Firmicutes	480	14	Cap4_short	0			
929	WP_117753998.1		Blautia obeum	Firmicutes	480	14	Cap4_short	0			
930	WP_119234989.1		Ruminococcus sp. AF41-9	Firmicutes	480	14	Cap4_short	0			
931	RGU28480.1		[Eubacterium] rectale	Firmicutes	480	14	Cap4_short	0			
932	WP_044921014.1		Anaerostipes hadrus	Firmicutes	480	14	Cap4_short	0			
933	HAS06749.1		Eubacterium sp.	Firmicutes	480	14	Cap4_short	0			
934	WP_118091839.1		Roseburia inulinivorans	Firmicutes	480	14	Cap4_short	0			
935	WP_019382434.1		Cytobacillus oceanisediminis	Firmicutes	481	1	Cap4_short	1			1250
936	WP_098319282.1		Bacillus cereus	Firmicutes	482	2	Cap4_short	1			1254
937	WP_098645925.1		Bacillus toyonensis	Firmicutes	482	2	Cap4_short	0			
938	MBO5210148.1		Lachnospiraceae bacterium	Firmicutes	483	1	Cap4_short	1			1239
939	WP_054697971.1		Syntrophomonas palmitatica	Firmicutes	484	1	Cap4_short	1			1271

940	WP_004622570.1		Ruminiclostridium papyrosolvens	Firmicutes	485	1	Cap4_short	1			1289
941	WP_126048499.1		Bacillus acidinfaciens	Firmicutes	486	1	Cap4_short	1			1280
942	MBS4814975.1		Clostridium sp.	Firmicutes	487	1	Cap4_short	1			1276
943	WP_025581139.1		Blautia wexlerae	Firmicutes	488	1	Cap4_short	1			1274
944	WP_025580498.1		Blautia wexlerae	Firmicutes	489	1	Cap4_short	1			1279
945	WP_072976247.1		Tissierella praeacuta	Firmicutes	490	1	Cap4_short	1			1248
946	WP_000203570.1		Leptospira interrogans	Spirochaetes	491	8	Cap4	1			1692
947	WP_192505140.1		Leptospira interrogans	Spirochaetes	491	8	Cap4	0			
948	WP_000203568.1		Leptospira interrogans	Spirochaetes	491	8	Cap4	0			
949	WP_192485961.1		Leptospira interrogans	Spirochaetes	491	8	Cap4	0			
950	WP_061236989.1		Leptospira interrogans	Spirochaetes	491	8	Cap4	0			
951	WP_192485982.1		Leptospira interrogans	Spirochaetes	491	8	Cap4	0			
952	WP_061252050.1		Leptospira interrogans	Spirochaetes	491	8	Cap4	0			
953	WP_061234000.1		Leptospira interrogans	Spirochaetes	491	8	Cap4	0			
954	WP_173907186.1		Blautia obeum	Firmicutes	492	1	Cap4	1			1698
955	WP_016359846.1		Eisenbergiella tayi	Firmicutes	493	1	Cap4	1			1696
956	MBQ6960994.1		Clostridia bacterium	Firmicutes	494	1	Cap4	1			1669
957	MBQ6273143.1		Oscillospiraceae bacterium	Firmicutes	495	1	Cap4	1			1670
958	WP_117660579.1		Dorea formicigenerans	Firmicutes	496	2	Cap4	1			1689
959	WP_118224810.1		Dorea longicatena	Firmicutes	496	2	Cap4	0			
960	WP_027430486.1		Lachnospira multipara	Firmicutes	497	1	Cap4	1			1675
961	AVM70506.1		Lachnospiraceae bacterium oral taxon 500	Firmicutes	498	1	Cap4	1			1677
962	WP_102198449.1		Anaerococcus hydrogenalis	Firmicutes	499	4	Cap4	1			1677
963	WP_070610524.1		Peptoniphilus sp. HMSC062D09	Firmicutes	499	4	Cap4	0			
964	WP_198400721.1		Finegoldia magna	Firmicutes	499	4	Cap4	0			
965	EFL54219.1		Finegoldia magna BVS033A4	Firmicutes	499	4	Cap4	0			
966	EGG91738.1		Lachnospiraceae oral taxon 107 str. F0167	Firmicutes	500	1	Cap4	1			1676
967	WP_087174758.1		Lachnoclostridium sp. An118	Firmicutes	501	1	Cap4	1			1681
968	TGX99773.1		Lachnospiraceae bacterium	Firmicutes	502	1	Cap4	1			1678
969	WP_130862154.1		Bacilliculturomica massiliensis	Firmicutes	503	1	Cap4	1			1678
970	WP_047774584.1		Faecalicatena fissicatena	Firmicutes	504	1	Cap4	1			1678
971	WP_153918681.1		Faecalicatena contorta	Firmicutes	505	1	Cap4	1			1678

972	WP_038572035.1		Paenibacillus odorifer	Firmicutes	506	2	Cap4	1			1706
973	WP_071221439.1		Paenibacillus sp. LC231	Firmicutes	506	2	Cap4	0			
974	PKO18731.1		candidate division BRC1 bacterium HGW-BRC1-1	Unclassified	507	1	Cap4	1			1724
975	WP_110603190.1		Salinicola lusitanus	Proteobacteria	508	1	Cap4	1			1690
976	WP_082820126.1		Desulfuromonas sp. DDH964	Proteobacteria	509	2	Cap4	1			1712
977	AMV70735.1		Desulfuromonas sp. DDH964	Proteobacteria	509	2	Cap4	0			
978	MBI5551332.1		Desulfobacterales bacterium	Proteobacteria	510	1	Cap4	1			1713
979	MBN8457812.1		Verrucomicrobia bacterium	Verrucomicrobia	511	1	Cap4	1			1716
980	WP_145122294.1		Rosistilla oblonga	Planctomycetes	512	1	Cap4	1			1621
981	WP_197451388.1		Rosistilla oblonga	Planctomycetes	513	2	Cap4	1			1642
982	WP_197452660.1		Rosistilla oblonga	Planctomycetes	513	2	Cap4	0			
983	WP_145427571.1		Symmachiella dynata	Planctomycetes	514	1	Cap4	1			1642
984	WP_196591984.1		Pectinatus frisingensis	Firmicutes	515	1	Cap4	1			1735
985	WP_076557829.1		Salimicrobium flavidum	Firmicutes	516	1	Cap4	1			1735
986	WP_074103177.1		Paenibacillus sp. P3E	Firmicutes	517	1	Cap4	1			1734
987	WP_046180959.1		Domibacillus tundrae	Firmicutes	518	1	Cap4	1			1739
988	WP_021728395.1		Bacillus cereus group	Firmicutes	519	2	Cap4	1			1722
989	WP_159146497.1		Bacillus sp. 71	Firmicutes	519	2	Cap4	0			
990	WP_153406742.1		Gracilibacillus sp. TD8	Firmicutes	520	1	Cap4	1			1722
991	WP_214917588.1		Bacillus sp. ISL-32	Firmicutes	521	1	Cap4	1			1721
992	WP_149868163.1		Bacillus sp. S3	Firmicutes	522	1	Cap4	1			1721
993	WP_095149410.1		Bacillus sp. BO	Firmicutes	523	1	Cap4	1			1723
994	WP_179879327.1		Bacillus cereus	Firmicutes	524	4	Cap4	1			1722
995	PFC52036.1		Bacillus cereus	Firmicutes	524	4	Cap4	0			
996	WP_180230524.1		Bacillus cereus	Firmicutes	524	4	Cap4	0			
997	PFW02675.1		Bacillus cereus	Firmicutes	524	4	Cap4	0			
998	WP_167655136.1		Pseudomonas grimontii	Proteobacteria	525	1	Cap4	1			1855
999	WP_116641807.1		Pseudomonas kilonensis	Proteobacteria	526	1	Cap4	1			1855
1000	WP_038848927.1		Pseudomonas	Proteobacteria	527	6	Cap4	1			1870
1001	WP_177052516.1		Pseudomonas reactans	Proteobacteria	527	6	Cap4	0			
1002	WP_016978904.1		Pseudomonas	Proteobacteria	527	6	Cap4	0			
1003	WP_178910887.1		Pseudomonas tolaasii	Proteobacteria	527	6	Cap4	0			
1004	WP_177111834.1		Pseudomonas reactans	Proteobacteria	527	6	Cap4	0			
1005	WP_177075510.1		Pseudomonas sp. P7759	Proteobacteria	527	6	Cap4	0			

1006	WP_132940487.1		Vibrio crassostreae	Proteobacteria	528	1	Cap4	1			1867
1007	MBN8852660.1		Sphingobacteriales bacterium	Bacteroidetes	529	1	Cap4	1			1864
1008	WP_188242832.1		Maribacter sp. ANRC-HE7	Bacteroidetes	530	1	Cap4	1			1897
1009	MBP8993597.1		Bacteroidales bacterium	Bacteroidetes	531	1	Cap4	1			1905
1010	WP_126651917.1		Chryseobacterium aureum	Bacteroidetes	532	2	Cap4	1			1904
1011	WP_065396164.1		Chryseobacterium artocarpi	Bacteroidetes	532	2	Cap4	0			
1012	WP_057119467.1		Flavobacterium sp. Root935	Bacteroidetes	533	1	Cap4	1			1925
1013	NDP22685.1		Paludibacter sp.	Bacteroidetes	534	1	Cap4	1			1906
1014	WP_026984383.1		Flavobacterium sp. URHB0058	Bacteroidetes	535	1	Cap4	1			1905
1015	WP_196853640.1		Flavobacterium sp. CAN_S2	Bacteroidetes	536	1	Cap4	1			1902
1016	WP_116547483.1		Chryseobacterium sp. HMWF035	Bacteroidetes	537	1	Cap4	1			1926
1017	WP_165688627.1		Flavobacterium columnare	Bacteroidetes	538	2	Cap4	1			1907
1018	OOB83897.1		Flavobacterium columnare	Bacteroidetes	538	2	Cap4	0			
1019	WP_195157348.1		Flavobacterium columnare	Bacteroidetes	539	2	Cap4	1			1899
1020	WP_107317786.1		Flavobacterium columnare	Bacteroidetes	539	2	Cap4	0			
1021	WP_164002027.1		Pyxidicoccus caerfyrddinensis	Proteobacteria	540	1	MBL + serine protease	1			1741
1022	MBQ2982324.1		Lachnospiraceae bacterium	Firmicutes	541	1	MBL + serine protease	1			1712
1023	WP_077894775.1		Clostridium felsineum	Firmicutes	542	1	MBL + serine protease	1			1745
1024	MBE6048191.1		Clostridium sp.	Firmicutes	543	1	MBL + serine protease	1			1746
1025	WP_077306018.1		Clostridium beijerinckii	Firmicutes	544	1	MBL + serine protease	1			1693
1026	WP_213416595.1		Wohlfahrtiimonas chitiniclastica	Proteobacteria	545	1	MBL + serine protease short	1			1530
1027	WP_192508876.1		Pseudoalteromonas aurantia	Proteobacteria	546	1	MBL + serine protease	1			1729
1028	WP_067205420.1		Sporosarcina psychrophila	Firmicutes	547	1	MBL + serine protease	1			1741
1029	WP_028560371.1		Paenibacillus pinihumi	Firmicutes	548	3	MBL + serine protease	1			1749
1030	RAV19112.1		Paenibacillus contaminans	Firmicutes	548	3	MBL + serine protease	0			
1031	WP_158560620.1		Paenibacillus contaminans	Firmicutes	548	3	MBL + serine protease	0			
1032	WP_201632810.1		Tumebacillus sp. ITR2	Firmicutes	549	1	MBL + serine protease	1			1745
1033	WP_179924659.1		Priestia megaterium	Firmicutes	550	2	MBL + serine protease	1			1729

1034	PFA94009.1		Priestia megaterium	Firmicutes	550	2	MBL + serine protease	0			
1035	WP_045852208.1		Domibacillus enclensis	Firmicutes	551	1	MBL + serine protease	1			1730
1036	WP_194758341.1		Priestia megaterium	Firmicutes	552	1	MBL + serine protease	1			1725
1037	WP_000582846.1		Bacillus cereus	Firmicutes	553	3	MBL + serine protease	1			1727
1038	WP_088351337.1		Bacillus cereus	Firmicutes	553	3	MBL + serine protease	0			
1039	WP_098056186.1		Bacillus toyonensis	Firmicutes	553	3	MBL + serine protease	0			
1040	WP_132109374.1		Flavobacterium hiemivividum	Bacteroidetes	554	1	MBL + serine protease	1			1715
1041	WP_094177926.1		Flavobacterium psychrophilum	Bacteroidetes	555	2	MBL + serine protease	1			1720
1042	WP_134391819.1		Flavobacterium psychrophilum	Bacteroidetes	555	2	MBL + serine protease	0			
1043	WP_104715569.1		Siphonobacter curvatus	Bacteroidetes	556	1	MBL + serine protease	1			1722
1044	WP_188686625.1		Rufibacter glacialis	Bacteroidetes	557	2	MBL + serine protease	1			1727
1045	KAA6434746.1		Rufibacter glacialis	Bacteroidetes	557	2	MBL + serine protease	0			
1046	OYD40174.1		Sphingobacterium cellulitidis	Bacteroidetes	558	2	MBL + serine protease	1			1728
1047	WP_134090161.1		Olivibacter sp. XZL3	Bacteroidetes	558	2	MBL + serine protease	0			
1048	WP_072922723.1		Chryseobacterium sp. OV279	Bacteroidetes	559	1	MBL + serine protease	1			1731
1049	WP_105703209.1		Chryseobacterium sp. MYb7	Bacteroidetes	560	1	MBL + serine protease	1			1727
1050	WP_122547944.1		Chryseobacterium nematophagum	Bacteroidetes	561	1	MBL + serine protease	1			1727
1051	WP_106004045.1		Elizabethkingia anophelis	Bacteroidetes	562	1	MBL + serine protease	1			1726
1052	WP_123872511.1		Chryseobacterium bernardetii	Bacteroidetes	563	2	MBL + serine protease	1			1727
1053	WP_080780211.1		Chryseobacterium phocaeense	Bacteroidetes	563	2	MBL + serine protease	0			
1054	WP_196292824.1		Hymenobacter sp. BT662	Bacteroidetes	564	1	MBL + serine protease short	1			1509
1055	WP_111480539.1		Hymenobacter edaphi	Bacteroidetes	565	1	MBL + serine protease short	1			1536
1056	WP_019949103.1		Hymenobacter aerophilus	Bacteroidetes	566	1	MBL + serine	1			1536

							protease _short				
1057	WP_1427 12806.1		Aliifodinibius sediminis	Balneolaeota	56 7	1	MBL + serine protease _short	1			15 34
1058	WP_0911 29950.1		Flavobacterium anhuiense	Bacteroidetes	56 8	1	MBL + serine protease _short	1			15 37
1059	WP_0885 83967.1		Chryseobacteri um indologenes	Bacteroidetes	56 9	2	MBL + serine protease _short	1			15 36
1060	WP_0473 78087.1		Chryseobacteri um sp. YR459	Bacteroidetes	56 9	2	MBL + serine protease _short	0			
1061	WP_1238 71764.1		Chryseobacteri um bernardetii	Bacteroidetes	57 0	2	MBL + serine protease _short	1			15 36
1062	HCM354 38.1		Chryseobacteri um sp.	Bacteroidetes	57 0	2	MBL + serine protease _short	0			
1063	MBR0500 344.1		Bacteroidales bacterium	Bacteroidetes	57 1	1	MBL + serine protease _short	1			14 49
1064	CDC9982 5.1		Prevotella sp. CAG:474	Bacteroidetes	57 2	1	MBL + serine protease _short	1			14 49
1065	MBQ278 4424.1		Alistipes sp.	Bacteroidetes	57 3	1	MBL + serine protease _short	1			14 49
1066	WP_1530 82097.1		Prevotella copri	Bacteroidetes	57 4	3	MBL + serine protease _short	1			14 47
1067	WP_1175 87920.1		Prevotella copri	Bacteroidetes	57 4	3	MBL + serine protease _short	0			
1068	MBN294 1615.1		Streptococcus sp.	Firmicutes	57 4	3	MBL + serine protease _short	0			
1069	TWV143 64.1		Bacteroidaceae bacterium HV4-6-C5C	Bacteroidetes	57 5	1	MBL + serine protease _short	1			14 48
1070	WP_1308 93272.1		Paraprevotella xylaniphila	Bacteroidetes	57 6	1	MBL + serine protease _short	1			14 49
1071	WP_0532 22412.1		Roseivirga seohaensis	Bacteroidetes	57 7	1	MBL + serine protease _short	1			15 06
1072	WP_0026 96562.1		Microscilla marina	Bacteroidetes	57 8	1	MBL + serine protease _short	1			15 07
1073	WP_0349 20876.1		Gillisia sp. CAL575	Bacteroidetes	57 9	1	MBL + serine protease _short	1			15 05

1074	WP_027393240.1		Aquimarina latercula	Bacteroidetes	580	1	MBL + serine protease short	1			1503
1075	WP_127345025.1		Ancylomarina longa	Bacteroidetes	581	1	MBL + serine protease short	1			1507
1076	WP_017496413.1		Flavobacterium sp. WG21	Bacteroidetes	582	1	MBL + serine protease short	1			1510
1077	WP_125012980.1		Flavobacterium macacae	Bacteroidetes	583	1	MBL + serine protease short	1			1507
1078	WP_017496772.1		Flavobacterium sp. WG21	Bacteroidetes	584	1	MBL + serine protease short	1			1508
1079	WP_171568516.1		Flavobacterium sp. CLA17	Bacteroidetes	585	1	MBL + serine protease short	1			1508
1080	WP_019977050.1		Empedobacter brevis	Bacteroidetes	586	1	MBL + serine protease short	1			1509
1081	WP_034675215.1		Chryseobacterium formosense	Bacteroidetes	587	1	MBL + serine protease short	1			1508
1082	WP_116097052.1		Chryseobacterium sp. 5_R23647	Bacteroidetes	588	2	MBL + serine protease short	1			1506
1083	WP_123904577.1		Chryseobacterium haifense	Bacteroidetes	588	2	MBL + serine protease short	0			
1084	WP_076552671.1		Flavobacteriales	Bacteroidetes	589	1	MBL + serine protease short	1			1506
1085	WP_138982518.1		Elizabethkingia sp. JS20170427COW	Bacteroidetes	590	1	MBL + serine protease short	1			1507
1086	WP_094281607.1		Sphingobacterium cellulitidis	Bacteroidetes	591	1	MBL + serine protease short	1			1507
1087	WP_078407482.1		Elizabethkingia anophelis	Bacteroidetes	592	1	MBL + serine protease short	1			1507
1088	WP_115000498.1		Empedobacter falsenii	Bacteroidetes	593	1	MBL + serine protease short	1			1507
1089	WP_092048403.1		Planctomicrobium piriforme	Planctomycetes	594	1	MBL + serine protease	1			1731
1090	OJU97783.1		Verrucomicrobia bacterium 61-8	Verrucomicrobia	595	1	MBL + serine protease	1			1737
1091	WP_153248553.1		Candidatus Methylospira mobilis	Proteobacteria	596	1	MBL + serine protease	1			1738

1092	MBA4317623.1		Flavobacterium sp.	Bacteroidetes	597	1	MBL + serine protease	1			1750
1093	WP_081149270.1		Colwelliaceae	Proteobacteria	598	1	MBL + serine protease	1			1723
1094	WP_182716635.1		Pseudoalteromonas sp. SG45-2	Proteobacteria	599	1	MBL + serine protease	1			1723
1095	WP_021186232.1		Pantoea sp. AS-PWVM4	Proteobacteria	600	1	MBL + serine protease	1			1714
1096	WP_048696970.1	EpAvs1	Erwinia piriflorinigrans	Proteobacteria	601	2	MBL + serine protease	1			1722
1097	CCG86505.1		Erwinia piriflorinigrans CFBP 5888	Proteobacteria	601	2	MBL + serine protease	0			
1098	WP_203418049.1		Pseudomonas granadensis	Proteobacteria	602	1	MBL + serine protease	1			1732
1099	WP_095129143.1		Pseudomonas sp. Irchel s3h14	Proteobacteria	603	1	MBL + serine protease	1			1735
1100	WP_101409975.1		Vibrio cholerae	Proteobacteria	604	5	MBL + serine protease	1			1727
1101	WP_114742670.1		Vibrio cholerae	Proteobacteria	604	5	MBL + serine protease	0			
1102	WP_148504005.1		Vibrio cholerae	Proteobacteria	604	5	MBL + serine protease	0			
1103	HAS3608186.1		Vibrio cholerae	Proteobacteria	604	5	MBL + serine protease	0			
1104	WP_126512209.1		Shewanella putrefaciens	Proteobacteria	604	5	MBL + serine protease	0			
1105	WP_198486583.1		Achromobacter deleyi	Proteobacteria	605	1	MBL + serine protease	1			1726
1106	WP_215398796.1		Rheinheimera sp. GL-53	Proteobacteria	606	1	MBL + serine protease	1			1720
1107	WP_165861676.1		Pseudomonas sp. GL93	Proteobacteria	607	2	MBL + serine protease	1			1723
1108	RFD24354.1		Pseudomonas sp. GL93	Proteobacteria	607	2	MBL + serine protease	0			
1109	WP_163033706.1		Pseudomonas viridiflava	Proteobacteria	608	1	MBL + serine protease	1			1724
1110	WP_131024349.1		Klebsiella variicola	Proteobacteria	609	3	MBL + serine protease	1			1710
1111	WP_168980525.1		Enterobacter asburiae	Proteobacteria	609	3	MBL + serine protease	0			
1112	HBC5385327.1		Citrobacter koseri	Proteobacteria	609	3	MBL + serine protease	0			
1113	WP_186530605.1		Serratia fonticola	Proteobacteria	610	1	MBL + serine protease	1			1701
1114	WP_039345188.1		Pectobacterium fontis	Proteobacteria	611	3	MBL + serine protease	1			1707

1115	WP_211023884.1		Pectobacterium versatile	Proteobacteria	611	3	MBL + serine protease	0			
1116	WP_205595145.1		Pectobacterium brasiliense	Proteobacteria	611	3	MBL + serine protease	0			
1117	WP_210905080.1		Citrobacter sp. On2M	Proteobacteria	612	1	MBL + serine protease	1			1704
1118	WP_204999891.1		Shewanella sp. 202IG2-18	Proteobacteria	613	1	MBL + serine protease	1			1715
1119	WP_204084550.1		Acinetobacter pittii	Proteobacteria	614	1	MBL + serine protease	1			1713
1120	WP_125699533.1		Acinetobacter lactucae	Proteobacteria	615	2	MBL + serine protease	1			1713
1121	WP_056479770.1		Acinetobacter sp. Leaf130	Proteobacteria	615	2	MBL + serine protease	0			
1122	BCX75212.1		Acinetobacter sp. Tol 5	Proteobacteria	616	1	MBL + serine protease	1			1715
1123	WP_165724365.1		unclassified Pseudoalteromonas	Proteobacteria	617	3	MBL + serine protease	1			1718
1124	WP_182758730.1		Pseudoalteromonas sp. SR41-4	Proteobacteria	617	3	MBL + serine protease	0			
1125	HEA15251.1		Pseudoalteromonas prydzensis	Proteobacteria	617	3	MBL + serine protease	0			
1126	WP_186507224.1		Neptunicella marina	Proteobacteria	618	1	MBL + serine protease	1			1713
1127	WP_089367668.1		Pseudoalteromonas nigrifaciens	Proteobacteria	619	1	MBL + serine protease	1			1716
1128	WP_016247042.1		Escherichia coli	Proteobacteria	620	6	MBL + serine protease	1			1715
1129	WP_124036340.1		Escherichia coli	Proteobacteria	620	6	MBL + serine protease	0			
1130	WP_064594036.1		Mangrovibacter phragmitis	Proteobacteria	620	6	MBL + serine protease	0			
1131	MBB9298199.1		Escherichia coli	Proteobacteria	620	6	MBL + serine protease	0			
1132	EGZ5754955.1		Salmonella enterica	Proteobacteria	620	6	MBL + serine protease	0			
1133	EAM7167742.1		Salmonella enterica	Proteobacteria	620	6	MBL + serine protease	0			
1134	EAW9405837.1		Salmonella enterica	Proteobacteria	621	1	MBL + serine protease	1			1715
1135	HBA3762423.1		Escherichia coli	Proteobacteria	622	4	MBL + serine protease	1			1714
1136	WP_160399310.1		Escherichia coli	Proteobacteria	622	4	MBL + serine protease	0			
1137	MBB7656035.1		Escherichia coli	Proteobacteria	622	4	MBL + serine protease	0			

1138	EBR3968469.1		Salmonella enterica	Proteobacteria	622	4	MBL + serine protease	0			
1139	WP_047741172.1		Enterobacter hormaechei	Proteobacteria	623	2	MBL + serine protease	1			1714
1140	WP_181352270.1		Klebsiella quasipneumoniae	Proteobacteria	623	2	MBL + serine protease	0			
1141	WP_033071176.1		Pectobacterium parmentieri	Proteobacteria	624	4	MBL + serine protease	1			1715
1142	WP_043135270.1		Pectobacterium atrosepticum	Proteobacteria	624	4	MBL + serine protease	0			
1143	WP_129936319.1		Serratia proteamaculans	Proteobacteria	624	4	MBL + serine protease	0			
1144	WP_141827970.1		Klebsiella pneumoniae	Proteobacteria	624	4	MBL + serine protease	0			
1145	EGQ8007170.1		Vibrio parahaemolyticus	Proteobacteria	625	1	MBL + serine protease	1			1709
1146	WP_106728654.1		Zobellella endophytica	Proteobacteria	626	1	MBL + serine protease	1			1710
1147	WP_171626732.1		Pseudoalteromonas caenipelagi	Proteobacteria	627	1	MBL + serine protease	1			1705
1148	WP_086997916.1		Pseudoalteromonas sp. JB197	Proteobacteria	628	2	MBL + serine protease	1			1709
1149	WP_208834122.1		Pseudoalteromonas sp. K222D	Proteobacteria	628	2	MBL + serine protease	0			
1150	WP_208128816.1		Shewanella algae	Proteobacteria	629	1	MBL + serine protease	1			1709
1151	WP_058667628.1		Vibrio vulnificus	Proteobacteria	630	12	MBL + serine protease	1			1709
1152	WP_208445042.1		Vibrio sp. B1ASS3	Proteobacteria	630	12	MBL + serine protease	0			
1153	WP_171392418.1		Vibrio mediterranei	Proteobacteria	630	12	MBL + serine protease	0			
1154	WP_132960215.1		Vibrio crassostreae	Proteobacteria	630	12	MBL + serine protease	0			
1155	WP_155690963.1		Aliivibrio fischeri	Proteobacteria	630	12	MBL + serine protease	0			
1156	EGQ9820091.1		Vibrio parahaemolyticus	Proteobacteria	630	12	MBL + serine protease	0			
1157	WP_101111692.1		Vibrio sp. vnigr-6D03	Proteobacteria	630	12	MBL + serine protease	0			
1158	WP_050922758.1		Vibrio harveyi	Proteobacteria	630	12	MBL + serine protease	0			
1159	WP_210473106.1		Vibrio crassostreae	Proteobacteria	630	12	MBL + serine protease	0			
1160	WP_102534847.1		Vibrio splendidus	Proteobacteria	630	12	MBL + serine protease	0			

1161	HAS6495802.1		Vibrio parahaemolyticus	Proteobacteria	630	12	MBL + serine protease	0			
1162	WP_108173273.1		Vibrio splendidus	Proteobacteria	630	12	MBL + serine protease	0			
1163	WP_063607275.1		Vibrio cholerae	Proteobacteria	631	12	MBL + serine protease	1			1709
1164	EGQ8087864.1		Vibrio vulnificus	Proteobacteria	631	12	MBL + serine protease	0			
1165	EGR0206928.1		Vibrio vulnificus	Proteobacteria	631	12	MBL + serine protease	0			
1166	HAS8124965.1		Vibrio vulnificus	Proteobacteria	631	12	MBL + serine protease	0			
1167	WP_130354990.1		Vibrio vulnificus	Proteobacteria	631	12	MBL + serine protease	0			
1168	WP_103176634.1		Vibrio vulnificus	Proteobacteria	631	12	MBL + serine protease	0			
1169	EGQ8592523.1		Vibrio cholerae	Proteobacteria	631	12	MBL + serine protease	0			
1170	WP_140291658.1		Vibrio parahaemolyticus	Proteobacteria	631	12	MBL + serine protease	0			
1171	WP_191685971.1		Vibrio parahaemolyticus	Proteobacteria	631	12	MBL + serine protease	0			
1172	HAS6330735.1		Vibrio vulnificus	Proteobacteria	631	12	MBL + serine protease	0			
1173	EGR2435113.1		Vibrio cholerae	Proteobacteria	631	12	MBL + serine protease	0			
1174	WP_130193929.1		Vibrio vulnificus	Proteobacteria	631	12	MBL + serine protease	0			
1175	WP_144726627.1		Cobetia crustatorum	Proteobacteria	632	1	MBL + serine protease	1			1709
1176	WP_153390642.1		Pseudomonas sp. FSL R10-1350	Proteobacteria	633	1	MBL + serine protease	1			1724
1177	WP_150743186.1		Pseudomonas fluorescens	Proteobacteria	634	1	MBL + serine protease	1			1727
1178	WP_043614348.1		Chromobacterium violaceum	Proteobacteria	635	1	MBL + serine protease	1			1726
1179	WP_052696090.1		Aeromonas hydrophila	Proteobacteria	636	2	MBL + serine protease	1			1728
1180	AKA19445.1		Aeromonas hydrophila	Proteobacteria	636	2	MBL + serine protease	0			
1181	WP_125117528.1		Aeromonas caviae	Proteobacteria	637	1	MBL + serine protease	1			1726
1182	WP_059853220.1		Burkholderia ubonensis	Proteobacteria	638	2	MBL + serine protease	1			1730
1183	WP_059866732.1		Burkholderia ubonensis	Proteobacteria	638	2	MBL + serine protease	0			

1184	WP_011154079.1		Cupriavidus necator	Proteobacteria	639	1	MBL + serine protease	1			1732
1185	WP_197342546.1		Ralstonia solanacearum	Proteobacteria	640	1	MBL + serine protease	1			1732
1186	WP_053283375.1		Comamonas testosteroni	Proteobacteria	641	1	MBL + serine protease	1			1736
1187	VUJ49867.1		Stenotrophomonas maltophilia	Proteobacteria	642	1	MBL + serine protease	1			1732
1188	OHC21966.1		Pseudomonadales bacterium RIFCSPHIGHO2_02_FULL_60_43	Proteobacteria	643	1	MBL + serine protease	1			1732
1189	WP_090432260.1		Pseudomonas guguanensis	Proteobacteria	644	1	MBL + serine protease	1			1732
1190	WP_033978793.1		Pseudomonas aeruginosa	Proteobacteria	645	4	MBL + serine protease	1			1732
1191	WP_121351934.1		Pseudomonas aeruginosa	Proteobacteria	645	4	MBL + serine protease	0			
1192	WP_058173212.1		Pseudomonas aeruginosa	Proteobacteria	645	4	MBL + serine protease	0			
1193	WP_124138180.1		Pseudomonas aeruginosa	Proteobacteria	645	4	MBL + serine protease	0			
1194	WP_183636306.1		Pseudoxanthomonas sp. OG2	Proteobacteria	646	1	MBL + serine protease	1			1740
1195	WP_164088557.1		Stenotrophomonas maltophilia	Proteobacteria	647	4	MBL + serine protease	1			1736
1196	MBA0275184.1		Stenotrophomonas maltophilia	Proteobacteria	647	4	MBL + serine protease	0			
1197	WP_164106282.1		Stenotrophomonas maltophilia	Proteobacteria	647	4	MBL + serine protease	0			
1198	MBH1659532.1		Stenotrophomonas maltophilia	Proteobacteria	647	4	MBL + serine protease	0			
1199	WP_089451525.1		Burkholderia aenigmatica	Proteobacteria	648	1	MBL + serine protease	1			1740
1200	WP_158537023.1		Dyella sp. AtDHG13	Proteobacteria	649	1	MBL + serine protease	1			1740
1201	KRC70135.1		Achromobacter sp. Root83	Proteobacteria	650	2	MBL + serine protease	1			1736
1202	WP_200964473.1		Achromobacter sp. Root83	Proteobacteria	650	2	MBL + serine protease	0			
1203	WP_004372559.1		Thauera phenylacetica	Proteobacteria	651	1	MBL + serine protease	1			1735
1204	WP_069270851.1		Burkholderia cepacia complex	Proteobacteria	652	2	MBL + serine protease	1			1737
1205	WP_107709301.1		unclassified Paraburkholderia	Proteobacteria	652	2	MBL + serine protease	0			
1206	WP_186052204.1		Burkholderia gladioli	Proteobacteria	653	1	MBL + serine protease	1			1737

1207	WP_090684885.1		Paraburkholderia phenazinium	Proteobacteria	654	1	MBL + serine protease	1			1737
1208	WP_039406935.1		Pandoraea pulmonicola	Proteobacteria	655	1	MBL + serine protease	1			1737
1209	WP_056840938.1		Pseudomonas sp. Leaf127	Proteobacteria	656	1	MBL + serine protease	1			1729
1210	WP_169894251.1		Pseudomonas lundensis	Proteobacteria	657	3	MBL + serine protease	1			1729
1211	NMZ99812.1		Pseudomonas lundensis	Proteobacteria	657	3	MBL + serine protease	0			
1212	WP_175559581.1		Pseudomonas sp. ok266	Proteobacteria	657	3	MBL + serine protease	0			
1213	WP_019819735.1		Pseudomonas	Proteobacteria	658	3	MBL + serine protease	1			1729
1214	WP_052963207.1		Pseudomonas syringae	Proteobacteria	658	3	MBL + serine protease	0			
1215	WP_214917778.1		Pseudomonas fluorescens	Proteobacteria	658	3	MBL + serine protease	0			
1216	WP_196298657.1		Pseudomonas stutzeri	Proteobacteria	659	2	MBL + serine protease	1			1729
1217	KJH82084.1		Pseudomonas stutzeri	Proteobacteria	659	2	MBL + serine protease	0			
1218	WP_122570080.1		Pseudomonas viridiflava	Proteobacteria	660	7	MBL + serine protease	1			1728
1219	WP_029290806.1		Pseudomonas fluorescens	Proteobacteria	660	7	MBL + serine protease	0			
1220	WP_153649623.1		Pseudomonas sp. CFSAN084952	Proteobacteria	660	7	MBL + serine protease	0			
1221	WP_191956326.1		Pseudomonas fluorescens	Proteobacteria	660	7	MBL + serine protease	0			
1222	WP_135196262.1		Pseudomonas fluorescens	Proteobacteria	660	7	MBL + serine protease	0			
1223	WP_191948584.1		Pseudomonas fluorescens	Proteobacteria	660	7	MBL + serine protease	0			
1224	WP_122228043.1		Pseudomonas syringae group genomosp. 3	Proteobacteria	660	7	MBL + serine protease	0			
1225	AZE17455.1		Pseudomonas chlororaphis subsp. aureofaciens	Proteobacteria	661	4	MBL + serine protease	1			1728
1226	WP_036997227.1		Pseudomonas donghuensis	Proteobacteria	661	4	MBL + serine protease	0			
1227	WP_140895079.1		Pseudomonas caspiana	Proteobacteria	661	4	MBL + serine protease	0			
1228	WP_139650086.1		Pseudomonas sp. ICMP22404	Proteobacteria	661	4	MBL + serine protease	0			
1229	WP_059397426.1		Pseudomonas libanensis	Proteobacteria	662	1	MBL + serine protease	1			1728

1230	WP_119713658.1		Pseudomonas putida	Proteobacteria	663	1	MBL + serine protease	1			1723
1231	VVM75134.1		Pseudomonas fluorescens	Proteobacteria	664	2	MBL + serine protease	1			1728
1232	WP_150706722.1		Pseudomonas fluorescens	Proteobacteria	664	2	MBL + serine protease	0			
1233	WP_092032346.1		Marinobacter sp. DSM 26671	Proteobacteria	665	1	Cap4	1			1737
1234	WP_026598226.1		Methylocystaceae	Proteobacteria	666	1	Cap4	1			1743
1235	WP_156043571.1		Mesorhizobium sp. URHC0008	Proteobacteria	667	1	Cap4	1			1747
1236	WP_184511994.1		Tardiphaga robiniae	Proteobacteria	668	1	Cap4	1			1738
1237	WP_212427677.1		Bradyrhizobium manausense	Proteobacteria	669	1	Cap4	1			1742
1238	ATG45096.1		Phaeobacter piscinae	Proteobacteria	670	1	Cap4	1			1742
1239	WP_146545055.1		Reyranella sp. CPCC 100927	Proteobacteria	671	1	Cap4	1			1741
1240	RWN97948.1		Mesorhizobium sp.	Proteobacteria	672	1	Cap4	1			1737
1241	WP_062489999.1		Labrenzia sp. CP4	Proteobacteria	673	1	Cap4	1			1743
1242	WP_063095900.1		Thalassospira	Proteobacteria	674	1	Cap4	1			1737
1243	WP_173998905.1		Agrobacterium rubi	Proteobacteria	675	1	Cap4	1			1753
1244	WP_195177208.1		Mesorhizobium sp. INR15	Proteobacteria	676	1	Cap4	1			1743
1245	WP_166620157.1		Pseudomonas sp. ODNR1LW	Proteobacteria	677	1	Cap4	1			1702
1246	SEC68812.1		Rhizobiales bacterium GAS188	Proteobacteria	678	1	Cap4	1			1737
1247	MBL8559445.1		Hyphomonadaceae bacterium	Proteobacteria	679	1	Cap4	1			1729
1248	PZT92007.1		Sphingomonas sp.	Proteobacteria	680	1	Cap4	1			1761
1249	WP_138921368.1		Novosphingobium pentaromativorans	Proteobacteria	681	2	Cap4	1			1751
1250	AIT79786.1		Novosphingobium pentaromativorans US6-1	Proteobacteria	681	2	Cap4	0			
1251	SDX64003.1		Ruegeria halocynthiae	Proteobacteria	682	2	Cap4	1			1756
1252	WP_143030598.1		Ruegeria halocynthiae	Proteobacteria	682	2	Cap4	0			
1253	WP_205700617.1		Croceicoccus sediminis	Proteobacteria	683	1	Cap4	1			1752
1254	WP_196845870.1		Erythrobacter sp. A30-3	Proteobacteria	684	1	Cap4	1			1752
1255	WP_146235674.1		Novosphingobium sp. GV010	Proteobacteria	685	2	Cap4	1			1758
1256	PYG17066.1		Novosphingobium sp. GV010	Proteobacteria	685	2	Cap4	0			
1257	WP_028967904.1		Sphingomonas phyllosphaerae	Proteobacteria	686	1	Cap4	1			1800
1258	MBG52202.1		Rhodobiaceae bacterium	Proteobacteria	687	1	Cap4	1			1766
1259	WP_067747657.1		Novosphingobium sp. Fuku2-ISO-50	Proteobacteria	688	1	Cap4	1			1772
1260	MBJ7440600.1		Sphingopyxis sp.	Proteobacteria	689	1	Cap4	1			1755

1261	WP_055821167.1		Xanthomonas sp. Leaf131	Proteobacteria	690	1	Cap4	1			1733
1262	WP_045758392.1		Xanthomonas albilineans	Proteobacteria	691	2	Cap4	1			1733
1263	WP_012917082.1		Xanthomonas albilineans	Proteobacteria	691	2	Cap4	0			
1264	WP_147627318.1		Rhodanobacter glycinis	Proteobacteria	692	1	Cap4	1			1726
1265	MBH1816835.1		Stenotrophomonas maltophilia	Proteobacteria	693	1	Cap4	1			1732
1266	WP_107232867.1		unclassified Stenotrophomonas	Proteobacteria	694	1	Cap4	1			1732
1267	WP_167708972.1		Xanthomonas arboricola	Proteobacteria	695	1	Cap4	1			1732
1268	WP_011483552.1		Polaromonas sp. JS666	Proteobacteria	696	1	Cap4	1			1745
1269	WP_154363458.1		Duganella	Proteobacteria	697	1	Cap4	1			1749
1270	WP_069084268.1		Pseudomonas sp. TCU-HL1	Proteobacteria	698	1	Cap4	1			1742
1271	WP_143136616.1		Burkholderia ubonensis	Proteobacteria	699	6	Cap4	1			1743
1272	WP_157653592.1		Burkholderia ubonensis	Proteobacteria	699	6	Cap4	0			
1273	WP_157645336.1		Burkholderia ubonensis	Proteobacteria	699	6	Cap4	0			
1274	KWE58879.1		Burkholderia ubonensis	Proteobacteria	699	6	Cap4	0			
1275	KVM68022.1		Burkholderia ubonensis	Proteobacteria	699	6	Cap4	0			
1276	KWI18341.1		Burkholderia ubonensis	Proteobacteria	699	6	Cap4	0			
1277	VWC77445.1		Burkholderia lata	Proteobacteria	700	4	Cap4	1			1742
1278	WP_175011124.1		Burkholderia lata	Proteobacteria	700	4	Cap4	0			
1279	WP_143331053.1		Burkholderia aenigmatica	Proteobacteria	700	4	Cap4	0			
1280	OXI43822.1		Burkholderia aenigmatica	Proteobacteria	700	4	Cap4	0			
1281	MBU2828950.1		Acidithiobacillus ferriphilus	Proteobacteria	701	1	Cap4	1			1737
1282	WP_031364623.1		Caballeronia sordidicola	Proteobacteria	702	1	Cap4	1			1737
1283	WP_212118715.1		Burkholderia ambifaria	Proteobacteria	703	1	Cap4	1			1736
1284	WP_076885700.1		Burkholderia pseudomallei	Proteobacteria	704	11	Cap4	1			1736
1285	AYX08386.1		Burkholderia pseudomallei	Proteobacteria	704	11	Cap4	0			
1286	KWK43181.1		Burkholderia stagnalis	Proteobacteria	704	11	Cap4	0			
1287	WP_038744349.1		Burkholderia pseudomallei	Proteobacteria	704	11	Cap4	0			
1288	WP_038761575.1		Burkholderia pseudomallei	Proteobacteria	704	11	Cap4	0			
1289	KGX96317.1		Burkholderia pseudomallei A79D	Proteobacteria	704	11	Cap4	0			
1290	WP_147486490.1		Burkholderia pseudomallei	Proteobacteria	704	11	Cap4	0			
1291	KVB49116.1		Burkholderia cepacia	Proteobacteria	704	11	Cap4	0			
1292	WP_059498165.1		Burkholderia cepacia	Proteobacteria	704	11	Cap4	0			
1293	WP_063900478.1		Burkholderia stagnalis	Proteobacteria	704	11	Cap4	0			
1294	KWK53134.1		Burkholderia stagnalis	Proteobacteria	704	11	Cap4	0			

1295	WP_168268582.1		Rhizobium leguminosarum	Proteobacteria	705	1	Cap4	1			1734
1296	HCN61923.1		Pseudomonas sp.	Proteobacteria	706	1	Cap4	1			1722
1297	WP_139273480.1		Pseudomonas syringae group	Proteobacteria	707	3	Cap4	1			1721
1298	RMN50163.1		Pseudomonas syringae pv. apii	Proteobacteria	707	3	Cap4	0			
1299	WP_146048283.1		Pseudomonas syringae	Proteobacteria	707	3	Cap4	0			
1300	WP_163129885.1		Aeromonas allosaccharophila	Proteobacteria	708	1	Cap4	1			1726
1301	WP_137211825.1		Pseudomonas sp. CFBP13509	Proteobacteria	709	1	Cap4	1			1721
1302	WP_070095644.1		Pseudomonas sp. NBRC 111139	Proteobacteria	710	1	Cap4	1			1721
1303	WP_010567742.1		Pseudomonas	Proteobacteria	711	5	Cap4	1			1721
1304	WP_034128272.1		Pseudomonas fluorescens	Proteobacteria	711	5	Cap4	0			
1305	WP_198834785.1		Pseudomonas	Proteobacteria	711	5	Cap4	0			
1306	WP_169878473.1		Pseudomonas proteolytica	Proteobacteria	711	5	Cap4	0			
1307	WP_169868558.1		Pseudomonas rhodesiae	Proteobacteria	711	5	Cap4	0			
1308	WP_052963161.1		Pseudomonas syringae	Proteobacteria	712	1	Cap4	1			1746
1309	WP_149911872.1		Citrobacter freundii	Proteobacteria	713	3	Cap4	1			1747
1310	WP_049045964.1		Klebsiella aerogenes	Proteobacteria	713	3	Cap4	0			
1311	WP_063622297.1		Enterobacter cloacae	Proteobacteria	713	3	Cap4	0			
1312	WP_060615938.1	EcAvs1	Escherichia coli	Proteobacteria	714	48	Cap4	1			1746
1313	WP_187225713.1		Escherichia coli	Proteobacteria	714	48	Cap4	0			
1314	WP_060659204.1		Escherichia coli	Proteobacteria	714	48	Cap4	0			
1315	EAY1318667.1		Salmonella enterica	Proteobacteria	714	48	Cap4	0			
1316	WP_195377718.1		Klebsiella michiganensis	Proteobacteria	714	48	Cap4	0			
1317	WP_097357842.1		Raoultella ornithinolytica	Proteobacteria	714	48	Cap4	0			
1318	WP_115489518.1		Citrobacter freundii	Proteobacteria	714	48	Cap4	0			
1319	WP_096974699.1		Escherichia coli	Proteobacteria	714	48	Cap4	0			
1320	WP_172692825.1		Enterobacter hormaechei	Proteobacteria	714	48	Cap4	0			
1321	WP_188040321.1		Escherichia coli	Proteobacteria	714	48	Cap4	0			
1322	WP_063430228.1		Enterobacter genomosp. O	Proteobacteria	714	48	Cap4	0			
1323	WP_152952918.1		Escherichia coli	Proteobacteria	714	48	Cap4	0			
1324	HAS0829832.1		Enterobacter hormaechei subsp. hoffmannii	Proteobacteria	714	48	Cap4	0			
1325	WP_149463459.1		Pseudocitrobacter sp. 73	Proteobacteria	714	48	Cap4	0			
1326	WP_063860714.1		Enterobacteriaceae	Proteobacteria	714	48	Cap4	0			

1327	WP_104010312.1		Enterobacteriaceae	Proteobacteria	714	48	Cap4	0			
1328	WP_095474296.1		Escherichia coli	Proteobacteria	714	48	Cap4	0			
1329	WP_135731116.1		Klebsiella quasipneumoniae	Proteobacteria	714	48	Cap4	0			
1330	WP_199017102.1		Enterobacter hormaechei	Proteobacteria	714	48	Cap4	0			
1331	WP_105322749.1		Enterobacter hormaechei	Proteobacteria	714	48	Cap4	0			
1332	WP_072643907.1		Enterobacteriaceae	Proteobacteria	714	48	Cap4	0			
1333	HBC9538170.1		Escherichia coli	Proteobacteria	714	48	Cap4	0			
1334	HAJ6529121.1		Escherichia fergusonii	Proteobacteria	714	48	Cap4	0			
1335	WP_058672601.1		Enterobacter cloacae complex	Proteobacteria	714	48	Cap4	0			
1336	WP_126765119.1		Enterobacter hormaechei	Proteobacteria	714	48	Cap4	0			
1337	WP_071032963.1		Klebsiella pneumoniae	Proteobacteria	714	48	Cap4	0			
1338	EFO1274989.1		Escherichia coli	Proteobacteria	714	48	Cap4	0			
1339	WP_158126600.1		Escherichia coli	Proteobacteria	714	48	Cap4	0			
1340	WP_049192781.1		Klebsiella	Proteobacteria	714	48	Cap4	0			
1341	WP_211652631.1		Klebsiella pneumoniae	Proteobacteria	714	48	Cap4	0			
1342	WP_064159938.1		Klebsiella pneumoniae	Proteobacteria	714	48	Cap4	0			
1343	EFE9155070.1		Escherichia coli	Proteobacteria	714	48	Cap4	0			
1344	HAT1691207.1		Raoultella planticola	Proteobacteria	714	48	Cap4	0			
1345	WP_129543086.1		Klebsiella pneumoniae	Proteobacteria	714	48	Cap4	0			
1346	AVJ79923.1		Enterobacter hormaechei subsp. hoffmannii	Proteobacteria	714	48	Cap4	0			
1347	WP_205894327.1		Escherichia coli	Proteobacteria	714	48	Cap4	0			
1348	WP_080819620.1		Klebsiella quasipneumoniae	Proteobacteria	714	48	Cap4	0			
1349	WP_103798286.1		Citrobacter freundii complex sp. CFNIH11	Proteobacteria	714	48	Cap4	0			
1350	EAO8181294.1		Salmonella enterica	Proteobacteria	714	48	Cap4	0			
1351	WP_150434117.1		Brenneria sp. L3-3HA	Proteobacteria	714	48	Cap4	0			
1352	HBD1043271.1		Escherichia coli	Proteobacteria	714	48	Cap4	0			
1353	WP_128317970.1		Klebsiella pneumoniae	Proteobacteria	714	48	Cap4	0			
1354	WP_043875277.1		Klebsiella	Proteobacteria	714	48	Cap4	0			
1355	WP_174329285.1		Citrobacter freundii	Proteobacteria	714	48	Cap4	0			
1356	WP_201507796.1		Klebsiella pneumoniae	Proteobacteria	714	48	Cap4	0			
1357	WP_086581269.1		Enterobacter hormaechei	Proteobacteria	714	48	Cap4	0			
1358	WP_132468237.1		Raoultella planticola	Proteobacteria	714	48	Cap4	0			

1359	WP_115455978.1		Enterobacteriaceae	Proteobacteria	714	48	Cap4	0			
1360	WP_207855862.1		Sulfurisoma sediminicola	Proteobacteria	715	2	Cap4	1			1746
1361	RLJ64612.1		Sulfurisoma sediminicola	Proteobacteria	715	2	Cap4	0			
1362	MBU2804258.1		Acidithiobacillus ferridurans	Proteobacteria	716	1	Cap4	1			1751
1363	WP_130188032.1		Massilia lutea	Proteobacteria	717	1	Cap4	1			1745
1364	OQC17776.1		Alphaproteobacteria bacterium ADurb.Bin100	Proteobacteria	718	1	Cap4	1			1747
1365	WP_026284329.1		Variovorax paradoxus	Proteobacteria	719	3	Cap4	1			1747
1366	WP_172438473.1		Variovorax boronicumulans	Proteobacteria	719	3	Cap4	0			
1367	PBI82848.1		Variovorax boronicumulans	Proteobacteria	719	3	Cap4	0			
1368	WP_165941191.1		unclassified Variovorax	Proteobacteria	720	2	Cap4	1			1747
1369	WP_108136698.1		Variovorax sp. GV051	Proteobacteria	720	2	Cap4	0			
1370	WP_093443769.1		Variovorax sp. 770b2	Proteobacteria	721	1	Cap4	1			1747
1371	WP_110256472.1		Undibacterium pigrum	Proteobacteria	722	1	Cap4	1			1747
1372	CAD6556365.1		Paraburkholderia hiiakae	Proteobacteria	723	2	Cap4	1			1748
1373	WP_201699627.1		Paraburkholderia hiiakae	Proteobacteria	723	2	Cap4	0			
1374	WP_193093315.1		Burkholderia cepacia	Proteobacteria	724	2	Cap4	1			1747
1375	WP_208448988.1		Burkholderia ambifaria	Proteobacteria	724	2	Cap4	0			
1376	WP_099411785.1		Janthinobacterium sp. BJB304	Proteobacteria	725	2	Cap4	1			1747
1377	WP_092715413.1		Janthinobacterium sp. OK676	Proteobacteria	725	2	Cap4	0			
1378	WP_038487621.1		Janthinobacterium agaricidamnosum	Proteobacteria	726	2	Cap4	1			1741
1379	WP_197539820.1		Janthinobacterium agaricidamnosum	Proteobacteria	726	2	Cap4	0			
1380	WP_203570004.1		Deefgea	Proteobacteria	727	1	Cap4	1			1751
1381	SFU56376.1		Nitrosomonas eutropha	Proteobacteria	728	2	Cap4	1			1749
1382	WP_074928179.1		Nitrosomonas eutropha	Proteobacteria	728	2	Cap4	0			
1383	WP_089034438.1		Alcanivorax sp. N3-2A	Proteobacteria	729	1	Cap4	1			1749
1384	WP_124467512.1		Burkholderia cepacia	Proteobacteria	730	2	Cap4	1			1749
1385	WP_124475431.1		Burkholderia cepacia	Proteobacteria	730	2	Cap4	0			
1386	WP_063598041.1		Burkholderia pseudomallei	Proteobacteria	731	5	Cap4	1			1749
1387	WP_060149162.1		Burkholderia stagnalis	Proteobacteria	731	5	Cap4	0			
1388	WP_175684860.1		Burkholderia anthina	Proteobacteria	731	5	Cap4	0			
1389	WP_186098543.1		Burkholderia gladioli	Proteobacteria	731	5	Cap4	0			
1390	WP_059512890.1		Burkholderia cepacia complex	Proteobacteria	731	5	Cap4	0			

1391	WP_126472711.1		Variovorax gossypii	Proteobacteria	732	1	Cap4	1			1749
1392	WP_175167093.1		Achromobacter dolens	Proteobacteria	733	1	Cap4	1			1749
1393	WP_134838848.1		Kerstersia gyiorum	Proteobacteria	734	1	Cap4	1			1749
1394	WP_121088905.1		Pararobbsia silviterrae	Proteobacteria	735	3	Cap4	1			1750
1395	WP_198409295.1		Caballeronia insecticola	Proteobacteria	735	3	Cap4	0			
1396	BAN26507.1		Caballeronia insecticola	Proteobacteria	735	3	Cap4	0			
1397	WP_211206913.1		Thiobacillus denitrificans	Proteobacteria	736	1	Cap4	1			1747
1398	MBO3708833.1		Candidatus Accumulibacter sp.	Proteobacteria	737	1	Cap4	1			1749
1399	WP_053843440.1		Acidovorax avenae	Proteobacteria	738	1	Cap4	1			1747
1400	WP_056423412.1		Acidovorax sp. Leaf84	Proteobacteria	739	1	Cap4	1			1749
1401	TCJ12225.1		Parasulfuritortus cantonensis	Proteobacteria	740	2	Cap4	1			1749
1402	WP_206199428.1		Parasulfuritortus cantonensis	Proteobacteria	740	2	Cap4	0			
1403	WP_018227469.1		Methyloversatilis universalis	Proteobacteria	741	2	Cap4	1			1749
1404	PZU53479.1		Thauera sp.	Proteobacteria	741	2	Cap4	0			
1405	WP_208081744.1		Stenotrophomonas indicatrix	Proteobacteria	742	1	Cap4	1			1749
1406	WP_009049374.1		Pseudomonas chlororaphis	Proteobacteria	743	33	Cap4	1			1749
1407	WP_126632436.1		Pseudomonas aeruginosa	Proteobacteria	743	33	Cap4	0			
1408	WP_124127870.1		Pseudomonas aeruginosa	Proteobacteria	743	33	Cap4	0			
1409	WP_116802142.1		Pseudomonas aeruginosa	Proteobacteria	743	33	Cap4	0			
1410	WP_115290755.1		Pseudomonas mendocina	Proteobacteria	743	33	Cap4	0			
1411	WP_210799159.1		Pseudomonas aeruginosa	Proteobacteria	743	33	Cap4	0			
1412	WP_187682569.1		Pseudomonas aeruginosa	Proteobacteria	743	33	Cap4	0			
1413	WP_186599437.1		Pseudomonas mosselii	Proteobacteria	743	33	Cap4	0			
1414	WP_043107338.1		Pseudomonas aeruginosa	Proteobacteria	743	33	Cap4	0			
1415	WP_134233876.1		Pseudomonas aeruginosa	Proteobacteria	743	33	Cap4	0			
1416	WP_049305718.1		Pseudomonas aeruginosa	Proteobacteria	743	33	Cap4	0			
1417	WP_034065761.1		Pseudomonas aeruginosa	Proteobacteria	743	33	Cap4	0			
1418	WP_213909000.1		Pseudomonas nitrititolerans	Proteobacteria	743	33	Cap4	0			
1419	WP_092413574.1		Pseudomonas sp. NFACC07-1	Proteobacteria	743	33	Cap4	0			
1420	WP_043311240.1		Pseudomonas sp. ML96	Proteobacteria	743	33	Cap4	0			
1421	WP_055134768.1		Pseudomonas mediterranea	Proteobacteria	743	33	Cap4	0			
1422	WP_169408239.1		Pseudomonas sp. gcc21	Proteobacteria	743	33	Cap4	0			
1423	WP_192294507.1		Pseudomonas sp. PDM11	Proteobacteria	743	33	Cap4	0			
1424	WP_017708917.1		Pseudomonas syringae	Proteobacteria	743	33	Cap4	0			

1425	WP_038661162.1		Pseudomonas stutzeri	Proteobacteria	743	33	Cap4	0			
1426	WP_087518660.1		Pseudomonas sp. M30-35	Proteobacteria	743	33	Cap4	0			
1427	WP_210671277.1		Pseudomonas protegens	Proteobacteria	743	33	Cap4	0			
1428	WP_200659604.1		Pseudomonas sp. MF6754	Proteobacteria	743	33	Cap4	0			
1429	WP_003155919.1		Pseudomonas aeruginosa	Proteobacteria	743	33	Cap4	0			
1430	WP_214329604.1		Pseudomonas nitrititolerans	Proteobacteria	743	33	Cap4	0			
1431	WP_096010310.1		Pseudomonas plecoglossicida	Proteobacteria	743	33	Cap4	0			
1432	WP_096133929.1		Pseudomonas syringae	Proteobacteria	743	33	Cap4	0			
1433	WP_125882519.1		Pseudomonas stutzeri	Proteobacteria	743	33	Cap4	0			
1434	WP_186626080.1		Pseudomonas sp. SWRI52	Proteobacteria	743	33	Cap4	0			
1435	MBI8948688.1		Pseudomonas aeruginosa	Proteobacteria	743	33	Cap4	0			
1436	WP_109933077.1		Pseudomonas sp. RW407	Proteobacteria	743	33	Cap4	0			
1437	WP_152220507.1		Pseudomonas sp. SCB32	Proteobacteria	743	33	Cap4	0			
1438	WP_095968244.1		Pseudomonas sp. ACN5	Proteobacteria	743	33	Cap4	0			
1439	WP_026070503.1		Stenotrophomonas	Proteobacteria	744	9	Cap4	1			1749
1440	WP_197656301.1		Stenotrophomonas maltophilia	Proteobacteria	744	9	Cap4	0			
1441	WP_180868648.1		Stenotrophomonas maltophilia	Proteobacteria	744	9	Cap4	0			
1442	VUO15661.1		Stenotrophomonas maltophilia	Proteobacteria	744	9	Cap4	0			
1443	WP_180855123.1		Stenotrophomonas maltophilia	Proteobacteria	744	9	Cap4	0			
1444	WP_208043165.1		Stenotrophomonas maltophilia	Proteobacteria	744	9	Cap4	0			
1445	WP_106548585.1		Stenotrophomonas maltophilia	Proteobacteria	744	9	Cap4	0			
1446	VUJ52758.1		Stenotrophomonas maltophilia	Proteobacteria	744	9	Cap4	0			
1447	WP_197598276.1		Stenotrophomonas maltophilia	Proteobacteria	744	9	Cap4	0			
1448	WP_180877560.1		Stenotrophomonas maltophilia	Proteobacteria	745	1	Cap4	1			1749
1449	VUP55277.1		Stenotrophomonas maltophilia	Proteobacteria	746	1	Cap4	1			1749
1450	NLS96430.1		Planctomycetaceae bacterium	Planctomycetes	747	1	Cap4	1			1759
1451	WP_092051676.1		Planctomicrobium piriforme	Planctomycetes	748	1	Cap4	1			1763
1452	WP_133875762.1		Actinoplanes brasiliensis	Actinobacteria	749	2	Cap4	1			2009
1453	GID29950.1		Actinoplanes brasiliensis	Actinobacteria	749	2	Cap4	0			
1454	WP_148710008.1		Micromonospora sp. M71_S20	Actinobacteria	750	2	Cap4	1			2032
1455	RLK24444.1		Micromonospora sp. M71_S20	Actinobacteria	750	2	Cap4	0			
1456	WP_157509761.1		Leifsonia sp. Root4	Actinobacteria	751	2	Cap4	1			1772
1457	KQW05552.1		Leifsonia sp. Root4	Actinobacteria	751	2	Cap4	0			
1458	WP_134546913.1		Cryobacterium sp. Hb1	Actinobacteria	752	1	Cap4	1			1769
1459	WP_116228919.1		Microbacterium sp. 1294	Actinobacteria	753	1	Cap4	1			1768

1460	WP_116228755.1		Microbacterium sp. 1294	Actinobacteria	754	1	Cap4	1			1763
1461	MBL7177937.1		Desulfobacteraceae bacterium	Proteobacteria	755	1	Cap4	1			1743
1462	HBX80614.1		Propionibacteriaceae bacterium	Actinobacteria	756	2	Cap4	1			1741
1463	HBY22102.1		Propionibacteriaceae bacterium	Actinobacteria	756	2	Cap4	0			
1464	WP_208500369.1		Microbacterium sp. NEAU-LLB	Actinobacteria	757	1	Cap4	1			1765
1465	WP_157811271.1		Microbacterium	Actinobacteria	758	1	Cap4	1			1765
1466	WP_144999359.1		Kocuria indica	Actinobacteria	759	2	Cap4	1			1806
1467	GHD85994.1		Kocuria marina	Actinobacteria	759	2	Cap4	0			
1468	WP_127783653.1		Rhodococcus sp. X156	Actinobacteria	760	1	Cap4	1			1798
1469	WP_020385222.1		Kribbella catacumbae	Actinobacteria	761	1	Cap4	1			1927
1470	WP_174914124.1		Rhodococcus sp. W8901	Actinobacteria	762	1	Cap4	1			1948
1471	WP_094616405.1		unclassified Rhodococcus	Actinobacteria	763	1	Cap4	1			1934
1472	WP_189169553.1		Pilimelia anulata	Actinobacteria	764	1	Cap4	1			1790
1473	WP_089245801.1		Asanoa hainanensis	Actinobacteria	765	2	Cap4	1			1791
1474	SNS98425.1		Asanoa hainanensis	Actinobacteria	765	2	Cap4	0			
1475	WP_189198447.1		Micromonospora fulviviridis	Actinobacteria	766	2	Cap4	1			1795
1476	GGS09964.1		Micromonospora fulviviridis	Actinobacteria	766	2	Cap4	0			
1477	WP_151558490.1		Actinomadura rudentiformis	Actinobacteria	767	1	Cap4	1			1817
1478	WP_157968532.1		Streptomyces geranii	Actinobacteria	768	1	Cap4	1			1803
1479	WP_148666629.1		Streptomyces sp. MOE7	Actinobacteria	769	2	Cap4	1			1801
1480	ARH90039.1		Streptomyces sp. MOE7	Actinobacteria	769	2	Cap4	0			
1481	EJJ01628.1		Rhodococcus sp. JVH1	Actinobacteria	770	2	Cap4	1			1783
1482	WP_157173919.1		Rhodococcus sp. JVH1	Actinobacteria	770	2	Cap4	0			
1483	WP_213570944.1		Rhodococcus sp. USK13	Actinobacteria	771	1	Cap4	1			1785
1484	WP_143981739.1		Nocardia otitidiscaviarum	Actinobacteria	772	1	Cap4	1			1799
1485	GEM29561.1		Nocardia neocaledoniensis NBRC 108232	Actinobacteria	773	3	Cap4	1			1782
1486	PWV68921.1		Nocardia neocaledoniensis	Actinobacteria	773	3	Cap4	0			
1487	WP_146229466.1		Nocardia neocaledoniensis	Actinobacteria	773	3	Cap4	0			
1488	OAM91260.1		Opitutaceae bacterium TSB47	Verrucomicrobia	774	1	Cap4	1			1763
1489	WP_151884660.1		Qipengyuania flava	Proteobacteria	775	1	Cap4	1			1791
1490	WP_174278338.1		Sphingomonas bacterium	Proteobacteria	776	1	Cap4	1			1779

1491	WP_203328824.1		Pseudomonas aeruginosa	Proteobacteria	777	3	Cap4	1			1782
1492	WP_147403425.1		Pseudomonas aeruginosa	Proteobacteria	777	3	Cap4	0			
1493	RIY52184.1		Pseudomonas aeruginosa	Proteobacteria	777	3	Cap4	0			
1494	WP_074070668.1		Rhizobium gallicum	Proteobacteria	778	1	Cap4	1			1777
1495	PVE52267.1		Agrobacterium rhizogenes	Proteobacteria	779	2	Cap4	1			1777
1496	WP_133178076.1		Agrobacterium	Proteobacteria	779	2	Cap4	0			
1497	WP_174173494.1		Agrobacterium rhizogenes	Proteobacteria	780	1	Cap4	1			1777
1498	WP_159992921.1		Roseomonas sp. 18066	Proteobacteria	781	1	Cap4	1			1803
1499	WP_200474803.1		Azospirillum brasilense	Proteobacteria	782	1	Cap4	1			1782
1500	MBI5589210.1		Deltaproteobacteria bacterium	Proteobacteria	783	1	Cap4	1			1782
1501	MBF8275310.1		Candidatus Brocadiaceae bacterium	Planctomycetes	784	1	Cap4	1			1785
1502	WP_177080804.1		Pseudomonas sp. D5002	Proteobacteria	785	1	Cap4	1			1780
1503	WP_103974617.1		Methylovulum psychrotolerans	Proteobacteria	786	1	Cap4	1			1793
1504	WP_157770203.1		Methylomonas koyamae	Proteobacteria	787	1	Cap4	1			1790
1505	WP_013028141.1		Nitrosococcus halophilus	Proteobacteria	788	1	Cap4	1			1792
1506	NQY26166.1		Piscirickettsiaceae bacterium	Proteobacteria	789	1	Cap4	1			1776
1507	WP_096910783.1		Cycloclasticus sp. PY97N	Proteobacteria	790	1	Cap4	1			1775
1508	MAD16476.1		Alteromonadaceae bacterium	Proteobacteria	791	1	Cap4	1			1775
1509	WP_138553439.1		Pseudoalteromonas rubra	Proteobacteria	792	1	Cap4	1			1775
1510	NRA81055.1		Pseudoalteromonas sp.	Proteobacteria	793	1	Cap4	1			1774
1511	WP_196952908.1		Vibrio parahaemolyticus	Proteobacteria	794	1	Cap4	1			1775
1512	WP_047537327.1		Shewanella sp. ZOR0012	Proteobacteria	795	1	Cap4	1			1775
1513	WP_009847105.1		Vibrio sp. MED222	Proteobacteria	796	1	Cap4	1			1773
1514	WP_175249689.1		Vibrio cholerae	Proteobacteria	797	18	Cap4	1			1775
1515	MBE3709814.1		Vibrio parahaemolyticus	Proteobacteria	797	18	Cap4	0			
1516	HBC8235268.1		Escherichia coli	Proteobacteria	797	18	Cap4	0			
1517	WP_185838536.1		Vibrio cholerae	Proteobacteria	797	18	Cap4	0			
1518	WP_015297531.1		Vibrio parahaemolyticus	Proteobacteria	797	18	Cap4	0			
1519	WP_038885213.1		Vibrio harveyi group	Proteobacteria	797	18	Cap4	0			
1520	WP_086484025.1		Vibrio parahaemolyticus	Proteobacteria	797	18	Cap4	0			
1521	WP_165401625.1		Vibrio alginolyticus	Proteobacteria	797	18	Cap4	0			
1522	WP_202648357.1		Vibrio fluvialis	Proteobacteria	797	18	Cap4	0			

1523	WP_064490002.1		Vibrio parahaemolyticus	Proteobacteria	797	18	Cap4	0			
1524	HAT8503607.1		Vibrio vulnificus	Proteobacteria	797	18	Cap4	0			
1525	HAS6056408.1		Vibrio vulnificus	Proteobacteria	797	18	Cap4	0			
1526	WP_164979875.1		Aeromonas veronii	Proteobacteria	797	18	Cap4	0			
1527	WP_167394830.1		Proteus terrae	Proteobacteria	797	18	Cap4	0			
1528	WP_075384547.1		Aeromonas hydrophila	Proteobacteria	797	18	Cap4	0			
1529	WP_207356532.1		Shewanella sp. FJAT-51800	Proteobacteria	797	18	Cap4	0			
1530	HAS8263263.1		Vibrio vulnificus	Proteobacteria	797	18	Cap4	0			
1531	WP_199249720.1		Vibrio metschnikovii	Proteobacteria	797	18	Cap4	0			
1532	WP_028305187.1		Oceanospirillum maris	Proteobacteria	798	1	Cap4	1			1775
1533	MBK83615.1		Gammaproteobacteria bacterium	Proteobacteria	799	1	Cap4	1			1775
1534	WP_064081273.1		Klebsiella pneumoniae	Proteobacteria	800	3	Cap4	1			1774
1535	WP_064323944.1		Klebsiella variicola	Proteobacteria	800	3	Cap4	0			
1536	WP_117373136.1		Klebsiella pneumoniae	Proteobacteria	800	3	Cap4	0			
1537	WP_167557072.1		Aeromonas veronii	Proteobacteria	801	3	Cap4	1			1774
1538	WP_181223363.1		Aeromonas veronii	Proteobacteria	801	3	Cap4	0			
1539	WP_202608043.1		Aeromonas caviae	Proteobacteria	801	3	Cap4	0			
1540	MAO22299.1		Phycisphaerae bacterium	Planctomycetes	802	2	Cap4	1			1780
1541	MBM92737.1		Phycisphaerae bacterium	Planctomycetes	802	2	Cap4	0			
1542	WP_193885991.1		Oculatella sp. LEGE 06141	Cyanobacteria	803	1	Cap4	1			1823
1543	HEY60588.1		Anaerolineae bacterium	Chloroflexi	804	1	Cap4	1			1823
1544	MBN1316411.1		Anaerolineales bacterium	Chloroflexi	805	1	Cap4	1			1817
1545	RLC77770.1		Chloroflexi bacterium	Chloroflexi	806	1	Cap4	1			1826
1546	WP_184027775.1		Deinococcus budaensis	Deinococcus-Thermus	807	1	Cap4	1			1779
1547	MYC61808.1		Caldilineaceae bacterium SB0661_bin_34	Chloroflexi	808	1	Cap4	1			1781
1548	WP_048118692.1		Methanosarcina barkeri	Euryarchaeota	809	1	Cap4	1			1786
1549	KAF5417137.1		ANME-2 cluster archaeon	Euryarchaeota	810	1	Cap4	1			1813
1550	WP_188832140.1		Mucilaginibacter rubeus	Bacteroidetes	811	1	Cap4	1			1757
1551	WP_073079799.1		Chitinophaga jiangningensis	Bacteroidetes	812	1	Cap4	1			1755
1552	WP_121285097.1		Pedobacter alluvionis	Bacteroidetes	813	1	Cap4	1			1758
1553	WP_028788405.1		Terrimonas ferruginea	Bacteroidetes	814	1	Cap4	1			1762
1554	WP_090078373.1		Chryseobacterium frigidisoli	Bacteroidetes	815	1	Cap4	1			1758

1555	WP_198840408.1		Aureibaculum sp. A20	Bacteroidetes	816	1	Cap4	1			1748
1556	WP_173589351.1		Winogradskyella undariae	Bacteroidetes	817	1	Cap4	1			1749
1557	WP_188764654.1		Emticicia aquatilis	Bacteroidetes	818	1	Cap4	1			1767
1558	WP_162386971.1		Spirosoma endbachense	Bacteroidetes	819	1	Cap4	1			1781
1559	WP_100990106.1		Spirosoma pollinicola	Bacteroidetes	820	1	Cap4	1			1777
1560	WP_028524526.1		Runella limosa	Bacteroidetes	821	1	Cap4	1			1765
1561	WP_006045135.1		Prevotella pallens	Bacteroidetes	822	2	Cap4	1			1679
1562	ETK02486.1		Tannerella sp. oral taxon BU063 isolate Cell 2	Bacteroidetes	822	2	Cap4	0			
1563	WP_074833497.1		Prevotella ruminicola	Bacteroidetes	823	1	Cap4	1			1689
1564	MBO7636530.1		Paludibacteraceae bacterium	Bacteroidetes	824	1	Cap4	1			1688
1565	OHC84652.1		Sphingobacteriia bacterium RIFOXYC2_FULL_35_18	Bacteroidetes	825	1	Cap4	1			1761
1566	MBC7389318.1		Opitutaceae bacterium	Verrucomicrobia	826	1	Cap4	1			1777
1567	WP_116787805.1		Flavobacterium sp. CJ74	Bacteroidetes	827	1	Cap4	1			1749
1568	WP_215137445.1		Hymenobacter sp. ISL-91	Bacteroidetes	828	1	Cap4	1			1780
1569	WP_022826295.1		Hymenobacter norwichensis	Bacteroidetes	829	1	Cap4	1			1776
1570	WP_132122229.1		Arundinibacter roseus	Bacteroidetes	830	1	Cap4	1			1770
1571	MBL7965456.1		Prolixibacteraceae bacterium	Bacteroidetes	831	1	Cap4	1			1766
1572	WP_027374293.1		Chryseobacterium sp. UNC8MFCol	Bacteroidetes	832	1	Cap4	1			1759
1573	WP_079464817.1		Chryseobacterium balustinum	Bacteroidetes	833	1	Cap4	1			1767
1574	WP_074235101.1		Epilithonimonas zeae	Bacteroidetes	834	1	Cap4	1			1769
1575	WP_063969917.1		Chryseobacterium sp. FP211-J200	Bacteroidetes	835	1	Cap4	1			1767
1576	WP_115472276.1		Bacteroides fragilis	Bacteroidetes	836	2	Cap4	1			1771
1577	WP_186928601.1		Parabacteroides sp. N37	Bacteroidetes	836	2	Cap4	0			
1578	WP_194183543.1		Faecalibacter sp. WQ 117	Bacteroidetes	837	1	Cap4	1			1770
1579	MBL7780112.1		Saprospiraceae bacterium	Bacteroidetes	838	1	Cap4	1			1771
1580	WP_110359861.1		Marinifilum breve	Bacteroidetes	839	1	Cap4	1			1770
1581	WP_167897991.1		Wenyingzhuangia aestuarii	Bacteroidetes	840	1	Cap4	1			1769
1582	WP_095069390.1		Tenacibaculum jejuense	Bacteroidetes	841	1	Cap4	1			1770
1583	WP_020531911.1		Flexithrix dorotheae	Bacteroidetes	842	1	Cap4	1			1769
1584	WP_116695677.1		Marixanthomonas sp. HN-E44	Bacteroidetes	843	1	Cap4	1			1769

See also Data S4 of Gao et al., Science 377, eabm4096 (2022), which is incorporated herein by reference as if expressed in its entirety herein.
A - Tree Order
B - Cluster Size
C – Architecture

D – Representative
E - Gene 2 Accession (*sequence corrected manually)
G - Gene 2 Length
H- Length excluding NTD

Table 4. Avs2 homologs and domain architectures.

#	Accession	Name	Organism	Phylum	Tree Order	Cluster Size	Archit-ecture	Representa-tive	Length excluding NTD
1	WP_184204013.1		Armatimonas rosea	Armatimonadetes	1	1	Mrr	1	1232
2	MBT3269076.1		Candidatus Poribacteria bacterium	Unclassified	2	2	Mrr	1	1199
3	MBT7809369.1		Candidatus Poribacteria bacterium	Unclassified	2	2	Mrr	0	
4	MBN8655118.1		Anaerolineae bacterium	Chloroflexi	3	1	Mrr	1	1197
5	MBL8164702.1		Anaerolineae bacterium	Chloroflexi	4	1	Mrr	1	1208
6	OGU29812.1		Ignavibacteria bacterium GWA2_35_9	Ignavibacteriae	5	1	Mrr	1	1178
7	PKL64951.1		Methanomicrobiales archaeon HGW-Methanomicrobiales-3	Euryarchaeota	6	1	Mrr	1	1195
8	WP_128706478.1		Acetivibrio mesophilus	Firmicutes	7	1	Mrr	1	1173
9	WP_055665541.1		Desnuesiella massiliensis	Firmicutes	8	1	Mrr	1	1171
10	TAN44437.1		Nitrospirae bacterium	Nitrospirae	9	1	Mrr	1	1164
11	WP_006633034.1		Microcoleus vaginatus	Cyanobacteria	10	1	Mrr	1	1227
12	WP_193923023.1		Plectonema radiosum	Cyanobacteria	11	1	Short	1	1189
13	WP_073607356.1		Phormidium tenue	Cyanobacteria	12	1	Short	1	1185
14	OFZ12991.1		Bdellovibrionales bacterium GWA2_49_15	Proteobacteria	13	1	Mrr	1	1164
15	WP_166807615.1		Chlorobium sp. BLA1	Chlorobi	14	1	Mrr	1	1203
16	WP_137699170.1		Ruminiclostridium herbifermentans	Firmicutes	15	1	Unclassified	1	1181
17	HBB94844.1		Blastocatellia bacterium	Acidobacteria	16	1	Mrr	1	1185
18	PXF50790.1		Methanophagales archaeon	Euryarchaeota	17	1	Mrr	1	1090
19	MBO8152419.1		Candidatus Marinimicrobia bacterium	Candidatus Marinimicrobia	18	1	Mrr	1	1173
20	WP_015329078.1		Chamaesiphon minutus	Cyanobacteria	19	1	Mrr	1	1206
21	WP_083616928.1		Planktothrix paucivesiculata	Cyanobacteria	20	1	Mrr	1	1196
22	HEG44458.1		Phycisphaerales bacterium	Planctomycetes	21	1	Mrr	1	1170
23	PZM84741.1		Candidatus Melainabacteria bacterium	Candidatus Melainabacteria	22	1	Mrr	1	1089

24	TET47484.1		candidate division TA06 bacterium	Unclassified	23	1	Short	1	1169
25	KKS38086.1		Parcubacteria group bacterium GW2011_GWA2_42_14	Unclassified	24	1	Short	1	1183
26	TAK43185.1		Betaproteobacteria bacterium	Proteobacteria	25	1	Short	1	1185
27	MBP9852837.1		Candidatus Saccharibacteria bacterium	Unclassified	26	1	Short	1	1187
28	MBR3329355.1		Candidatus Saccharibacteria bacterium	Unclassified	27	1	Short	1	1184
29	OGH14236.1		Candidatus Levybacteria bacterium RIFCSPHIGHO2_01_FULL_38_26	Unclassified	28	1	Short	1	1193
30	MBK6861599.1		Saprospiraceae bacterium	Bacteroidetes	29	2	Short	1	1198
31	MBK8298752.1		Saprospiraceae bacterium	Bacteroidetes	29	2	Short	0	
32	WP_061469522.1		Legionella pneumophila	Proteobacteria	30	7	Short	1	1198
33	HAU2158094.1		Legionella pneumophila	Proteobacteria	30	7	Short	0	
34	HBD7143980.1		Legionella pneumophila	Proteobacteria	30	7	Short	0	
35	HAU1202360.1		Legionella pneumophila	Proteobacteria	30	7	Short	0	
36	HAT2056597.1		Legionella pneumophila	Proteobacteria	30	7	Short	0	
37	HAT1881270.1		Legionella pneumophila	Proteobacteria	30	7	Short	0	
38	WP_061489791.1		Legionella pneumophila	Proteobacteria	30	7	Short	0	
39	OFZ16958.1		Bdellovibrionales bacterium GWA2_49_15	Proteobacteria	31	1	Short	1	1211
40	RJR51410.1		Desulfobacteraceae bacterium	Proteobacteria	32	1	Short	1	1209
41	WP_121515292.1		Petrotoga sp. HKA.pet.4.5	Thermotogae	33	1	Short	1	1210
42	WP_206847413.1		Methylacidiphilum sp. IT5	Verrucomicrobia	34	1	Short	1	1093
43	NTW89708.1		Candidatus Moranbacteria bacterium	Unclassified	35	1	Short	1	1215
44	OFV66976.1		Candidatus Syntrophoarchaeum butanivorans	Euryarchaeota	36	1	Short	1	1096
45	MBT7087970.1		bacterium	Unclassified	37	1	Short	1	1218
46	MBP9036552.1		Candidatus Cloacimonas sp.	Candidatus Cloacimonetes	38	1	Short	1	1190
47	KKQ46297.1		Candidatus Moranbacteria bacterium GW2011_GWC2_37_8	Unclassified	39	1	Short	1	1178
48	CAD6491829.1		Candidatus Argoarchaeu	Euryarchaeota	40	1	Short	1	1212

			m ethanivorans						
49	MBL7067155.1		Candidatus Marinimicrobia bacterium	Candidatus Marinimicrobia	41	1	Short	1	1212
50	WP_121270861.1		Pedobacter sp. eg	Bacteroidetes	42	1	MBL + serine protease	1	1177
51	MBP7845404.1		Proteobacteria bacterium	Proteobacteria	43	1	Short	1	1168
52	WP_141731324.1		Oligoflexus tunisiensis	Proteobacteria	44	1	Short	1	1149
53	MBK6564685.1		Saprospiraceae bacterium	Bacteroidetes	45	1	MBL + serine protease	1	1203
54	TWV14116.1		Bacteroidaceae bacterium HV4-6-C5C	Bacteroidetes	46	1	MBL (3') + serine protease	1	1071
55	WP_093318426.1		Olivibacter domesticus	Bacteroidetes	47	1	MBL (3') + serine protease	1	1174
56	WP_160845850.1		Pedobacter sp. HMF7647	Bacteroidetes	48	1	MBL (3') + serine protease	1	1176
57	WP_184178022.1		Rhabdobacter roseus	Bacteroidetes	49	1	MBL (3') + serine protease	1	1183
58	WP_211977225.1		Chitinophaga sp. 2R12	Bacteroidetes	50	1	MBL (3') + serine protease	1	1172
59	MBS0646942.1		Verrucomicrobia bacterium	Verrucomicrobia	51	1	MBL (3') + serine protease	1	1176
60	WP_110582307.1		Mucilaginibacter sp. OV119	Bacteroidetes	52	2	MBL (3') + serine protease	1	1173
61	AMR32310.1		Mucilaginibacter sp. PAMC 26640	Bacteroidetes	52	2	MBL (3') + serine protease	0	
62	WP_068399402.1		Pedobacter cryoconitis	Bacteroidetes	53	1	MBL (3') + serine protease	1	1174
63	WP_099766822.1		Chryseobacterium sp. 52	Bacteroidetes	54	1	MBL (3') + serine protease	1	1174
64	WP_066759214.1		Chryseobacterium glaciei	Bacteroidetes	55	1	MBL (3') + serine	1	1174

							protease		
65	WP_161028924.1		Elizabethkingia anophelis	Bacteroidetes	56	1	MBL (3') + serine protease	1	1172
66	WP_078211081.1		Flavobacterium sp. LM5	Bacteroidetes	57	1	MBL (3') + serine protease	1	1174
67	WP_132003417.1		Flavobacterium cellulosilyticum	Bacteroidetes	58	1	MBL (3') + serine protease	1	1175
68	WP_155093336.1		Myroides albus	Bacteroidetes	59	1	MBL (3') + serine protease	1	1169
69	WP_132069940.1		Flavobacterium sp. LB2P22	Bacteroidetes	60	1	MBL (3') + serine protease	1	1170
70	GFD92211.1		Alteromonas sp. KUL154	Proteobacteria	61	1	MBL (3') + serine protease	1	1189
71	WP_106568277.1		Cecembia rubra	Bacteroidetes	62	1	MBL (3') + serine protease	1	1177
72	MBA4852006.1		Emticicia sp. BO119	Bacteroidetes	63	2	MBL (3') + serine protease	1	1181
73	WP_181841839.1		Emticicia sp. BO119	Bacteroidetes	63	2	MBL (3') + serine protease	0	
74	WP_089056834.1		Flavobacterium plurextorum	Bacteroidetes	64	1	MBL (3') + serine protease	1	1166
75	WP_091290868.1		Flavobacterium sp. fv08	Bacteroidetes	65	1	MBL (3') + serine protease	1	1174
76	RZM27270.1		Pedobacter sp.	Bacteroidetes	66	1	MBL (3') + serine protease	1	1188
77	WP_088594557.1		Capnocytophaga endodontalis	Bacteroidetes	67	2	MBL (3') + serine protease	1	1074
78	WP_203083317.1		Capnocytophaga genosp. AHN8471	Bacteroidetes	67	2	MBL (3') + serine	0	

							protease		
79	WP_201113679.1		Dysgonomonas sp. Marseille-P4677	Bacteroidetes	68	1	MBL (3') + serine protease	1	1179
80	WP_146262625.1		Apibacter muscae	Bacteroidetes	69	1	MBL (3') + serine protease	1	1085
81	MBR4488411.1		Bacteroidales bacterium	Bacteroidetes	70	1	Mrr	1	1079
82	MBO7528804.1		Bacteroidales bacterium	Bacteroidetes	71	1	Mrr	1	1156
83	MBQ7382039.1		Paludibacteraceae bacterium	Bacteroidetes	72	1	MBL (3') + serine protease	1	1149
84	MBR5215506.1		Bacteroidales bacterium	Bacteroidetes	73	1	MBL (3') + serine protease	1	1167
85	WP_124317422.1		Butyricimonas faecihominis	Bacteroidetes	74	1	MBL (3') + serine protease	1	1167
86	WP_078805564.1		Prevotella oulorum	Bacteroidetes	75	1	MBL (3') + serine protease	1	1166
87	MBR1889188.1		Alloprevotella sp.	Bacteroidetes	76	1	MBL (3') + serine protease	1	1167
88	MBQ7280899.1		Bacteroidales bacterium	Bacteroidetes	77	1	MBL (3') + serine protease	1	1167
89	WP_122379852.1		Parabacteroides sp. AF19-14	Bacteroidetes	78	3	MBL (3') + serine protease	1	1167
90	WP_195473414.1		Parabacteroides distasonis	Bacteroidetes	78	3	MBL (3') + serine protease	0	
91	WP_121962656.1		Parabacteroides sp. 20_3	Bacteroidetes	78	3	MBL (3') + serine protease	0	
92	WP_147351312.1		Bacteroides caccae	Bacteroidetes	79	5	MBL (3') + serine protease	1	1167
93	WP_195414801.1		Bacteroides caccae	Bacteroidetes	79	5	MBL (3') + serine protease	0	

94	RHD51082.1		Bacteroides caccae	Bacteroidetes	79	5	MBL (3') + serine protease	0	
95	WP_147351501.1		Bacteroides caccae	Bacteroidetes	79	5	MBL (3') + serine protease	0	
96	RHH87490.1		Bacteroides caccae	Bacteroidetes	79	5	MBL (3') + serine protease	0	
97	WP_118527245.1		Bacteroides sp. OF04-15BH	Bacteroidetes	80	1	MBL (3') + serine protease	1	1167
98	WP_087249106.1		Bacteroides sp. An19	Bacteroidetes	81	2	MBL (3') + serine protease	1	1167
99	WP_204402254.1		Bacteroides caecigallinarum	Bacteroidetes	81	2	MBL (3') + serine protease	0	
100	WP_074209401.1		Fibrobacter sp. UWB11	Fibrobacteres	82	1	Short	1	1050
101	MBR3070872.1		Fibrobacter sp.	Fibrobacteres	83	1	Short	1	1051
102	MBR2900575.1		Fibrobacter sp.	Fibrobacteres	84	1	Mrr	1	1054
103	WP_073424064.1		Fibrobacter sp. UWB7	Fibrobacteres	85	1	Mrr	1	1049
104	MBR6125338.1		Candidatus Saccharibacteria bacterium	Unclassified	86	1	Mrr	1	1050
105	WP_166209701.1		Pedobacter sp. HDW13	Bacteroidetes	87	1	MBL + serine protease	1	1025
106	WP_086947456.1		Sphingobacterium sp. JB170	Bacteroidetes	88	1	MBL + serine protease	1	1064
107	OJV54279.1		Bacteroidetes bacterium 43-16	Bacteroidetes	89	1	MBL + serine protease	1	1065
108	WP_115048315.1		Sphingobacterium multivorum	Bacteroidetes	90	1	MBL + serine protease	1	1064
109	WP_090155920.1		Dyadobacter soli	Bacteroidetes	91	1	MBL + serine protease	1	1062
110	WP_120231220.1		Chryseobacterium sp. AG363	Bacteroidetes	92	1	MBL + serine protease	1	1056
111	WP_164111989.1		unclassified Sphingobacterium	Bacteroidetes	93	1	MBL + serine protease	1	1059
112	WP_015336659.1		Maridesulfovibrio	Proteobacteria	94	1	Short	1	1062

			hydrothermalis						
113	WP_107289941.1		Photobacterium kishitanii	Proteobacteria	95	1	MBL + serine protease	1	1060
114	WP_146156695.1		Photobacterium swingsii	Proteobacteria	96	1	MBL + serine protease	1	1063
115	WP_180800799.1		Vibrio parahaemolyticus	Proteobacteria	97	1	MBL + serine protease	1	1091
116	WP_210797670.1		Pseudoaltero monas sp. hOe-66	Proteobacteria	98	1	MBL + serine protease	1	1058
117	WP_123012219.1		Vibrio cholerae	Proteobacteria	99	2	MBL + serine protease	1	1089
118	HAS360655 8.1		Vibrio cholerae	Proteobacteria	99	2	MBL + serine protease	0	
119	WP_142614577.1		Vibrio cholerae	Proteobacteria	100	2	MBL + serine protease	1	1088
120	WP_172775018.1		Vibrio cholerae	Proteobacteria	100	2	MBL + serine protease	0	
121	WP_162791034.1		Aeromonas hydrophila	Proteobacteria	101	2	MBL + serine protease	1	1088
122	WP_167566366.1		Aeromonas veronii	Proteobacteria	101	2	MBL + serine protease	0	
123	WP_193864139.1		Pseudomonas cyclaminis	Proteobacteria	102	1	Mrr	1	1071
124	WP_139214896.1		unclassified Pseudomonas	Proteobacteria	103	1	Mrr	1	1068
125	WP_046216662.1		Paenibacillus wulumuqiensis	Firmicutes	104	1	Mrr	1	1069
126	WP_209126512.1		Dickeya fangzhongdai	Proteobacteria	105	1	Mrr	1	1077
127	WP_143838418.1		Dickeya sp. ws52	Proteobacteria	106	2	Mrr	1	1077
128	WP_039308914.1		Pectobacterium betavasculorum	Proteobacteria	106	2	Mrr	0	
129	WP_000027912.1		Salmonella enterica	Proteobacteria	107	22	Mrr	1	1077
130	WP_182929612.1		unclassified Klebsiella	Proteobacteria	107	22	Mrr	0	
131	WP_136345909.1		Citrobacter freundii	Proteobacteria	107	22	Mrr	0	
132	WP_160179782.1		Citrobacter	Proteobacteria	107	22	Mrr	0	
133	WP_142972441.1		Citrobacter freundii	Proteobacteria	107	22	Mrr	0	
134	HBB672408 1.1		Citrobacter freundii	Proteobacteria	107	22	Mrr	0	
135	WP_191221403.1		Citrobacter freundii	Proteobacteria	107	22	Mrr	0	

136	WP_130715128.1		Citrobacter freundii	Proteobacteria	107	22	Mrr	0	
137	EAP2788425.1		Salmonella enterica	Proteobacteria	107	22	Mrr	0	
138	WP_167610546.1		Cedecea colo	Proteobacteria	107	22	Mrr	0	
139	EFQ7997788.1		Salmonella enterica	Proteobacteria	107	22	Mrr	0	
140	EAU9996656.1		Salmonella enterica	Proteobacteria	107	22	Mrr	0	
141	EEH8831737.1		Salmonella enterica	Proteobacteria	107	22	Mrr	0	
142	MFK51822.1		Salmonella enterica	Proteobacteria	107	22	Mrr	0	
143	WP_080152666.1		Salmonella enterica	Proteobacteria	107	22	Mrr	0	
144	EDE9521894.1		Salmonella enterica	Proteobacteria	107	22	Mrr	0	
145	EBR2488590.1		Salmonella enterica	Proteobacteria	107	22	Mrr	0	
146	EEO8783242.1		Salmonella enterica	Proteobacteria	107	22	Mrr	0	
147	EDW2494221.1		Salmonella enterica subsp. enterica serovar Oranienburg	Proteobacteria	107	22	Mrr	0	
148	WP_140198326.1		Salmonella enterica	Proteobacteria	107	22	Mrr	0	
149	SUG17267.1		Salmonella enterica subsp. arizonae	Proteobacteria	107	22	Mrr	0	
150	EFV4529721.1		Salmonella enterica	Proteobacteria	107	22	Mrr	0	
151	WP_079974080.1		Salmonella enterica	Proteobacteria	108	9	Mrr	1	1077
152	EDV4531620.1		Salmonella enterica subsp. enterica	Proteobacteria	108	9	Mrr	0	
153	EBW7590374.1		Salmonella enterica subsp. salamae serovar Sofia	Proteobacteria	108	9	Mrr	0	
154	WP_079797565.1		Salmonella enterica	Proteobacteria	108	9	Mrr	0	
155	ECF6014919.1		Salmonella enterica subsp. salamae	Proteobacteria	108	9	Mrr	0	
156	EAA4081384.1		Salmonella enterica subsp. salamae	Proteobacteria	108	9	Mrr	0	
157	AFO66380.1		Salmonella enterica subsp. salamae serovar Sofia	Proteobacteria	108	9	Mrr	0	
158	EEP9809207.1		Salmonella enterica subsp. diarizonae	Proteobacteria	108	9	Mrr	0	
159	EBU7538521.1		Salmonella enterica subsp. salamae serovar Sofia	Proteobacteria	108	9	Mrr	0	

160	WP_080150299.1		Salmonella enterica	Proteobacteria	109	15	Mrr	1	1077
161	EBF8291568.1		Salmonella enterica subsp. enterica	Proteobacteria	109	15	Mrr	0	
162	ECE8654369.1		Salmonella enterica subsp. enterica	Proteobacteria	109	15	Mrr	0	
163	EDA3140102.1		Salmonella enterica subsp. enterica serovar Joal	Proteobacteria	109	15	Mrr	0	
164	EBR5127163.1		Salmonella enterica	Proteobacteria	109	15	Mrr	0	
165	EBP1709950.1		Salmonella enterica	Proteobacteria	109	15	Mrr	0	
166	HAU6761898.1		Salmonella enterica subsp. enterica serovar Pasing	Proteobacteria	109	15	Mrr	0	
167	EDX0369977.1		Salmonella enterica	Proteobacteria	109	15	Mrr	0	
168	EHN2719530.1		Salmonella enterica subsp. enterica serovar Essen	Proteobacteria	109	15	Mrr	0	
169	EHG6571002.1		Salmonella enterica subsp. enterica serovar Pasing	Proteobacteria	109	15	Mrr	0	
170	EDQ9753196.1		Salmonella enterica subsp. enterica	Proteobacteria	109	15	Mrr	0	
171	EAA7108094.1		Salmonella enterica subsp. enterica	Proteobacteria	109	15	Mrr	0	
172	EBW5537325.1		Salmonella enterica subsp. enterica serovar Pasing	Proteobacteria	109	15	Mrr	0	
173	EEP1439491.1		Salmonella enterica subsp. enterica	Proteobacteria	109	15	Mrr	0	
174	EDH9166062.1		Salmonella enterica subsp. enterica serovar Fallowfield	Proteobacteria	109	15	Mrr	0	
175	WP_194968293.1		Methylobacter sp. BlB1	Proteobacteria	110	1	Mrr	1	1082
176	WP_127029120.1		Candidatus Methylobacter oryzae	Proteobacteria	111	1	Mrr	1	1081
177	WP_031438285.1		Methylobacter tundripaludum	Proteobacteria	112	1	Mrr	1	1080

178	WP_027390249.1		Chrysiogenes arsenatis	Chrysiogenetes	113	2	Mrr	1	1074
179	WP_012970053.1		Allochromatium vinosum	Proteobacteria	113	2	Mrr	0	
180	WP_098263543.1		Bacillus sp. AFS006103	Firmicutes	114	1	Mrr	1	1077
181	WP_215164210.1		Paenibacillus sp. ISL-20	Firmicutes	115	1	Mrr	1	1068
182	MBS3976254.1		Syntrophomonadaceae bacterium	Firmicutes	116	1	Mrr	1	1076
183	WP_211598614.1		Pasteurella atlantica	Proteobacteria	117	1	Mrr	1	1062
184	WP_210541534.1		Psychrobacter sp. K31L	Proteobacteria	118	1	HNH	1	1068
185	WP_079690942.1		Psychrobacter sp. DAB_AL43B	Proteobacteria	119	1	HNH	1	1069
186	WP_147223093.1		Psychrobacter frigidicola	Proteobacteria	120	1	HNH	1	1069
187	WP_100125817.1		Snodgrassella alvi	Proteobacteria	121	1	Mrr	1	1071
188	WP_065737360.1		Gilliamella apicola	Proteobacteria	122	1	Mrr	1	1069
189	WP_132022014.1		Bisgaardia hudsonensis	Proteobacteria	123	1	Mrr	1	1073
190	WP_201532033.1		Psychrobacter sp. LFX-11D	Proteobacteria	124	1	Mrr	1	1070
191	WP_031964135.1		Acinetobacter baumannii	Proteobacteria	125	4	Mrr	1	1080
192	WP_031976725.1		Acinetobacter baumannii	Proteobacteria	125	4	Mrr	0	
193	WP_086250535.1		Acinetobacter baumannii	Proteobacteria	125	4	Mrr	0	
194	WP_114172218.1		Acinetobacter baumannii	Proteobacteria	125	4	Mrr	0	
195	WP_045125471.1		Acinetobacter calcoaceticus	Proteobacteria	126	1	Mrr	1	1080
196	WP_209048102.1		Acinetobacter nosocomialis	Proteobacteria	127	1	Mrr	1	1107
197	WP_208753229.1		Acinetobacter pittii	Proteobacteria	128	1	Mrr	1	1081
198	WP_204084992.1		Acinetobacter pittii	Proteobacteria	129	2	Mrr	1	1107
199	HAV5769560.1		Acinetobacter baumannii	Proteobacteria	129	2	Mrr	0	
200	WP_159375831.1		Moraxella osloensis	Proteobacteria	130	1	Mrr	1	1069
201	WP_083106294.1		Moraxella osloensis	Proteobacteria	131	2	Mrr	1	1072
202	OBX57766.1		Moraxella osloensis	Proteobacteria	131	2	Mrr	0	
203	WP_151052044.1		Neisseria zalophi	Proteobacteria	132	1	Mrr	1	1071
204	WP_003759601.1		Neisseria sicca	Proteobacteria	133	1	Mrr	1	1069
205	WP_133542381.1		Mesocricetibacter intestinalis	Proteobacteria	134	1	Mrr	1	1068
206	WP_005821035.1		Actinobacillus minor	Proteobacteria	135	1	Mrr	1	1068
207	WP_021111871.1		Glaesserella parasuis	Proteobacteria	136	2	Mrr	1	1068
208	WP_075606270.1		Glaesserella parasuis	Proteobacteria	136	2	Mrr	0	
209	WP_154103593.1		Eikenella corrodens	Proteobacteria	137	2	Mrr	1	1068
210	RKV84069.1		Neisseria sp.	Proteobacteria	137	2	Mrr	0	

211	WP_016534437.1		Pasteurella multocida	Proteobacteria	138	5	Mrr	1	1070
212	WP_170449591.1		Pasteurella multocida	Proteobacteria	138	5	Mrr	0	
213	WP_099821844.1		Pasteurella multocida	Proteobacteria	138	5	Mrr	0	
214	WP_079157767.1		Pasteurella multocida	Proteobacteria	138	5	Mrr	0	
215	WP_170356796.1		Pasteurella multocida	Proteobacteria	138	5	Mrr	0	
216	WP_115097899.1		Pasteurella multocida	Proteobacteria	139	2	Mrr	1	1070
217	WP_096743303.1		Pasteurella multocida	Proteobacteria	139	2	Mrr	0	
218	MBS6017565.1		Haemophilus parainfluenzae	Proteobacteria	140	1	Mrr	1	1068
219	WP_025342502.1		Mannheimia varigena	Proteobacteria	141	1	Mrr	1	1071
220	WP_019520715.1		Moraxella boevrei	Proteobacteria	142	1	Mrr	1	1062
221	WP_006249440.1		Mannheimia haemolytica	Proteobacteria	143	8	Mrr	1	1068
222	WP_115262428.1		Mannheimia haemolytica	Proteobacteria	143	8	Mrr	0	
223	WP_061887347.1		Mannheimia haemolytica	Proteobacteria	143	8	Mrr	0	
224	SQE31487.1		Mannheimia haemolytica	Proteobacteria	143	8	Mrr	0	
225	WP_040080867.1		Mannheimia haemolytica	Proteobacteria	143	8	Mrr	0	
226	WP_142778268.1		Mannheimia haemolytica	Proteobacteria	143	8	Mrr	0	
227	WP_147009272.1		Mannheimia haemolytica	Proteobacteria	143	8	Mrr	0	
228	WP_147008897.1		Mannheimia haemolytica	Proteobacteria	143	8	Mrr	0	
229	AIS31580.1		Methanobacterium formicicum	Euryarchaeota	144	2	Mrr	1	1073
230	WP_048084866.1		Methanobacterium formicicum	Euryarchaeota	144	2	Mrr	0	
231	MBN8545544.1		Ignavibacteria bacterium	Ignavibacteriae	145	1	Mrr	1	1068
232	MBR4447685.1		Methanobrevibacter sp.	Euryarchaeota	146	1	CMP hydrolase	1	1100
233	WP_094549827.1		Petroclostridium xylanilyticum	Firmicutes	147	1	Unknown-2	1	1057
234	ESL13878.1		Firmicutes bacterium ASF500	Firmicutes	148	1	Unknown-2	1	1072
235	MBQ3194610.1		Oscillospiraceae bacterium	Firmicutes	149	1	Unknown-2	1	1067
236	WP_091752299.1		Propionispora vibrioides	Firmicutes	150	1	Unknown-2	1	1059
237	NBI89024.1		Lachnospiraceae bacterium	Firmicutes	151	1	Unknown-2	1	1074
238	WP_125136632.1		Petrocella atlantisensis	Firmicutes	152	1	SIR2	1	1057
239	WP_021362992.1		Clostridioides difficile	Firmicutes	153	1	SIR2	1	1061
240	WP_021392315.1		Clostridioides difficile	Firmicutes	154	12	SIR2	1	1061
241	HBF4875342.1		Clostridioides difficile	Firmicutes	154	12	SIR2	0	
242	WP_095891050.1		Clostridioides difficile	Firmicutes	154	12	SIR2	0	

243	WP_021373721.1		Clostridioides difficile	Firmicutes	154	12	SIR2	0	
244	MBH7278500.1		Clostridioides difficile	Firmicutes	154	12	SIR2	0	
245	WP_077716309.1		Clostridioides difficile	Firmicutes	154	12	SIR2	0	
246	HBF9625169.1		Clostridioides difficile	Firmicutes	154	12	SIR2	0	
247	HBF2172268.1		Clostridioides difficile	Firmicutes	154	12	SIR2	0	
248	HBG4111761.1		Clostridioides difficile	Firmicutes	154	12	SIR2	0	
249	HBF0793835.1		Clostridioides difficile	Firmicutes	154	12	SIR2	0	
250	HBF8210002.1		Clostridioides difficile	Firmicutes	154	12	SIR2	0	
251	EGT385072 2.1		Clostridioides difficile	Firmicutes	154	12	SIR2	0	
252	MBE6034141.1		Clostridiales bacterium	Firmicutes	155	1	SIR2	1	1059
253	NFJ04174.1		Clostridium botulinum	Firmicutes	156	3	SIR2	1	1060
254	NFJ53823.1		Clostridium botulinum	Firmicutes	156	3	SIR2	0	
255	NFD75934.1		Clostridium botulinum	Firmicutes	156	3	SIR2	0	
256	WP_045144665.1		Clostridium butyricum	Firmicutes	157	1	SIR2	1	1059
257	WP_016646431.1		Campylobacter ureolyticus	Proteobacteria	158	1	Mrr	1	1051
258	WP_124852330.1		Campylobacter rectus	Proteobacteria	159	2	Mrr	1	1051
259	WP_004320557.1		Campylobacter rectus	Proteobacteria	159	2	Mrr	0	
260	WP_112870338.1		Francisella adeliensis	Proteobacteria	160	1	Mrr	1	1066
261	WP_151951121.1		Aliarcobacter butzleri	Proteobacteria	161	1	Mrr	1	1092
262	WP_052746079.1		Sulfurovum lithotrophicum	Proteobacteria	162	1	Mrr	1	1094
263	NUN09210.1		Ignavibacteriaceae bacterium	Ignavibacteriae	163	1	Mrr	1	1057
264	WP_184178785.1		Rhabdobacter roseus	Bacteroidetes	164	1	Mrr	1	1131
265	OFX25212.1		Bacteroidetes bacterium GWA2_31_9	Bacteroidetes	165	1	Mrr	1	1077
266	WP_010853022.1		Lunatimonas lonarensis	Bacteroidetes	166	2	Mrr	1	1077
267	WP_025764605.1		Dyadobacter tibetensis	Bacteroidetes	166	2	Mrr	0	
268	WP_072408755.1		Chryseobacterium limigenitum	Bacteroidetes	167	1	Mrr	1	1092
269	WP_039372698.1		Chryseobacterium taiwanense	Bacteroidetes	168	1	Mrr	1	1104
270	SKC03095.1		Chryseobacterium balustinum	Bacteroidetes	169	2	Mrr	1	1104
271	WP_079466659.1		Chryseobacterium balustinum	Bacteroidetes	169	2	Mrr	0	
272	WP_078795687.1		Elizabethkingia miricola	Bacteroidetes	170	1	Mrr	1	1081
273	MBP6238444.1		Saprospiraceae bacterium	Bacteroidetes	171	1	Mrr	1	1099
274	WP_077492490.1		Porphyromonas gingivalis	Bacteroidetes	172	2	Mrr	1	1082

275	WP_099779211.1		Porphyromonas gingivalis	Bacteroidetes	172	2	Mrr	0	
276	WP_018359964.1		Porphyromonas macacae	Bacteroidetes	173	1	Mrr	1	1078
277	WP_114237173.1		Paracnuella aquatica	Bacteroidetes	174	1	MBL + serine protease	1	1099
278	WP_198850042.1		Alicyclobacillus sp. SO9	Firmicutes	175	2	Mrr	1	1058
279	QQE77268.1		Alicyclobacillus sp. SO9	Firmicutes	175	2	Mrr	0	
280	WP_090922622.1		Paenibacillus polysaccharolyticus	Firmicutes	176	1	Mrr	1	1105
281	WP_085516251.1		Marivirga sericea	Bacteroidetes	177	1	Mrr	1	1089
282	WP_109674963.1		Dyadobacter jejuensis	Bacteroidetes	178	1	Mrr	1	1092
283	WP_163286696.1		Cryomorpha ignava	Bacteroidetes	179	1	Mrr	1	1091
284	WP_142712813.1		Aliifodinibius sediminis	Balneolaeota	180	1	Mrr	1	1088
285	WP_143164815.1		Chryseolinea serpens	Bacteroidetes	181	1	Mrr	1	1127
286	WP_202008441.1		Chryseolinea sp. Jin1	Bacteroidetes	182	1	Mrr	1	1149
287	WP_188220213.1		Flavobacterium pokkalii	Bacteroidetes	183	1	Mrr	1	1084
288	WP_129752559.1		Flavobacterium	Bacteroidetes	184	2	Mrr	1	1084
289	WP_026985935.1		Flavobacterium sp. URHB0058	Bacteroidetes	184	2	Mrr	0	
290	WP_062184002.1		Dysgonomonas macrotermitis	Bacteroidetes	185	1	Mrr	1	1081
291	WP_090986291.1		Pedobacter suwonensis	Bacteroidetes	186	2	Mrr	1	1085
292	WP_184471351.1		Pedobacter sp. AK013	Bacteroidetes	186	2	Mrr	0	
293	WP_146740037.1		Legionella quinlivanii	Proteobacteria	187	2	Mrr	1	1051
294	RAP37143.1		Legionella quinlivanii	Proteobacteria	187	2	Mrr	0	
295	MBR2031995.1		Alistipes sp.	Bacteroidetes	188	1	Mrr	1	1095
296	MBE6320004.1		Bacteroidales bacterium	Bacteroidetes	189	1	Mrr	1	1120
297	MBR6363079.1		Bacteroidales bacterium	Bacteroidetes	190	1	Mrr	1	1070
298	WP_004364085.1		Prevotella nigrescens	Bacteroidetes	191	6	Mrr	1	1055
299	MBF1622244.1		Prevotella sp.	Bacteroidetes	191	6	Mrr	0	
300	MBF1596946.1		Prevotella sp.	Bacteroidetes	191	6	Mrr	0	
301	MBF1614297.1		Prevotella sp.	Bacteroidetes	191	6	Mrr	0	
302	WP_013264603.1		Prevotella melaninogenica	Bacteroidetes	191	6	Mrr	0	
303	WP_131752224.1		Bacteroides heparinolyticus	Bacteroidetes	191	6	Mrr	0	
304	WP_091957097.1		Prevotella sp. tc2-28	Bacteroidetes	192	1	Mrr	1	1059
305	MBR3699818.1		Bacteroidales bacterium	Bacteroidetes	193	1	Mrr	1	1060
306	WP_092073540.1		Prevotella sp. tf2-5	Bacteroidetes	194	1	Mrr	1	1060

307	WP_130059400.1		Bacteroides salyersiae	Bacteroidetes	195	1	Mrr	1	1060
308	WP_053873644.1		Bacteroides fragilis	Bacteroidetes	196	1	Mrr	1	1060
309	WP_195411903.1		Bacteroides eggerthii	Bacteroidetes	197	3	Mrr	1	1059
310	WP_117678353.1		Phocaeicola vulgatus	Bacteroidetes	197	3	Mrr	0	
311	WP_186968348.1		Bacteroides sp. M27	Bacteroidetes	197	3	Mrr	0	
312	WP_005816350.1		Bacteroides	Bacteroidetes	198	5	Mrr	1	1059
313	WP_061438424.1		Bacteroides intestinalis	Bacteroidetes	198	5	Mrr	0	
314	WP_032566357.1		Bacteroides fragilis	Bacteroidetes	198	5	Mrr	0	
315	WP_152635413.1		Bacteroides stercoris	Bacteroidetes	198	5	Mrr	0	
316	WP_065538134.1		Bacteroides caecimuris	Bacteroidetes	198	5	Mrr	0	
317	WP_195654942.1		Bacteroides sp. 1001136B_160425_E2	Bacteroidetes	199	1	Mrr	1	1059
318	WP_203257071.1		Hyunsoonleella sp. HU1-3	Bacteroidetes	200	1	Mrr	1	1079
319	WP_205431341.1		Flavobacterium arcticum	Bacteroidetes	201	2	Mrr	1	1084
320	AXG75196.1		Flavobacterium arcticum	Bacteroidetes	201	2	Mrr	0	
321	MBI2417951.1		Ignavibacteriales bacterium	Ignavibacteriae	202	1	Mrr	1	1077
322	WP_188599262.1		Polaribacter pacificus	Bacteroidetes	203	2	Mrr	1	1078
323	GGG97381.1		Polaribacter pacificus	Bacteroidetes	203	2	Mrr	0	
324	WP_211826043.1		Kistimonas asteriae	Proteobacteria	204	1	Mrr	1	1073
325	WP_181447818.1		unclassified Pseudoaltero monas	Proteobacteria	205	2	Mrr	1	1073
326	WP_175083478.1		Pseudoaltero monas sp. 1701	Proteobacteria	205	2	Mrr	0	
327	WP_155697270.1		Aliivibrio fischeri	Proteobacteria	206	1	Mrr	1	1074
328	WP_160337781.1		Elizabethkingia anophelis	Bacteroidetes	207	3	Unknown-1	1	1073
329	WP_061889417.1		Elizabethkingia anophelis	Bacteroidetes	207	3	Unknown-1	0	
330	EGT5873275.1		Elizabethkingia anophelis	Bacteroidetes	207	3	Unknown-1	0	
331	WP_166925154.1		Flavobacterium sp. JXAS1	Bacteroidetes	208	1	Unknown-1	1	1076
332	WP_056191773.1		Flavobacterium sp. Root420	Bacteroidetes	209	1	Unknown-1	1	1076
333	WP_165763210.1		Flavobacterium sp. AJR	Bacteroidetes	210	2	Unknown-1	1	1077
334	OUL62002.1		Flavobacterium sp. AJR	Bacteroidetes	210	2	Unknown-1	0	
335	WP_055092545.1		Flavobacterium aquidurense	Bacteroidetes	211	1	Unknown-1	1	1077
336	WP_198931823.1		Labilibacter marinus	Bacteroidetes	212	1	Unknown-1	1	1076
337	OFX65628.1		Bacteroidetes bacterium GWC2_40_13	Bacteroidetes	213	1	Unknown-1	1	1163

338	WP_199163413.1		Parabacteroides provencensis	Bacteroidetes	214	1	Transmembrane	1	1075
339	WP_025640271.1		Schnuerera ultunensis	Firmicutes	215	2	Unknown-1	1	1070
340	WP_034602223.1		Clostridiisalibacter paucivorans	Firmicutes	215	2	Unknown-1	0	
341	WP_194974215.1		Aquiflexum lacus	Bacteroidetes	216	1	Transmembrane	1	1071
342	NVK53669.1		Flavobacteriaceae bacterium	Bacteroidetes	217	1	Transmembrane	1	1076
343	HAR74186.1		Flavobacteriaceae bacterium	Bacteroidetes	218	1	Unknown-1	1	1075
344	WP_179218129.1		Dokdonia pacifica	Bacteroidetes	219	1	Unknown-1	1	1076
345	WP_208599573.1		Myroides odoratimimus	Bacteroidetes	220	2	Transmembrane	1	1076
346	ALU28251.1		Myroides odoratimimus	Bacteroidetes	220	2	Transmembrane	0	
347	WP_207509578.1		Telluribacter humicola	Bacteroidetes	221	1	Unknown-1	1	1163
348	WP_013552761.1		Cellulophaga algicola	Bacteroidetes	222	1	Unknown-1	1	1161
349	WP_184167923.1		Flavobacterium	Bacteroidetes	223	1	Unknown-1	1	1162
350	WP_173857103.1		Flavobacterium sp. 7E	Bacteroidetes	224	2	Unknown-1	1	1161
351	WP_192500866.1		Flavobacterium sp. PL002	Bacteroidetes	224	2	Unknown-1	0	
352	WP_008315274.1		Wohlfahrtiimonas chitiniclastica	Proteobacteria	225	1	MBL + serine protease	1	1201
353	WP_109201019.1		Ignatzschineria cameli	Proteobacteria	226	1	MBL + serine protease	1	1176
354	ENW24071.1		Acinetobacter lwoffii NCTC 5866 = CIP 64.10 = NIPH 512	Proteobacteria	227	1	Unknown-1	1	1170
355	WP_017050185.1		Vibrio anguillarum	Proteobacteria	228	20	Unknown-1	1	1167
356	WP_025767124.1		Vibrio parahaemolyticus	Proteobacteria	228	20	Unknown-1	0	
357	WP_053316851.1		Vibrio parahaemolyticus	Proteobacteria	228	20	Unknown-1	0	
358	WP_029809946.1		Vibrio parahaemolyticus	Proteobacteria	228	20	Unknown-1	0	
359	WP_064485911.1		Vibrio parahaemolyticus	Proteobacteria	228	20	Unknown-1	0	
360	WP_069212231.1		Vibrio anguillarum	Proteobacteria	228	20	Unknown-1	0	
361	WP_140167202.1		Vibrio parahaemolyticus	Proteobacteria	228	20	Unknown-1	0	
362	HAV136536 6.1		Vibrio parahaemolyticus	Proteobacteria	228	20	Unknown-1	0	

363	WP_213962086.1		Vibrio alginolyticus	Proteobacteria	228	20	Unknown-1	0	
364	WP_213880610.1		Vibrio alginolyticus	Proteobacteria	228	20	Unknown-1	0	
365	WP_086484041.1		Vibrio parahaemolyticus	Proteobacteria	228	20	Unknown-1	0	
366	WP_077284313.1		Vibrio	Proteobacteria	228	20	Unknown-1	0	
367	WP_102350412.1		Vibrio cyclitrophicus	Proteobacteria	228	20	Unknown-1	0	
368	EHD225014 3.1		Vibrio vulnificus	Proteobacteria	228	20	Unknown-1	0	
369	EGQ929051 4.1		Vibrio vulnificus	Proteobacteria	228	20	Unknown-1	0	
370	MBM51084 38.1		Vibrio parahaemolyticus	Proteobacteria	228	20	Unknown-1	0	
371	WP_208137768.1		Shewanella algae	Proteobacteria	228	20	Unknown-1	0	
372	WP_065099618.1		Vibrio cyclitrophicus	Proteobacteria	228	20	Unknown-1	0	
373	WP_208160343.1		Shewanella algae	Proteobacteria	228	20	Unknown-1	0	
374	WP_064354202.1		Vibrio alginolyticus	Proteobacteria	228	20	Unknown-1	0	
375	WP_182676551.1		Pseudoaltero monas sp. SR45-4	Proteobacteria	229	1	Unknown-1	1	1164
376	WP_065297001.1		Vibrio natriegens	Proteobacteria	230	1	Unknown-1	1	1149
377	WP_026024187.1		Enterovibrio norvegicus	Proteobacteria	231	1	Unknown-1	1	1163
378	WP_015958822.1		Klebsiella pneumoniae	Proteobacteria	232	16	Unknown-1	1	1163
379	WP_205920380.1		Enterobacter hormaechei	Proteobacteria	232	16	Unknown-1	0	
380	WP_108458028.1		Klebsiella pneumoniae	Proteobacteria	232	16	Unknown-1	0	
381	WP_114507000.1		Klebsiella pneumoniae	Proteobacteria	232	16	Unknown-1	0	
382	WP_211801581.1		Klebsiella michiganensis	Proteobacteria	232	16	Unknown-1	0	
383	WP_117260467.1		Klebsiella pneumoniae	Proteobacteria	232	16	Unknown-1	0	
384	WP_101968981.1		Klebsiella pneumoniae	Proteobacteria	232	16	Unknown-1	0	
385	WP_064141464.1		Klebsiella pneumoniae	Proteobacteria	232	16	Unknown-1	0	
386	WP_045346117.1		Enterobacter hormaechei	Proteobacteria	232	16	Unknown-1	0	
387	WP_181627786.1		Citrobacter sp. RHB35-C17	Proteobacteria	232	16	Unknown-1	0	
388	WP_058672803.1		Enterobacter hormaechei	Proteobacteria	232	16	Unknown-1	0	
389	WP_048226709.1		Enterobacter quasiroggenkampii	Proteobacteria	232	16	Unknown-1	0	
390	WP_080884843.1		Klebsiella pneumoniae	Proteobacteria	232	16	Unknown-1	0	
391	MBC42521 07.1		Klebsiella pneumoniae	Proteobacteria	232	16	Unknown-1	0	
392	MBK27335 40.1		Klebsiella pneumoniae	Proteobacteria	232	16	Unknown-1	0	
393	WP_087783528.1		Klebsiella pneumoniae	Proteobacteria	232	16	Unknown-1	0	
394	WP_086360599.1		Gilliamella apis	Proteobacteria	233	1	Unknown-1	1	1164

395	WP_125729330.1		Aeromonas salmonicida	Proteobacteria	234	1	Unknown-1	1	1164
396	MBG3153147.1		Proteus mirabilis	Proteobacteria	235	1	Unknown-1	1	1161
397	WP_119881916.1		Pectobacterium carotovorum	Proteobacteria	236	1	Unknown-1	1	1164
398	WP_172379316.1		Vibrio sp. Vb339	Proteobacteria	237	2	Unknown-1	1	1163
399	WP_108117333.1		Vibrio splendidus	Proteobacteria	237	2	Unknown-1	0	
400	WP_144387969.1		Vibrio algivorus	Proteobacteria	238	1	Unknown-1	1	1164
401	WP_047725505.1		Vibrio	Proteobacteria	239	8	Unknown-1	1	1164
402	HAS6263011.1		Vibrio vulnificus	Proteobacteria	239	8	Unknown-1	0	
403	WP_206636432.1		Vibrio vulnificus	Proteobacteria	239	8	Unknown-1	0	
404	MBM4926054.1		Vibrio parahaemolyticus	Proteobacteria	239	8	Unknown-1	0	
405	WP_157625313.1		Vibrio parahaemolyticus	Proteobacteria	239	8	Unknown-1	0	
406	EGQ8025797.1		Vibrio vulnificus	Proteobacteria	239	8	Unknown-1	0	
407	WP_140357297.1		Vibrio parahaemolyticus	Proteobacteria	239	8	Unknown-1	0	
408	WP_039464217.1		Vibrio navarrensis	Proteobacteria	239	8	Unknown-1	0	
409	WP_214506391.1		Shewanella sp. JM162201	Proteobacteria	240	1	Unknown-1	1	1163
410	WP_102322145.1		Vibrio breoganii	Proteobacteria	241	1	Unknown-1	1	1170
411	OZB93909.1		Alteromonas macleodii	Proteobacteria	242	1	Unknown-1	1	1164
412	WP_126802208.1		Aliidiomarina minuta	Proteobacteria	243	1	Unknown-1	1	1164
413	MBD7978408.1		Pseudomonadaceae bacterium Sa2CUA2	Proteobacteria	244	1	Unknown-1	1	1164
414	WP_094169002.1		Citrobacter	Proteobacteria	245	1	Unknown-1	1	1164
415	WP_160873420.1		Raoultella ornithinolytica	Proteobacteria	246	3	Unknown-1	1	1164
416	NLU18596.1		Serratia liquefaciens	Proteobacteria	246	3	Unknown-1	0	
417	WP_126124230.1		Rahnella aquatilis	Proteobacteria	246	3	Unknown-1	0	
418	WP_096073835.1		Pantoea	Proteobacteria	247	2	Unknown-1	1	1164
419	MBS6928379.1		Finegoldia magna	Firmicutes	247	2	Unknown-1	0	
420	MBI6296478.1		Proteus mirabilis	Proteobacteria	248	1	Unknown-1	1	1165
421	HAT3515072.1		Serratia marcescens	Proteobacteria	249	2	Unknown-1	1	1164
422	MBH2961216.1		Serratia marcescens	Proteobacteria	249	2	Unknown-1	0	
423	WP_039509378.1		Pectobacterium brasiliense	Proteobacteria	250	1	Unknown-1	1	1164
424	WP_063118745.1	EcAvs2	Escherichia coli	Proteobacteria	251	81	Unknown-1	1	1157
425	QEP62455.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
426	WP_080076515.1		Enterobacteriaceae	Proteobacteria	251	81	Unknown-1	0	

427	WP_116836614.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
428	WP_159178379.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
429	HBD2861917.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
430	WP_137579187.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
431	WP_088167376.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
432	WP_113383186.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
433	WP_186256192.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
434	MBJ3551091.1		Salmonella enterica subsp. enterica serovar Indiana	Proteobacteria	251	81	Unknown-1	0	
435	WP_130072542.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
436	WP_151160649.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
437	HAV8369792.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
438	WP_183057385.1		Salmonella enterica	Proteobacteria	251	81	Unknown-1	0	
439	EFO1352471.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
440	MBA2201653.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
441	EAQ0291027.1		Salmonella enterica	Proteobacteria	251	81	Unknown-1	0	
442	EAA6411354.1		Salmonella enterica subsp. enterica	Proteobacteria	251	81	Unknown-1	0	
443	WP_113445411.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
444	EBU6680158.1		Salmonella enterica subsp. enterica serovar Kimuenza	Proteobacteria	251	81	Unknown-1	0	
445	HBD5595896.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
446	EED0306499.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
447	EBR6397410.1		Salmonella enterica	Proteobacteria	251	81	Unknown-1	0	
448	HAL6807877.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
449	WP_073461897.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
450	WP_024250580.1		Enterobacteriaceae	Proteobacteria	251	81	Unknown-1	0	
451	EFN8183969.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
452	HAJ3451287.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
453	WP_061091478.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
454	WP_192522173.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
455	WP_000433377.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
456	WP_061359862.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	

457	WP_044325588.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
458	HBB4873077.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
459	EFE0886165.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
460	NPH42110.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
461	WP_203640994.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
462	HAV8259074.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
463	HAV7956343.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
464	HAV9184058.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
465	EFB4095673.1		Escherichia coli O101	Proteobacteria	251	81	Unknown-1	0	
466	WP_097457525.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
467	WP_074513070.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
468	HAI9831213.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
469	HAX8334665.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
470	WP_032199984.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
471	EEQ5903002.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
472	HAH1223270.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
473	HAM9425206.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
474	WP_042099759.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
475	EFJ3257616.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
476	WP_001394806.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
477	WP_001396531.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
478	WP_086598298.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
479	HBA6659751.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
480	EFD3131274.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
481	WP_048237157.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
482	HBE6470387.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
483	EFM3087507.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
484	WP_000433375.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
485	EFC9683671.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
486	EFS3848353.1		Shigella sonnei	Proteobacteria	251	81	Unknown-1	0	
487	WP_052978476.1		Enterobacteriaceae	Proteobacteria	251	81	Unknown-1	0	
488	MGR34492.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
489	WP_160391323.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
490	WP_061362275.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
491	EFJ5766452.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	

492	WP_160459745.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
493	WP_129959750.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
494	HAP2734117.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
495	HAI9571447.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
496	WP_097730583.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
497	EFC5419363.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
498	WP_053001144.1		Enterobacteriaceae	Proteobacteria	251	81	Unknown-1	0	
499	WP_077135241.1		Shigella sonnei	Proteobacteria	251	81	Unknown-1	0	
500	MBB8013336.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
501	EFZ0881642.1		Shigella sonnei	Proteobacteria	251	81	Unknown-1	0	
502	HAK9790810.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
503	HBA6180172.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
504	HAL7313383.1		Escherichia coli	Proteobacteria	251	81	Unknown-1	0	
505	WP_111961258.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	1	1157
506	WP_061356479.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
507	WP_130238033.1		unclassified Escherichia	Proteobacteria	252	48	Unknown-1	0	
508	WP_044687640.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
509	EFE8334592.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
510	EHL6267659.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
511	HAX5165561.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
512	ECY0836695.1		Salmonella enterica subsp. enterica	Proteobacteria	252	48	Unknown-1	0	
513	EGF6972849.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
514	WP_103723445.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
515	EGU0061854.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
516	WP_097339780.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
517	HBA4063945.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
518	HAW2890739.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
519	MBI9940425.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
520	EFO2887580.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
521	EEZ5641592.1		Escherichia coli O21	Proteobacteria	252	48	Unknown-1	0	
522	HAH5105541.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
523	EFC2774209.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
524	WP_187809968.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
525	WP_079948925.1		Enterobacteriaceae	Proteobacteria	252	48	Unknown-1	0	

526	EFN534183 8.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
527	EFB445932 6.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
528	HAV992508 2.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
529	MGQ58841. 1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
530	WP_089518 599.1		Escherichia	Proteobacteria	252	48	Unknown-1	0	
531	WP_139496 393.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
532	WP_089581 855.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
533	EFM003327 0.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
534	HBB389377 2.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
535	HAW90513 41.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
536	HAN803417 8.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
537	HAM53329 36.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
538	EFE599054 1.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
539	WP_069892 550.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
540	HAV859043 4.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
541	WP_139591 678.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
542	WP_098723 360.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
543	WP_130532 747.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
544	WP_139525 025.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
545	WP_180492 042.1		Escherichia fergusonii	Proteobacteria	252	48	Unknown-1	0	
546	EFB619990 0.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
547	EET691680 1.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
548	MBC03153 03.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
549	WP_097491 857.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
550	WP_115800 270.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
551	WP_000433 376.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
552	WP_053289 867.1		Escherichia coli	Proteobacteria	252	48	Unknown-1	0	
553	WP_117036 708.1		Klebsiella	Proteobacteria	253	1	Unknown-1	1	1164
554	WP_058675 881.1		Klebsiella aerogenes	Proteobacteria	254	4	Unknown-1	1	1164
555	WP_109220 861.1		Klebsiella oxytoca	Proteobacteria	254	4	Unknown-1	0	
556	HCZ39992. 1		Brochothrix thermosphacta	Firmicutes	254	4	Unknown-1	0	
557	WP_086816 738.1		Raoultella ornithinolytica	Proteobacteria	254	4	Unknown-1	0	
558	KFN39608. 1		Sulfuricurvum sp. MLSB	Proteobacteria	255	1	Transmembrane	1	1052

559	WP_029523333.1		Persephonella sp. KM09-Lau-8	Aquificae	256	1	TIR	1	1113
560	WP_195681335.1		Aliarcobacter butzleri	Proteobacteria	257	1	Short	1	1043
561	MBS4067095.1		Sulfurimonas sp.	Proteobacteria	258	1	Short	1	1142
562	MBR4629547.1		Treponema sp.	Spirochaetes	259	1	Mrr	1	1042
563	WP_059125418.1		Chryseobacterium sp. JAH	Bacteroidetes	260	1	Mrr	1	1059
564	WP_184093214.1		Anaerocolumna cellulosilytica	Firmicutes	261	1	Cap4	1	945
565	WP_011968819.1		Clostridium beijerinckii	Firmicutes	262	1	MBL + serine protease	1	990
566	NFL44733.1		Clostridium botulinum	Firmicutes	263	1	Unclassified	1	991
567	WP_056030549.1		Paenibacillus sp. Leaf72	Firmicutes	264	1	MBL + serine protease	1	1025
568	WP_068664479.1		Paenibacillus oryzisoli	Firmicutes	265	1	Cap4	1	1021
569	WP_068779286.1		Paenibacillus sp. GM2	Firmicutes	266	1	Unclassified	1	1028
570	WP_120464350.1		Paenibacillus aceti	Firmicutes	267	1	Unclassified	1	1029
571	WP_066375548.1		Bacillus sp. FJAT-27264	Firmicutes	268	1	Unclassified	1	1028
572	WP_060626512.1		Paenibacillus etheri	Firmicutes	269	1	Unclassified	1	1028
573	WP_057566418.1		Paeniclostridium sordellii	Firmicutes	270	2	MBL + serine protease	1	1085
574	WP_195924194.1		Clostridium sp. 1001271B_151109_D9	Firmicutes	270	2	MBL + serine protease	0	
575	WP_045539061.1		Clostridium botulinum	Firmicutes	271	6	Unclassified	1	1083
576	WP_154232133.1		Clostridium botulinum	Firmicutes	271	6	Unclassified	0	
577	WP_123955385.1		Clostridium botulinum	Firmicutes	271	6	Unclassified	0	
578	WP_158088111.1		Clostridium botulinum	Firmicutes	271	6	Unclassified	0	
579	APQ97766.1		Clostridium botulinum	Firmicutes	271	6	Unclassified	0	
580	OSB09889.1		Clostridium botulinum	Firmicutes	271	6	Unclassified	0	
581	WP_205455806.1		Clostridium botulinum	Firmicutes	272	1	Unclassified	1	1081
582	WP_191768948.1		Clostridium sp. Sa3CVN1	Firmicutes	273	1	SIR2	1	1083
583	MBN1054911.1		Clostridium botulinum	Firmicutes	274	1	Unclassified	1	1084
584	WP_077856663.1		Clostridium beijerinckii	Firmicutes	275	1	Unclassified	1	1084
585	WP_062239948.1		Fictibacillus sp. FJAT-27399	Firmicutes	276	1	Unknown-2	1	1080
586	WP_214891666.1		Exiguobacterium sp. s142	Firmicutes	277	1	Unknown-2	1	1021
587	WP_034756591.1		Rossellomorea vietnamensis	Firmicutes	278	1	Unknown-2	1	1076

588	WP_062187138.1		Bacillales	Firmicutes	279	1	Unknown-2	1	1084
589	WP_048718972.1		Bacillus sp. 522_BSPC	Firmicutes	280	1	Unknown-2	1	1084
590	WP_187046568.1		Sporosarcina sp. resist	Firmicutes	281	1	Unknown-2	1	1082
591	WP_079513709.1		Bacillus marisflavi	Firmicutes	282	1	Unknown-2	1	1079
592	WP_174497309.1		Salirhabdus euzebyi	Firmicutes	283	1	Unknown-2	1	1080
593	WP_108586182.1		Paenisporosarcina sp. OV554	Firmicutes	284	1	Unknown-2	1	1080
594	WP_149212634.1		Bacillus sp. AR2-1	Firmicutes	285	1	Unknown-2	1	1082
595	WP_000734833.1		Bacillus cereus group	Firmicutes	286	4	Unknown-2	1	1020
596	WP_003286106.1		Bacillus thuringiensis	Firmicutes	286	4	Unknown-2	0	
597	AIE33489.1		Bacillus thuringiensis serovar kurstaki str. HD-1	Firmicutes	286	4	Unknown-2	0	
598	WP_016099943.1		Bacillus cereus group	Firmicutes	286	4	Unknown-2	0	
599	EEL60383.1		Bacillus cereus Rock4-18	Firmicutes	287	5	Unknown-2	1	1020
600	WP_050427157.1		Bacillus cereus group	Firmicutes	287	5	Unknown-2	0	
601	PEN60497.1		Bacillus toyonensis	Firmicutes	287	5	Unknown-2	0	
602	WP_147740523.1		Bacillus sp. AR13-1	Firmicutes	287	5	Unknown-2	0	
603	WP_215598725.1		Bacillus mycoides	Firmicutes	287	5	Unknown-2	0	
604	WP_000229126.1		Bacillus cereus	Firmicutes	288	2	Unknown-2	1	1081
605	EEL55333.1		Bacillus cereus Rock4-2	Firmicutes	288	2	Unknown-2	0	
606	WP_215575298.1		Bacillus mycoides	Firmicutes	289	1	Unknown-2	1	1081
607	PFA26012.1		Bacillus cereus	Firmicutes	290	4	Unknown-2	1	1081
608	WP_176524921.1		Bacillus cereus	Firmicutes	290	4	Unknown-2	0	
609	WP_179880262.1		Bacillus cereus	Firmicutes	290	4	Unknown-2	0	
610	PET96541.1		Bacillus cereus	Firmicutes	290	4	Unknown-2	0	
611	WP_033671653.1		Bacillus cereus	Firmicutes	291	2	Unknown-2	1	1081
612	EOQ01289.1		Bacillus cereus VDM021	Firmicutes	291	2	Unknown-2	0	
613	WP_000229125.1		Bacillus cereus	Firmicutes	292	7	Unknown-2	1	1081
614	WP_001992824.1		Bacillus cereus	Firmicutes	292	7	Unknown-2	0	
615	WP_078438644.1		Bacillus cereus	Firmicutes	292	7	Unknown-2	0	
616	WP_098392054.1		Bacillus thuringiensis	Firmicutes	292	7	Unknown-2	0	
617	WP_098783989.1		Bacillus cereus	Firmicutes	292	7	Unknown-2	0	
618	WP_098542561.1		Bacillus	Firmicutes	292	7	Unknown-2	0	
619	WP_098906760.1		Bacillus cereus	Firmicutes	292	7	Unknown-2	0	

620	WP_098211518.1		Bacillus thuringiensis	Firmicutes	293	4	Unknown-2	1	1081
621	WP_029442959.1		Bacillus thuringiensis	Firmicutes	293	4	Unknown-2	0	
622	WP_098953493.1		Bacillus thuringiensis	Firmicutes	293	4	Unknown-2	0	
623	WP_196471714.1		Bacillus sp. SRB1LM	Firmicutes	293	4	Unknown-2	0	
624	AAY60470.1		Bacillus cereus E33L	Firmicutes	294	2	Unknown-2	1	1081
625	WP_049770107.1		Bacillus cereus	Firmicutes	294	2	Unknown-2	0	
626	WP_102377568.1		Raoultibacter timonensis	Actinobacteria	295	1	Putative REase	1	1094
627	WP_147556148.1		Olsenella timonensis	Actinobacteria	296	1	Putative REase	1	1094
628	MBE6472807.1		Coriobacteriaceae bacterium	Actinobacteria	297	1	Putative REase	1	1086
629	WP_114620037.1		Senegalimassilia anaerobia	Actinobacteria	298	1	Putative REase	1	1083
630	WP_040324504.1		Lancefieldella rimae	Actinobacteria	299	2	Putative REase	1	1106
631	EEE17364.1		Atopobium rimae ATCC 49626	Actinobacteria	299	2	Putative REase	0	
632	WP_157293469.1		Bhargavaea sp. CC-171006	Firmicutes	300	1	Putative REase	1	1083
633	WP_123258799.1		Bacillus circulans	Firmicutes	301	1	Putative REase	1	1083
634	WP_212947887.1		Bacillus terrae	Firmicutes	302	2	Putative REase	1	1083
635	WP_213020694.1		Bacillus terrae	Firmicutes	302	2	Putative REase	0	
636	WP_000831218.1		Staphylococcus	Firmicutes	303	170	Putative REase	1	1029
637	MBH4771124.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
638	CAC6829849.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
639	WP_107240618.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
640	WP_000831220.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
641	WP_148850300.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
642	WP_111762425.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
643	WP_130827174.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
644	HAR4283810.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	

645	CUD40550.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
646	CUF88264.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
647	WP_180148703.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
648	WP_060643304.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
649	MBN5677841.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
650	MBO8905030.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
651	WP_111021996.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
652	QNZ78543.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
653	WP_031881329.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
654	WP_190283091.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
655	CAC6315867.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
656	QNW69608.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
657	WP_050901426.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
658	WP_163989937.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
659	WP_112368364.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
660	MBD6843149.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
661	QFM39565.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
662	WP_031907024.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
663	MBO8611771.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
664	MBD6816385.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
665	WP_117221292.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
666	WP_086040509.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
667	MBH4544552.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	

668	WP_178890276.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
669	WP_020444746.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
670	WP_088591727.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
671	HAY6348442.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
672	WP_182078398.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
673	WP_208888033.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
674	HAR3447639.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
675	CAA3932771.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
676	WP_070957873.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
677	WP_193627209.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
678	HAR4049771.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
679	WP_111130628.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
680	WP_031836461.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
681	WP_043854015.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
682	WP_050975120.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
683	MBD6767353.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
684	WP_057511206.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
685	WP_094970737.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
686	WP_031885004.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
687	WP_061730476.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
688	WP_100250365.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
689	WP_052997892.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
690	WP_049280028.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	

691	WP_058102374.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
692	CAC7100794.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
693	NGA39475.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
694	HAR3199002.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
695	WP_183139306.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
696	WP_089536551.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
697	WP_031880834.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
698	WP_086097462.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
699	WP_053865143.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
700	WP_199004269.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
701	MBR8885435.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
702	NFX52868.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
703	WP_089520914.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
704	CAC6223864.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
705	WP_050964237.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
706	WP_072419819.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
707	MBG1048756.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
708	WP_053863046.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
709	CAC9293094.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
710	WP_100309878.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
711	WP_031893543.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
712	WP_064128622.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
713	WP_031784187.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	

714	HAR4409039.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
715	WP_048521815.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
716	WP_031784391.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
717	WP_117205273.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
718	WP_053871702.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
719	WP_130828151.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
720	WP_111141136.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
721	MBR9090169.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
722	WP_070044558.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
723	WP_050958245.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
724	WP_072526894.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
725	WP_094971332.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
726	MBS3387488.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
727	HAR3273513.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
728	WP_117214867.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
729	WP_094978238.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
730	CAC6411422.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
731	HBC8709154.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
732	WP_050965418.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
733	QNW16518.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
734	WP_050958930.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
735	WP_053866062.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
736	MBD3533028.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	

737	CAA4054467.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
738	WP_176290353.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
739	WP_108998647.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
740	HAR3996249.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
741	HBC4238540.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
742	CAA3931854.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
743	WP_086096799.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
744	WP_050959431.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
745	CAC8412998.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
746	WP_050976616.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
747	WP_159314348.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
748	WP_048666651.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
749	KDP47687.1		Staphylococcus aureus subsp. aureus CO-86	Firmicutes	303	170	Putative REase	0	
750	WP_111091625.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
751	MBR9405337.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
752	WP_153906411.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
753	MBG3565827.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
754	CAA3851911.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
755	WP_053862861.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
756	WP_112379994.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
757	KSA61853.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
758	MBG1102341.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
759	CAA4480560.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	

760	WP_111045577.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
761	WP_053867690.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
762	WP_031869418.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
763	WP_123073961.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
764	CAA3920541.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
765	WP_172844216.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
766	WP_183137629.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
767	WP_119107085.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
768	WP_000831215.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
769	WP_161475420.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
770	CAA4604820.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
771	WP_049280959.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
772	WP_183135947.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
773	WP_117207519.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
774	MBG1482299.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
775	WP_111015857.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
776	WP_138397279.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
777	HAR4862555.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
778	WP_119106891.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
779	WP_000831219.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
780	HAR5534250.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
781	WP_098691632.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
782	WP_208703208.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	

783	CAC6089338.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
784	WP_031905627.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
785	HAR5275964.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
786	WP_109039131.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
787	WP_053875652.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
788	WP_057511163.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
789	WP_031867146.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
790	MBG3823946.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
791	MBR8903263.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
792	MBN5688613.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
793	WP_031882044.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
794	MBG3235706.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
795	WP_031890994.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
796	HAR6905648.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
797	WP_049314120.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
798	WP_031911572.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
799	CAA3733639.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
800	WP_072468068.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
801	WP_086043809.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
802	HAR3415876.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
803	CAA3806228.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
804	MBD6935363.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	
805	WP_031882023.1		Staphylococcus aureus	Firmicutes	303	170	Putative REase	0	

806	WP_204253803.1		Staphylococcus sp. GDH8C146P	Firmicutes	304	1	Putative REase	1	1097
807	WP_157947185.1		Staphylococcus equorum	Firmicutes	305	1	Putative REase	1	1098
808	WP_072419803.1		Staphylococcus aureus	Firmicutes	306	3	Putative REase	1	1098
809	WP_206543481.1		Mammaliicoccus sciuri	Firmicutes	306	3	Putative REase	0	
810	HAR5458506.1		Staphylococcus aureus	Firmicutes	306	3	Putative REase	0	
811	WP_058306330.1		Gracilibacillus massiliensis	Firmicutes	307	1	Putative REase	1	1095
812	WP_159033196.1		Oceanobacillus iheyensis	Firmicutes	308	1	Putative REase	1	1098
813	WP_099225253.1		Listeria costaricensis	Firmicutes	309	1	Unknown-2	1	1082
814	WP_181343625.1		Listeria monocytogenes	Firmicutes	310	9	Unknown-2	1	1087
815	WP_185510341.1		Listeria innocua	Firmicutes	310	9	Unknown-2	0	
816	WP_194328251.1		Listeria marthii	Firmicutes	310	9	Unknown-2	0	
817	EAG5905227.1		Listeria monocytogenes	Firmicutes	310	9	Unknown-2	0	
818	ECJ9746443.1		Listeria monocytogenes	Firmicutes	310	9	Unknown-2	0	
819	EAF3625748.1		Listeria monocytogenes	Firmicutes	310	9	Unknown-2	0	
820	EGF6980377.1		Listeria monocytogenes	Firmicutes	310	9	Unknown-2	0	
821	WP_194344514.1		Listeria welshimeri	Firmicutes	310	9	Unknown-2	0	
822	MBC2348019.1		Listeria welshimeri	Firmicutes	310	9	Unknown-2	0	
823	WP_185491199.1		Listeria booriae	Firmicutes	311	2	Unknown-2	1	1092
824	WP_185588566.1		Listeria booriae	Firmicutes	311	2	Unknown-2	0	
825	WP_104846046.1		Enterococcus faecalis	Firmicutes	312	7	Unknown-2	1	1081
826	EEU79966.1		Enterococcus faecalis Fly1	Firmicutes	312	7	Unknown-2	0	
827	WP_141753534.1		Enterococcus	Firmicutes	312	7	Unknown-2	0	
828	HAR2508595.1		Enterococcus faecalis	Firmicutes	312	7	Unknown-2	0	
829	WP_025187243.1		Enterococcus faecalis	Firmicutes	312	7	Unknown-2	0	
830	EPI34319.1		Enterococcus faecalis SLO2C-1	Firmicutes	312	7	Unknown-2	0	
831	EGO8527747.1		Enterococcus faecalis	Firmicutes	312	7	Unknown-2	0	
832	WP_079505170.1		Lactococcus piscium	Firmicutes	313	1	Unknown-2	1	1083
833	WP_002301728.1		Enterococcus faecium	Firmicutes	314	6	Unknown-2	1	1083
834	WP_123862539.1		Enterococcus durans	Firmicutes	314	6	Unknown-2	0	

835	HAQ467090 8.1		Enterococcus faecium	Firmicutes	314	6	Unkno wn-2	0	
836	WP_155853 353.1		Enterococcus gallinarum	Firmicutes	314	6	Unkno wn-2	0	
837	WP_136770 837.1		Enterococcus hirae	Firmicutes	314	6	Unkno wn-2	0	
838	HAZ098680 1.1		Enterococcus faecium	Firmicutes	314	6	Unkno wn-2	0	
839	WP_095803 592.1		Enterococcus thailandicus	Firmicutes	315	1	Unkno wn-2	1	1081
840	EAH263794 5.1		Listeria monocytogen es	Firmicutes	316	1	Unkno wn-2	1	1060
841	WP_185587 111.1		Listeria cossartiae	Firmicutes	317	1	Unkno wn-2	1	1055
842	WP_095345 506.1		Lactococcus lactis	Firmicutes	318	3	Unkno wn-2	1	1062
843	WP_190992 208.1		Lactococcus lactis	Firmicutes	318	3	Unkno wn-2	0	
844	QLF89377.1		Lactococcus lactis subsp. lactis	Firmicutes	318	3	Unkno wn-2	0	
845	WP_076768 532.1		Jeotgalibaca sp. PTS2502	Firmicutes	319	2	Unkno wn-2	1	1063
846	WP_100084 143.1		Streptococcus parauberis	Firmicutes	319	2	Unkno wn-2	0	
847	WP_121260 538.1		Enterococcus casseliflavus	Firmicutes	320	1	Unkno wn-2	1	1062
848	WP_207703 288.1		Enterococcus sp. 665A	Firmicutes	321	1	Unkno wn-2	1	1062
849	WP_185501 448.1		Listeria innocua	Firmicutes	322	1	Unkno wn-2	1	1062
850	EGP545927 7.1		Enterococcus faecium	Firmicutes	323	1	Unkno wn-2	1	1062
851	OLA01742. 1		Clostridium sp. CAG:62_40_ 43	Firmicutes	324	1	Mrr	1	1050
852	NLP31470.1		Clostridiales bacterium	Firmicutes	325	1	Mrr	1	1057
853	HHX60383. 1		Epulopiscium sp.	Firmicutes	326	1	Mrr	1	1051
854	MBD52396 53.1		Bacteroidales bacterium	Bacteroidete s	327	1	Transm embran e	1	1100
855	WP_027202 185.1		Butyricimona s virosa	Bacteroidete s	328	1	Transm embran e	1	1083
856	WP_132216 521.1		Phocaeicola massiliensis	Bacteroidete s	329	1	Transm embran e	1	1068
857	WP_007855 885.1		Phocaeicola dorei	Bacteroidete s	330	2	Transm embran e	1	1068
858	WP_118037 838.1		Bacteroidales	Bacteroidete s	330	2	Transm embran e	0	
859	HAT62701. 1		Prevotella sp.	Bacteroidete s	331	1	Transm embran e	1	1102
860	RHG31319. 1		Prevotella copri	Bacteroidete s	332	3	Transm embran e	1	1081
861	WP_147379 947.1		Prevotella copri	Bacteroidete s	332	3	Transm embran e	0	
862	WP_147380 225.1		Prevotella copri	Bacteroidete s	332	3	Transm embran e	0	

863	MBO74936 66.1		Bacteroidales bacterium	Bacteroidetes	333	1	Transmembrane	1	1078
864	MBR57819 36.1		Bacteroidales bacterium	Bacteroidetes	334	1	Transmembrane	1	1083
865	MBD92614 28.1		Prevotella sp.	Bacteroidetes	335	1	Transmembrane	1	1094
866	WP_131927 580.1		Bacteroides heparinolyticus	Bacteroidetes	336	3	Transmembrane	1	1111
867	TCO86115. 1		Bacteroides heparinolyticus	Bacteroidetes	336	3	Transmembrane	0	
868	WP_139145 433.1		Prevotella sp. HMSC073D09	Bacteroidetes	336	3	Transmembrane	0	
869	AVM50283. 1		Capnocytophaga sp. oral taxon 878	Bacteroidetes	337	2	Transmembrane	1	1058
870	WP_129588 215.1		Capnocytophaga sp. oral taxon 878	Bacteroidetes	337	2	Transmembrane	0	
871	WP_028899 599.1		Prevotella sp. HJM029	Bacteroidetes	338	2	Transmembrane	1	1112
872	WP_021669 636.1		Prevotella disiens	Bacteroidetes	338	2	Transmembrane	0	
873	WP_178957 639.1		Capnocytophaga sp. oral taxon 903	Bacteroidetes	339	1	Transmembrane	1	1058
874	WP_099811 528.1		Porphyromonas gingivalis	Bacteroidetes	340	1	Transmembrane	1	1109
875	WP_025843 059.1		Porphyromonas gingivicanis	Bacteroidetes	341	1	Transmembrane	1	1109
876	WP_101638 437.1		Brachyspira hyodysenteriae	Spirochaetes	342	1	Short	1	1070
877	WP_135106 196.1		Dysgonomonas mossii	Bacteroidetes	343	1	MBL + serine protease	1	1031
878	WP_018668 502.1		Bacteroides gallinarum	Bacteroidetes	344	1	MBL + serine protease	1	1029
879	WP_172775 691.1		Phocaeicola vulgatus	Bacteroidetes	345	1	MBL + serine protease	1	1030
880	WP_195456 099.1		Bacteroides ovatus	Bacteroidetes	346	1	MBL + serine protease	1	1028
881	WP_005803 321.1		Bacteroides fragilis	Bacteroidetes	347	1	MBL + serine protease	1	1028
882	WP_021981 164.1		Paraprevotella clara	Bacteroidetes	348	1	MBL + serine protease	1	1028
883	MBQ96748 75.1		Bacteroidaceae bacterium	Bacteroidetes	349	1	MBL + serine protease	1	1083

884	WP_103915756.1		Prevotella ruminicola	Bacteroidetes	350	1	MBL + serine protease	1	1097
885	WP_115321296.1		Porphyromonas endodontalis	Bacteroidetes	351	2	MBL + serine protease	1	1046
886	EEN81970.1		Porphyromonas endodontalis ATCC 35406	Bacteroidetes	351	2	MBL + serine protease	0	
887	WP_004323295.1		Bacteroides	Bacteroidetes	352	4	MBL + serine protease	1	1065
888	WP_113748259.1		Bacteroides thetaiotaomicron	Bacteroidetes	352	4	MBL + serine protease	0	
889	WP_118337112.1		Bacteroides thetaiotaomicron	Bacteroidetes	352	4	MBL + serine protease	0	
890	WP_011108364.1		Bacteroides thetaiotaomicron	Bacteroidetes	352	4	MBL + serine protease	0	
891	WP_057253524.1		Bacteroides uniformis	Bacteroidetes	353	1	MBL + serine protease	1	1071
892	WP_011965437.1		Bacteroidales	Bacteroidetes	354	5	MBL + serine protease	1	1067
893	WP_149970508.1		Bacteroides ovatus	Bacteroidetes	354	5	MBL + serine protease	0	
894	WP_149940681.1		Phocaeicola dorei	Bacteroidetes	354	5	MBL + serine protease	0	
895	WP_117985916.1		Bacteroides stercoris	Bacteroidetes	354	5	MBL + serine protease	0	
896	WP_122264288.1		Bacteroides caccae	Bacteroidetes	354	5	MBL + serine protease	0	
897	CDD20515.1		Prevotella sp. CAG:732	Bacteroidetes	355	1	MBL + serine protease	1	1074
898	MBR2030654.1		Alistipes sp.	Bacteroidetes	356	1	MBL + serine protease	1	1029
899	WP_120174868.1		Prevotella melaninogenica	Bacteroidetes	357	1	MBL + serine protease	1	1072
900	WP_107625451.1		Prevotella sp. oral taxon 313	Bacteroidetes	358	1	MBL + serine protease	1	1077
901	WP_156100352.1		Prevotella denticola	Bacteroidetes	359	3	MBL + serine	1	1075

							protease		
902	KGF42117.1		Prevotella denticola DNF00960	Bacteroidetes	359	3	MBL + serine protease	0	
903	WP_211796708.1		Prevotella denticola	Bacteroidetes	359	3	MBL + serine protease	0	
904	WP_094448219.1		Prevotella bryantii	Bacteroidetes	360	1	MBL + serine protease	1	1030
905	MBQ6038136.1		Bacteroidaceae bacterium	Bacteroidetes	361	1	MBL + serine protease	1	1056
906	WP_182127237.1		Bacteroides fragilis	Bacteroidetes	362	1	MBL + serine protease	1	1045
907	WP_211420628.1		Bacteroides xylanisolvens	Bacteroidetes	363	2	MBL + serine protease	1	1017
908	MBQ2884208.1		Alphaproteobacteria bacterium	Proteobacteria	363	2	MBL + serine protease	0	
909	HAT61518.1		Prevotella sp.	Bacteroidetes	364	1	MBL + serine protease	1	1016
910	RGW77185.1		Prevotella copri	Bacteroidetes	365	2	MBL + serine protease	1	1022
911	WP_181992477.1		Prevotella copri	Bacteroidetes	365	2	MBL + serine protease	0	
912	MBD5304460.1		Bacteroides sp.	Bacteroidetes	366	1	MBL + serine protease	1	1022
913	WP_062695471.1		Bacteroides thetaiotaomicron	Bacteroidetes	367	1	MBL + serine protease	1	1030
914	MBF1645747.1		Prevotella sp.	Bacteroidetes	368	1	MBL + serine protease	1	1019
915	WP_005874375.1		Porphyromonas gingivalis	Bacteroidetes	369	2	MBL + serine protease	1	1018
916	AKV64330.1		Porphyromonas gingivalis	Bacteroidetes	369	2	MBL + serine protease	0	
917	WP_016420980.1		Capnocytophaga granulosa	Bacteroidetes	370	1	MBL + serine protease	1	1026
918	MBP5551268.1		Bacilli bacterium	Firmicutes	371	1	Mrr	1	1082
919	WP_156286358.1		Oceanivirga salmonicida	Fusobacteria	372	1	Mrr	1	1059

920	WP_071408891.1		Globicatella sp. HMSC072A10	Firmicutes	373	2	Mrr	1	1098
921	WP_138404684.1		Facklamia tabacinasalis	Firmicutes	373	2	Mrr	0	
922	WP_083070137.1		Aerococcus viridans	Firmicutes	374	2	Mrr	1	1098
923	WP_048729165.1		Carnobacterium sp. 1290_CSPC	Firmicutes	374	2	Mrr	0	
924	WP_102198414.1		Anaerococcus hydrogenalis	Firmicutes	375	1	Mrr	1	1082
925	WP_209814239.1		Anaerococcus nagyae	Firmicutes	376	1	Mrr	1	1082
926	MBE6698835.1		Ruminococcaceae bacterium	Firmicutes	377	1	Mrr	1	1078
927	MBO5481645.1		Clostridia bacterium	Firmicutes	378	1	Mrr	1	1092
928	WP_159160499.1		Enterococcus	Firmicutes	379	1	Mrr	1	1094
929	WP_161637152.1		Erysipelothrix rhusiopathiae	Firmicutes	380	1	Mrr	1	1094
930	WP_195674910.1		Streptococcus sp. HF-1907	Firmicutes	381	1	Mrr	1	1092
931	WP_168751325.1		Streptococcus mutans	Firmicutes	382	2	Mrr	1	1090
932	WP_002266448.1		Streptococcus mutans	Firmicutes	382	2	Mrr	0	
933	WP_000790890.1		Streptococcus	Firmicutes	383	1	Mrr	1	1084
934	WP_000791089.1		Streptococcus sp. SK643	Firmicutes	384	1	Mrr	1	1084
935	WP_015604235.1		Streptococcus cristatus	Firmicutes	385	1	Mrr	1	1089
936	WP_074595911.1		Streptococcus gallolyticus	Firmicutes	386	1	Mrr	1	1085
937	WP_141597172.1		Streptococcus suis	Firmicutes	387	3	Mrr	1	1087
938	WP_105242307.1		Streptococcus suis	Firmicutes	387	3	Mrr	0	
939	WP_205041337.1		Streptococcus suis	Firmicutes	387	3	Mrr	0	
940	WP_053861650.1		Streptococcus suis	Firmicutes	388	3	Mrr	1	1087
941	NQK38637.1		Streptococcus suis	Firmicutes	388	3	Mrr	0	
942	NQL53096.1		Streptococcus suis	Firmicutes	388	3	Mrr	0	
943	WP_044693978.1		Streptococcus suis	Firmicutes	389	4	Mrr	1	1087
944	NQP19822.1		Streptococcus suis	Firmicutes	389	4	Mrr	0	
945	WP_105122361.1		Streptococcus suis	Firmicutes	389	4	Mrr	0	
946	WP_044757781.1		Streptococcus suis	Firmicutes	389	4	Mrr	0	
947	NCB80241.1		Bacilli bacterium	Firmicutes	390	2	Mrr	1	1088
948	NQI72987.1		Streptococcus suis	Firmicutes	390	2	Mrr	0	
949	WP_174850271.1		Streptococcus suis	Firmicutes	391	1	Mrr	1	1088
950	WP_115883961.1		Jeotgalicoccus halotolerans	Firmicutes	392	1	Mrr	1	1086
951	WP_139916693.1		Lactococcus lactis	Firmicutes	393	1	Mrr	1	1094
952	WP_014570424.1		Lactococcus lactis	Firmicutes	394	4	Mrr	1	1100

953	MSU87837.1		Streptococcus dysgalactiae subsp. dysgalactiae	Firmicutes	394	4	Mrr	0	
954	KSU30207.1		Lactococcus lactis subsp. lactis	Firmicutes	394	4	Mrr	0	
955	WP_058207089.1		Lactococcus lactis	Firmicutes	394	4	Mrr	0	
956	MBR3140184.1		Methanobrevibacter sp.	Euryarchaeota	395	1	Mrr	1	1097
957	NTK90559.1		Enterococcus faecium	Firmicutes	396	1	Mrr	1	1105
958	WP_080146958.1		Marinilactibacillus piezotolerans	Firmicutes	397	1	Mrr	1	1098
959	WP_076648268.1		Latilactobacillus sakei	Firmicutes	398	1	Mrr	1	1089
960	EGS116331.1		Enterococcus faecalis	Firmicutes	399	1	Mrr	1	1094
961	MBD5153606.1		Oscillibacter sp.	Firmicutes	400	1	Mrr	1	1093
962	MBQ9838223.1		Oscillospiraceae bacterium	Firmicutes	401	1	Mrr	1	1098
963	MBE6947947.1		Ruminococcaceae bacterium	Firmicutes	402	1	Mrr	1	1098
964	MBQ0090439.1		Prevotellaceae bacterium	Bacteroidetes	403	1	Mrr	1	1057
965	NLD51079.1		Clostridiaceae bacterium	Firmicutes	404	1	Mrr	1	1055
966	MBS5190201.1		Lachnospiraceae bacterium	Firmicutes	405	1	Mrr	1	1047
967	WP_012864545.1		Veillonella parvula	Firmicutes	406	1	Mrr	1	1051
968	WP_085022297.1		Anaerovibrio sp. JC8	Firmicutes	407	1	Mrr	1	1053
969	WP_027209040.1		Butyrivibrio hungatei	Firmicutes	408	1	Mrr	1	1054
970	WP_087175154.1		Lachnoclostridium sp. An118	Firmicutes	409	1	Mrr	1	1051
971	WP_117699261.1		Roseburia sp. AF42-8	Firmicutes	410	1	Mrr	1	1079
972	MBQ5445788.1		Lachnospiraceae bacterium	Firmicutes	411	1	Mrr	1	1049
973	WP_087233958.1		Pseudoflavonifractor sp. An176	Firmicutes	412	1	Mrr	1	1116
974	MBS4815820.1		Clostridium sp.	Firmicutes	413	1	Mrr	1	1112
975	WP_118325411.1		Holdemanella biformis	Firmicutes	414	1	Mrr	1	1114
976	RGI92213.1		Blautia obeum	Firmicutes	415	3	Mrr	1	1112
977	WP_147324851.1		Blautia obeum	Firmicutes	415	3	Mrr	0	
978	TGX97904.1		Lachnospiraceae bacterium	Firmicutes	415	3	Mrr	0	
979	WP_055195524.1		Dorea longicatena	Firmicutes	416	2	Mrr	1	1109
980	WP_173792661.1		Dorea longicatena	Firmicutes	416	2	Mrr	0	
981	MBS5115448.1		Erysipelotrichaceae bacterium	Firmicutes	417	1	Mrr	1	1116
982	WP_087169081.1		Drancourtella sp. An12	Firmicutes	418	1	Mrr	1	1113
983	WP_150311049.1		Mediterraneibacter sp. SW178	Firmicutes	419	1	Mrr	1	1113

984	WP_055056604.1		Blautia obeum	Firmicutes	420	2	Mrr	1	1109
985	WP_118591515.1		Ruminococcus sp. AM29-26	Firmicutes	420	2	Mrr	0	
986	WP_118743365.1		Blautia sp. AF13-16	Firmicutes	421	1	Mrr	1	1113
987	WP_117585186.1		Megamonas	Firmicutes	422	1	Mrr	1	1050
988	WP_122788314.1		Intestinibacillus sp. Marseille-P6563	Firmicutes	423	1	Mrr	1	1127
989	HAM92927.1		Phascolarctobacterium succinatutens	Firmicutes	424	1	Mrr	1	1103
990	WP_158388290.1		Faecalibacterium prausnitzii	Firmicutes	425	1	Mrr	1	1101
991	WP_113992220.1		Faecalibacterium prausnitzii	Firmicutes	426	1	Mrr	1	1100
992	CDA47966.1		Dialister sp. CAG:486	Firmicutes	427	1	Mrr	1	1052
993	WP_090529044.1		Pseudobutyrivibrio sp. UC1225	Firmicutes	428	1	Mrr	1	1010
994	CUP42593.1		[Eubacterium] rectale	Firmicutes	429	6	Mrr	1	1102
995	RHL02996.1		[Eubacterium] rectale	Firmicutes	429	6	Mrr	0	
996	RGT10827.1		[Eubacterium] rectale	Firmicutes	429	6	Mrr	0	
997	RHL77199.1		[Eubacterium] rectale	Firmicutes	429	6	Mrr	0	
998	RGZ75199.1		[Eubacterium] rectale	Firmicutes	429	6	Mrr	0	
999	RHD97928.1		[Eubacterium] rectale	Firmicutes	429	6	Mrr	0	
1000	WP_172455242.1		Acetatifactor muris	Firmicutes	430	3	Mrr	1	1117
1001	NBH20026.1		Clostridiaceae bacterium	Firmicutes	430	3	Mrr	0	
1002	SOY31730.1		Acetatifactor muris	Firmicutes	430	3	Mrr	0	
1003	HAY04329.1		Lachnospiraceae bacterium	Firmicutes	431	1	Mrr	1	1079
1004	RKI31113.1		bacterium D16-36	Unclassified	432	1	Mrr	1	1072
1005	RPH96551.1		candidate division KSB1 bacterium	Unclassified	433	1	Mrr	1	1138
1006	MXW61621.1		Acidimicrobiaceae bacterium	Actinobacteria	434	1	Mrr	1	1176
1007	MBL8475058.1		Methyloversatilis sp.	Proteobacteria	435	1	Mrr	1	1182
1008	WP_194273463.1		Alcaligenes faecalis	Proteobacteria	436	4	Mrr	1	1180
1009	WP_164461348.1		Alcaligenes faecalis	Proteobacteria	436	4	Mrr	0	
1010	QFY77598.1		Alcaligenes faecalis	Proteobacteria	436	4	Mrr	0	
1011	AYR19454.1		Alcaligenes faecalis	Proteobacteria	436	4	Mrr	0	
1012	WP_185993547.1		Variovorax sp. KBS0712	Proteobacteria	437	1	Mrr	1	1178
1013	WP_057927205.1		Burkholderia ambifaria	Proteobacteria	438	1	Mrr	1	1180
1014	WP_173940876.1		Ralstonia solanacearum	Proteobacteria	439	2	Mrr	1	1179

1015	WP_173940774.1		Ralstonia solanacearum	Proteobacteria	439	2	Mrr	0	
1016	WP_200030761.1		Burkholderia multivorans	Proteobacteria	440	1	Mrr	1	1180
1017	OJV43370.1		Acidobacteriales bacterium 59-55	Acidobacteria	441	1	Mrr	1	1182
1018	HAZ63961.1		Armatimonadetes bacterium	Armatimonadetes	442	1	Mrr	1	1099
1019	WP_045689378.1		Hymenobacter sp. AT01-02	Bacteroidetes	443	2	Mrr	1	1214
1020	WP_125433766.1		Hymenobacter metallilatus	Bacteroidetes	443	2	Mrr	0	
1021	WP_139252148.1		Hymenobacter psychrotolerans	Bacteroidetes	444	1	Mrr	1	1212
1022	HHW86159.1		Chloroflexi bacterium	Chloroflexi	445	2	Mrr	1	1121
1023	HAJ37911.1		Chloroflexi bacterium	Chloroflexi	445	2	Mrr	0	
1024	WP_132812165.1		Sulfuricurvum sp. IAE1	Proteobacteria	446	1	Mrr	1	1187
1025	MYC30576.1		Chloroflexi bacterium	Chloroflexi	447	1	Mrr	1	1226
1026	MXV85303.1		Candidatus Poribacteria bacterium	Unclassified	448	1	Mrr	1	1234
1027	WP_144295528.1		Rhodothermus marinus	Bacteroidetes	449	2	Mrr	1	1191
1028	ACY49764.1		Rhodothermus marinus DSM 4252	Bacteroidetes	449	2	Mrr	0	
1029	AVX31521.1		Carboxydocella thermautotrophica	Firmicutes	450	1	Mrr	1	1196
1030	WP_075517736.1		Moorella thermoacetica	Firmicutes	451	1	Mrr	1	1196
1031	MBI1814243.1		Deltaproteobacteria bacterium	Proteobacteria	452	1	Mrr	1	1216
1032	WP_141736409.1		Oligoflexus tunisiensis	Proteobacteria	453	1	Mrr	1	1290
1033	MBI1212898.1		Alphaproteobacteria bacterium	Proteobacteria	454	1	Mrr	1	1207
1034	WP_012826254.1		Haliangium ochraceum	Proteobacteria	455	1	Mrr	1	1217
1035	MBK7411705.1		Ignavibacteria bacterium	Ignavibacteriae	456	1	Mrr	1	1208
1036	MBL9165622.1		Planctomycetaceae bacterium	Planctomycetes	457	1	Mrr	1	1154
1037	WP_145111557.1		Botrimarina mediterranea	Planctomycetes	458	1	Mrr	1	1161
1038	WP_155121367.1		Bryobacter aggregatus	Acidobacteria	459	1	Mrr	1	1176
1039	WP_147452066.1		Corallococcus	Proteobacteria	460	2	Mrr	1	1181
1040	WP_171810979.1		Corallococcus exiguus	Proteobacteria	460	2	Mrr	0	
1041	WP_153018441.1		Moorella mulderi	Firmicutes	461	2	Mrr	1	1067
1042	KYH30690.1		Moorella mulderi DSM 14980	Firmicutes	461	2	Mrr	0	
1043	CQR49627.1		Haloferax massiliensis	Euryarchaeota	462	2	Mrr	1	1064
1044	WP_089777503.1		Haloferax massiliensis	Euryarchaeota	462	2	Mrr	0	

1045	WP_013035348.1		Haloferax volcanii	Euryarchaeota	463	1	Mrr	1	1060
1046	WP_013440480.1		Halogeometricum borinquense	Euryarchaeota	464	2	Mrr	1	1062
1047	ELY27710.1		Halogeometricum borinquense DSM 11551	Euryarchaeota	464	2	Mrr	0	
1048	WP_204746780.1		Haloterrigena salifodinae	Euryarchaeota	465	1	HNH	1	1063
1049	ELY93140.1		Natrialba taiwanensis DSM 12281	Euryarchaeota	466	2	HNH	1	1064
1050	WP_160164226.1		Natrialba taiwanensis	Euryarchaeota	466	2	HNH	0	
1051	WP_092907621.1		Halostagnicola kamekurae	Euryarchaeota	467	1	HNH	1	1064
1052	WP_206876357.1		Corallococcus exiguus	Proteobacteria	468	1	Mrr	1	1181
1053	WP_020741386.1		Sorangium cellulosum	Proteobacteria	469	1	Mrr	1	1223
1054	MBL1200944.1		Nostoc sp. GBBB01	Cyanobacteria	470	1	Mrr	1	1115
1055	WP_015081596.1		Anabaena sp. 90	Cyanobacteria	471	1	Mrr	1	1115
1056	WP_015225263.1		Halothece sp. PCC 7418	Cyanobacteria	472	1	Short	1	1204
1057	TAG97156.1		Oscillatoriales cyanobacterium	Cyanobacteria	473	1	Nucleosidase	1	1193
1058	WP_105222173.1		Gloeocapsopsis sp. AAB1 = 1H9	Cyanobacteria	474	1	Short	1	1211
1059	WP_194027528.1		Lusitaniella coriacea	Cyanobacteria	475	1	Short	1	1201
1060	WP_136933689.1		Polyangium fumosum	Proteobacteria	476	1	Mrr	1	1248
1061	WP_153817453.1		Polyangium spumosum	Proteobacteria	477	1	Mrr	1	1241
1062	KYG02965.1		Sorangium cellulosum	Proteobacteria	478	1	Mrr	1	1281
1063	KYG08320.1		Sorangium cellulosum	Proteobacteria	479	1	Mrr	1	1352
1064	KYF77114.1		Sorangium cellulosum	Proteobacteria	480	1	Mrr	1	1311
1065	KYF88137.1		Sorangium cellulosum	Proteobacteria	481	1	Mrr	1	1285
1066	WP_020740786.1		Sorangium cellulosum	Proteobacteria	482	1	Mrr	1	1282
1067	WP_129353935.1		Sorangium cellulosum	Proteobacteria	483	1	Mrr	1	1283
1068	WP_129579733.1		Sorangium cellulosum	Proteobacteria	484	1	Mrr	1	1280
1069	KYF52360.1		Sorangium cellulosum	Proteobacteria	485	1	Mrr	1	1281
1070	WP_076763213.1		Domibacillus epiphyticus	Firmicutes	486	1	Mrr	1	1087
1071	WP_131533506.1		Pedobacter sp. RP-3-21	Bacteroidetes	487	1	Mrr	1	1062
1072	WP_133659106.1		Hymenobacter sp. UV11	Bacteroidetes	488	1	Mrr	1	1115
1073	WP_188752141.1		Parapedobacter defluvii	Bacteroidetes	489	1	Mrr	1	1067
1074	WP_182920692.1		Pedobacter planticolens	Bacteroidetes	490	1	Mrr	1	1078
1075	WP_074489327.1		Mucilaginibacter polytrichastri	Bacteroidetes	491	1	Mrr	1	1082
1076	WP_078351619.1		Mucilaginibacter pedocola	Bacteroidetes	492	1	Mrr	1	1090

1077	WP_188094334.1		Chitinophaga varians	Bacteroidetes	493	1	Mrr	1	1125
1078	WP_197429475.1		Flavobacterium rakeshii	Bacteroidetes	494	1	Mrr	1	1114
1079	WP_169192007.1		Chitinophaga sp. Ak27	Bacteroidetes	495	1	Mrr	1	1120
1080	WP_014218049.1		Niastella koreensis	Bacteroidetes	496	1	Mrr	1	1284
1081	HEK20310.1		Bacteroidetes bacterium	Bacteroidetes	497	1	Mrr	1	1291
1082	WP_130867371.1		Chryseobacterium indologenes	Bacteroidetes	498	1	Mrr	1	1098
1083	HAO07922.1		Chryseobacterium sp.	Bacteroidetes	499	1	Mrr	1	1088
1084	WP_076351134.1		Chryseobacterium joostei	Bacteroidetes	500	4	Mrr	1	1088
1085	WP_123861412.1		Chryseobacterium indologenes	Bacteroidetes	500	4	Mrr	0	
1086	WP_076394391.1		Chryseobacterium gambrini	Bacteroidetes	500	4	Mrr	0	
1087	WP_120334278.1		Sphingobacterium siyangense	Bacteroidetes	500	4	Mrr	0	
1088	WP_184555281.1		Chryseobacterium sediminis	Bacteroidetes	501	2	Mrr	1	1088
1089	WP_062648073.1		Chryseobacterium kwangjuense	Bacteroidetes	501	2	Mrr	0	
1090	WP_156196498.1		Labilibaculum euxinus	Bacteroidetes	502	1	Mrr	1	1126
1091	OJX90997.1		Paludibacter sp. 47-17	Bacteroidetes	503	1	Mrr	1	1117
1092	WP_209403806.1		Pseudozobellia sp. WGM2	Bacteroidetes	504	1	Mrr	1	1118
1093	WP_140996397.1		Flavobacterium sp. TP390	Bacteroidetes	505	1	Mrr	1	1109
1094	WP_123280325.1		Epilithonimonas hominis	Bacteroidetes	506	1	Mrr	1	1110
1095	WP_059154626.1		Elizabethkingia miricola	Bacteroidetes	507	1	Mrr	1	1109
1096	WP_188621552.1		Flavobacterium suaedae	Bacteroidetes	508	1	Mrr	1	1111
1097	WP_116093315.1		Chryseobacterium sp. 5_R23647	Bacteroidetes	509	1	Mrr	1	1111
1098	WP_088264195.1		Chryseobacterium haifense	Bacteroidetes	510	1	Mrr	1	1111
1099	WP_058699949.1		Myroides odoratimimus	Bacteroidetes	511	1	Mrr	1	1111
1100	WP_148796710.1		Bacillus marisflavi	Firmicutes	512	1	Mrr	1	1212
1101	WP_094942935.1		Tetzosporium hominis	Firmicutes	513	1	Mrr	1	1212
1102	WP_163204513.1		Clostridium sporogenes	Firmicutes	514	1	MBL + serine protease	1	1179
1103	WP_021127257.1		Paeniclostridium sordellii	Firmicutes	515	1	MBL + serine protease	1	1179
1104	WP_090089103.1		Clostridium uliginosum	Firmicutes	516	1	Mrr	1	1116
1105	WP_019849870.1		Desulfitobacterium sp. PCE1	Firmicutes	517	1	Mrr	1	1202
1106	MBS4024137.1		Clostridia bacterium	Firmicutes	518	1	Mrr	1	1179

1107	WP_098056183.1		Bacillus toyonensis	Firmicutes	519	1	Mrr	1	1195
1108	WP_097853922.1		Bacillus thuringiensis	Firmicutes	520	1	Mrr	1	1196
1109	WP_098834206.1		Bacillus thuringiensis	Firmicutes	521	7	Mrr	1	1196
1110	PEP91682.1		Bacillus toyonensis	Firmicutes	521	7	Mrr	0	
1111	WP_098257365.1		Bacillus toyonensis	Firmicutes	521	7	Mrr	0	
1112	WP_098277089.1		Bacillus thuringiensis	Firmicutes	521	7	Mrr	0	
1113	WP_098807313.1		Bacillus toyonensis	Firmicutes	521	7	Mrr	0	
1114	EJR31141.1		Bacillus mycoides	Firmicutes	521	7	Mrr	0	
1115	WP_162533053.1		Bacillus mycoides	Firmicutes	521	7	Mrr	0	
1116	WP_097960012.1		Bacillus cereus	Firmicutes	522	12	Mrr	1	1188
1117	WP_098843295.1		Bacillus thuringiensis	Firmicutes	522	12	Mrr	0	
1118	WP_179889673.1		Bacillus thuringiensis	Firmicutes	522	12	Mrr	0	
1119	WP_098836047.1		Bacillus thuringiensis	Firmicutes	522	12	Mrr	0	
1120	PGW64714.1		Bacillus cereus	Firmicutes	522	12	Mrr	0	
1121	WP_141556078.1		Bacillus cereus	Firmicutes	522	12	Mrr	0	
1122	WP_098841673.1		Bacillus thuringiensis	Firmicutes	522	12	Mrr	0	
1123	WP_200462662.1		Bacillus cereus	Firmicutes	522	12	Mrr	0	
1124	PHD23352.1		Bacillus wiedmannii	Firmicutes	522	12	Mrr	0	
1125	WP_142346284.1		Bacillus wiedmannii	Firmicutes	522	12	Mrr	0	
1126	WP_098866799.1		Bacillus thuringiensis	Firmicutes	522	12	Mrr	0	
1127	WP_053564078.1		Bacillus cereus	Firmicutes	522	12	Mrr	0	
1128	WP_142347996.1		Bacillus cereus	Firmicutes	523	6	Mrr	1	1188
1129	WP_142331183.1		Bacillus cereus	Firmicutes	523	6	Mrr	0	
1130	WP_098209392.1		Bacillus cereus	Firmicutes	523	6	Mrr	0	
1131	WP_098607625.1		Bacillus cereus	Firmicutes	523	6	Mrr	0	
1132	WP_098284878.1		Bacillus cereus	Firmicutes	523	6	Mrr	0	
1133	PGY28496.1		Bacillus cereus	Firmicutes	523	6	Mrr	0	
1134	MBD5137178.1		Lachnospiraceae bacterium	Firmicutes	524	1	Mrr	1	1083
1135	WP_164947982.1		Clostridium sp. FP1	Firmicutes	525	1	Mrr	1	1084
1136	WP_119196155.1		Dorea formicigenerans	Firmicutes	526	1	Mrr	1	1086
1137	WP_195553839.1		[Ruminococcus] gnavus	Firmicutes	527	1	Mrr	1	1081
1138	MBR5180411.1		Lachnospiraceae bacterium	Firmicutes	528	1	Mrr	1	1079
1139	NLI72036.1		Bacteroidales bacterium	Bacteroidetes	529	1	Mrr	1	1086
1140	WP_019229262.1		Sedimentibacter sp. B4	Firmicutes	530	1	Mrr	1	1093

1141	WP_110461196.1		Ruminiclostridium sufflavum	Firmicutes	531	1	Mrr	1	1091
1142	MBO5462940.1		Clostridia bacterium	Firmicutes	532	1	Mrr	1	1087
1143	MBQ7543163.1		Clostridia bacterium	Firmicutes	533	1	Mrr	1	1086
1144	MBQ8241672.1		Bacteroides sp.	Bacteroidetes	534	1	Mrr	1	1094
1145	MBQ2819116.1		Clostridia bacterium	Firmicutes	535	1	Mrr	1	1088
1146	WP_156062980.1		Collinsella aerofaciens	Actinobacteria	536	2	Mrr	1	1107
1147	WP_117855807.1		Collinsella sp. AF14-35	Actinobacteria	536	2	Mrr	0	
1148	WP_195361987.1		Collinsella aerofaciens	Actinobacteria	537	1	Mrr	1	1110
1149	WP_040656782.1		Olsenella sp. oral taxon 809	Actinobacteria	538	2	Mrr	1	1124
1150	EHF01190.1		Olsenella sp. oral taxon 809 str. F0356	Actinobacteria	538	2	Mrr	0	
1151	WP_070699280.1		Atopobium sp. HMSC064B08	Actinobacteria	539	1	Mrr	1	1123
1152	WP_021725373.1		Olsenella profusa	Actinobacteria	540	1	Mrr	1	1123
1153	MBR3356225.1		Solobacterium sp.	Firmicutes	541	1	Mrr	1	1119
1154	WP_173755178.1		Anaerostipes hadrus	Firmicutes	542	1	Mrr	1	1081
1155	MBS6168951.1		Firmicutes bacterium	Firmicutes	543	1	Mrr	1	1084
1156	MBD5096877.1		Lachnospiraceae bacterium	Firmicutes	544	1	Mrr	1	1088
1157	RGI68947.1		[Eubacterium] rectale	Firmicutes	545	7	Mrr	1	1088
1158	MBD9142417.1		[Eubacterium] rectale	Firmicutes	545	7	Mrr	0	
1159	RHA94546.1		[Eubacterium] rectale	Firmicutes	545	7	Mrr	0	
1160	CDC69391.1		Eubacterium rectale CAG:36	Firmicutes	545	7	Mrr	0	
1161	ACR75086.1		[Eubacterium rectale] ATCC 33656	Firmicutes	545	7	Mrr	0	
1162	RGT77023.1		[Eubacterium] rectale	Firmicutes	545	7	Mrr	0	
1163	RGK45736.1		[Eubacterium] rectale	Firmicutes	545	7	Mrr	0	
1164	WP_117509703.1		Blautia sp. AM23-13AC	Firmicutes	546	1	Mrr	1	1086
1165	WP_118158391.1		Ruminococcus sp. AM31-15AC	Firmicutes	547	1	Mrr	1	1065
1166	MBE5878914.1		Lachnospiraceae bacterium	Firmicutes	548	1	Mrr	1	1087
1167	OYZ34272.1		Campylobacterales bacterium 16-40-21	Proteobacteria	549	1	Short	1	1003
1168	PKL92936.1		Candidatus Goldbacteria bacterium HGW-Goldbacteria-1	Unclassified	550	1	Short	1	1095

1169	SYZ73761.1		candidate division Zixibacteria bacterium	candidate division Zixibacteria	551	1	Short	1	1113
1170	NVM55358.1		Candidatus Helarchaeota archaeon	Candidatus Helarchaeota	552	1	Short	1	1110
1171	OHD82122.1		Spirochaetes bacterium RIFOXYB1_FULL_32_8	Spirochaetes	553	1	Short	1	1116
1172	OGG13237.1		Candidatus Gottesmanbacteria bacterium RIFCSPHIGHO2_01_FULL_39_10	Unclassified	554	1	Short	1	1111
1173	HAU0512755.1		Legionella pneumophila	Proteobacteria	555	1	Short	1	1141
1174	WP_161517150.1		Legionella pneumophila	Proteobacteria	556	1	Short	1	1141
1175	WP_154694819.1		Legionella pneumophila	Proteobacteria	557	2	Short	1	1147
1176	WP_061637844.1		Legionella pneumophila	Proteobacteria	557	2	Short	0	
1177	HBD7469429.1		Legionella pneumophila	Proteobacteria	558	1	Short	1	1147
1178	PJI41847.1		Ferrovibrio sp.	Proteobacteria	559	1	Short	1	1172
1179	WP_133057880.1		Mycolicibacterium conceptionense	Actinobacteria	560	2	Short	1	1077
1180	ORV24701.1		Mycolicibacterium conceptionense	Actinobacteria	560	2	Short	0	
1181	WP_142392598.1		Mycobacterium sp. 3519A	Actinobacteria	561	1	Short	1	1075
1182	WP_204131904.1		Phycicoccus sp. MQZ13P-5	Actinobacteria	562	1	Short	1	1068
1183	WP_207006013.1		Nocardioides aromaticivorans	Actinobacteria	563	1	Short	1	1068
1184	TAK69202.1		Actinobacteria bacterium	Actinobacteria	564	1	Short	1	1068
1185	WP_006165197.1		Natrialba chahannaoensis	Euryarchaeota	565	1	Short	1	1070
1186	PKB83444.1		SAR202 cluster bacterium Io17-Chloro-G9	Chloroflexi	566	1	Short	1	1086
1187	MAN81008.1		Magnetovibrio sp.	Proteobacteria	567	1	Short	1	1130
1188	KQW72856.1		Phenylobacterium sp. Root1277	Proteobacteria	568	2	Short	1	1127
1189	WP_156378236.1		unclassified Phenylobacterium	Proteobacteria	568	2	Short	0	
1190	GFO83192.1		Methyloceanibacter sp.	Proteobacteria	569	1	Short	1	1113
1191	WP_191249976.1		Kordiimonas sediminis	Proteobacteria	570	1	Short	1	1116
1192	WP_135202880.1		Duganella callida	Proteobacteria	571	1	Short	1	1149

1193	MYA67891.1		Gammaproteobacteria bacterium	Proteobacteria	572	3	Short	1	1120
1194	MXX07861.1		Gammaproteobacteria bacterium	Proteobacteria	572	3	Short	0	
1195	MYK05209.1		Gammaproteobacteria bacterium	Proteobacteria	572	3	Short	0	
1196	AZD29366.1		Pseudomonas chlororaphis	Proteobacteria	573	3	Short	1	1120
1197	ETD37948.1		Pseudomonas chlororaphis subsp. aurantiaca PB-St2	Proteobacteria	573	3	Short	0	
1198	WP_031286254.1		Pseudomonas chlororaphis	Proteobacteria	573	3	Short	0	
1199	WP_134372531.1		Methylacidiphilum fumariolicum	Verrucomicrobia	574	3	Short	1	1115
1200	TFE68005.1		Methylacidiphilum fumariolicum	Verrucomicrobia	574	3	Short	0	
1201	WP_208526845.1		Methylacidiphilum fumariolicum	Verrucomicrobia	574	3	Short	0	
1202	MBI4911676.1		Acidobacteria bacterium	Acidobacteria	575	1	Short	1	1121
1203	WP_056261883.1		unclassified Pelomonas	Proteobacteria	576	1	Short	1	1119
1204	WP_127684259.1		Inhella crocodyli	Proteobacteria	577	1	Short	1	1119
1205	WP_111519994.1		Cupriavidus	Proteobacteria	578	2	Short	1	1120
1206	PZX26807.1		Cupriavidus alkaliphilus	Proteobacteria	578	2	Short	0	
1207	WP_150599071.1		Pandoraea fibrosis	Proteobacteria	579	1	Short	1	1120
1208	WP_169339836.1		Ralstonia insidiosa	Proteobacteria	580	1	Short	1	1120
1209	WP_064802386.1		Ralstonia insidiosa	Proteobacteria	581	2	Short	1	1120
1210	WP_102079905.1		Ralstonia mannitolilytica	Proteobacteria	581	2	Short	0	
1211	WP_124494266.1		Burkholderia contaminans	Proteobacteria	582	1	Short	1	1120
1212	WP_212250580.1		Burkholderia cenocepacia	Proteobacteria	583	1	Short	1	1120
1213	WP_175738074.1		Burkholderia ambifaria	Proteobacteria	584	1	Short	1	1120
1214	WP_174982329.1		Burkholderia lata	Proteobacteria	585	2	Short	1	1120
1215	VWC87547.1		Burkholderia lata	Proteobacteria	585	2	Short	0	
1216	WP_157782508.1		Virgibacillus halodenitrificans	Firmicutes	586	1	Mrr	1	1103
1217	WP_050184068.1		Domibacillus robiginosus	Firmicutes	587	1	Mrr	1	1109
1218	WP_098238855.1		Priestia megaterium	Firmicutes	588	1	Mrr	1	1110
1219	WP_158666627.1		Salinibacter altiplanensis	Bacteroidetes	589	1	Mrr	1	1105
1220	WP_166178564.1		Rubrobacter sp. SCSIO 52909	Actinobacteria	590	1	Mrr	1	1109
1221	WP_168499869.1		Anabaena sp. UHCC 0204	Cyanobacteria	591	1	Mrr	1	1095

1222	HHI30087.1		Candidatus Methanoperedenaceae archaeon	Euryarchaeota	592	1	Mrr	1	1066
1223	WP_073404447.1		Mucilaginibacter sp. OK098	Bacteroidetes	593	1	Mrr	1	1061
1224	WP_157262267.1		Pedobacter sp. PACM 27299	Bacteroidetes	594	2	Mrr	1	1063
1225	ALL05586.1		Pedobacter sp. PACM 27299	Bacteroidetes	594	2	Mrr	0	
1226	WP_184495502.1		Algoriphagus iocasae	Bacteroidetes	595	1	Mrr	1	1099
1227	WP_205638121.1		Algoriphagus aestuarii	Bacteroidetes	596	1	Mrr	1	1099
1228	MBS1759266.1		Bacteroidetes bacterium	Bacteroidetes	597	1	Mrr	1	1096
1229	WP_132118489.1		Arundinibacter roseus	Bacteroidetes	598	1	Mrr	1	1097
1230	WP_088583965.1		Chryseobacterium indologenes	Bacteroidetes	599	4	Mrr	1	1096
1231	WP_123871758.1		Chryseobacterium bernardetii	Bacteroidetes	599	4	Mrr	0	
1232	WP_047378083.1		Chryseobacterium sp. YR459	Bacteroidetes	599	4	Mrr	0	
1233	WP_130867394.1		Chryseobacterium indologenes	Bacteroidetes	599	4	Mrr	0	
1234	HAO07924.1		Chryseobacterium sp.	Bacteroidetes	600	1	Mrr	1	1096
1235	HCM35443.1		Chryseobacterium sp.	Bacteroidetes	601	1	Mrr	1	1096
1236	WP_014468876.1		Aliarcobacter butzleri	Proteobacteria	602	2	Mrr	1	1169
1237	WP_151947679.1		Aliarcobacter butzleri	Proteobacteria	602	2	Mrr	0	
1238	WP_128359253.1		Arcobacter venerupis	Proteobacteria	603	1	Mrr	1	1172
1239	WP_198855769.1		Flavobacterium sp. IB48	Bacteroidetes	604	2	Mrr	1	1059
1240	WP_072953499.1		Flavobacterium aquidurense	Bacteroidetes	604	2	Mrr	0	
1241	WP_126515696.1		Flavobacterium sp. RSP15	Bacteroidetes	605	1	Mrr	1	1063
1242	APA00341.1		Flavobacterium commune	Bacteroidetes	606	2	Mrr	1	1060
1243	WP_157499554.1		Flavobacterium commune	Bacteroidetes	606	2	Mrr	0	
1244	WP_014164638.1		Flavobacterium columnare	Bacteroidetes	607	7	Mrr	1	1060
1245	WP_195157910.1		Flavobacterium columnare	Bacteroidetes	607	7	Mrr	0	
1246	WP_195157345.1		Flavobacterium columnare	Bacteroidetes	607	7	Mrr	0	
1247	WP_107317784.1		Flavobacterium columnare	Bacteroidetes	607	7	Mrr	0	
1248	WP_077225221.1		Flavobacterium columnare	Bacteroidetes	607	7	Mrr	0	
1249	WP_195157749.1		Flavobacterium columnare	Bacteroidetes	607	7	Mrr	0	
1250	MBL7886829.1		Flavobacterium sp.	Bacteroidetes	607	7	Mrr	0	
1251	WP_132794494.1		Tenacibaculum skagerrakense	Bacteroidetes	608	1	Mrr	1	1065

1252	WP_035324763.1		Dokdonia donghaensis	Bacteroidetes	609	1	Mrr	1	1064
1253	WP_013752141.1		unclassified Dokdonia	Bacteroidetes	610	1	Mrr	1	1066
1254	HEC37766.1		bacterium	Unclassified	611	1	Mrr	1	1017
1255	NOT73992.1		Cyclobacteriaceae bacterium	Bacteroidetes	612	1	Mrr	1	1103
1256	WP_068891254.1		Pedobacter panaciterrae	Bacteroidetes	613	1	Mrr	1	1121
1257	SEO12119.1		bacterium A37T11	Unclassified	614	1	Mrr	1	1132
1258	WP_199076214.1		Pedobacter sp. ASV19	Bacteroidetes	615	1	Mrr	1	1135
1259	WP_205057390.1		Pedobacter terrae	Bacteroidetes	616	1	Mrr	1	1129
1260	WP_169610084.1		Mucilaginibacter sp. F39-2	Bacteroidetes	617	1	Mrr	1	1127
1261	WP_134091280.1		Olivibacter sp. XZL3	Bacteroidetes	618	1	Mrr	1	1127
1262	WP_192348835.1		Algoriphagus sp. Y33	Bacteroidetes	619	1	Mrr	1	1127
1263	PZR18281.1		Flavobacterium psychrophilum	Bacteroidetes	620	1	Transmembrane	1	1030
1264	WP_188953736.1		Filimonas zeae	Bacteroidetes	621	1	Transmembrane (5' split)	1	1043
1265	WP_166566084.1		Emticicia sp. CRIBPO	Bacteroidetes	622	1	Transmembrane (5' split)	1	1045
1266	WP_158798017.1		Pedobacter sp. L105	Bacteroidetes	623	1	Transmembrane (5' split)	1	1050
1267	WP_034925169.1		Gillisia sp. CAL575	Bacteroidetes	624	1	Transmembrane	1	1041
1268	WP_065318975.1		Polaribacter vadi	Bacteroidetes	625	1	Transmembrane	1	1074
1269	PLX10728.1		Marinilabiliales bacterium	Bacteroidetes	626	1	Transmembrane	1	1072
1270	WP_172282601.1		Lentimicrobium sp. S6	Bacteroidetes	627	1	Transmembrane (5' split)	1	1054
1271	WP_027124940.1		Gelidibacter mesophilus	Bacteroidetes	628	1	Transmembrane (5' split)	1	1050
1272	WP_084843389.1		Zunongwangia atlantica	Bacteroidetes	629	1	Transmembrane (5' split)	1	1052
1273	WP_036580987.1		Nonlabens ulvanivorans	Bacteroidetes	630	1	Transmembrane (5' split)	1	1049
1274	MBI2489978.1		Planctomycetes bacterium	Planctomycetes	631	1	Mrr	1	989
1275	WP_028844911.1		Thermodesulfovibrio thiophilus	Nitrospirae	632	1	Mrr	1	996

1276	OIN88582.1		Candidatus Aenigmarchaeota archaeon CG1_02_38_14	Candidatus Aenigmarchaeota	633	1	Mrr	1	999
1277	WP_113646138.1		Pedobacter chitinilyticus	Bacteroidetes	634	1	Mrr	1	993
1278	WP_040996281.1		Chryseobacterium oranimense	Bacteroidetes	635	1	Mrr	1	994
1279	WP_014165288.1		Flavobacterium columnare	Bacteroidetes	636	4	Mrr	1	993
1280	WP_065213521.1		Flavobacterium columnare	Bacteroidetes	636	4	Mrr	0	
1281	WP_192188518.1		Flavobacterium columnare	Bacteroidetes	636	4	Mrr	0	
1282	WP_146917101.1		Flavobacterium columnare	Bacteroidetes	636	4	Mrr	0	
1283	MAB49339.1		Flavobacteriaceae bacterium	Bacteroidetes	637	1	Mrr	1	995
1284	WP_072990090.1		Pseudozobellia thermophila	Bacteroidetes	638	1	Mrr	1	993
1285	MBT5305188.1		Candidatus Scalindua sp.	Planctomycetes	639	1	Mrr	1	998
1286	WP_093921000.1		Sunxiuqinia elliptica	Bacteroidetes	640	1	Mrr	1	1018
1287	WP_130305840.1		Ancylomarina subtilis	Bacteroidetes	641	1	Mrr	1	1012
1288	MBP5476569.1		Paludibacteraceae bacterium	Bacteroidetes	642	1	Mrr	1	966
1289	MBQ2886599.1		Alphaproteobacteria bacterium	Proteobacteria	643	1	Mrr	1	959
1290	WP_032566562.1		Bacteroidales	Bacteroidetes	644	4	Mrr	1	965
1291	WP_057252619.1		Bacteroides	Bacteroidetes	644	4	Mrr	0	
1292	WP_008157802.1		Bacteroidales	Bacteroidetes	644	4	Mrr	0	
1293	WP_181215015.1		Bacteroides sp. AGMB03916	Bacteroidetes	644	4	Mrr	0	
1294	NLI35284.1		Bacteroidales bacterium	Bacteroidetes	645	1	Mrr	1	962
1295	WP_184020395.1		Haloferula luteola	Verrucomicrobia	646	1	Mrr	1	1111
1296	MSU49148.1		Opitutus sp.	Verrucomicrobia	647	1	Mrr	1	1118
1297	WP_164111996.1		unclassified Sphingobacterium	Bacteroidetes	648	1	Mrr	1	1131
1298	WP_136841627.1		Pedobacter sp. RP-3-22	Bacteroidetes	649	1	Mrr	1	1123
1299	MAS87830.1		Micavibrio sp.	Proteobacteria	650	1	Mrr	1	1113
1300	MBL1146186.1		Proteobacteria bacterium	Proteobacteria	651	1	Mrr	1	1078
1301	WP_215434519.1		Akkermansia muciniphila	Verrucomicrobia	652	2	Mrr	1	1107
1302	QWO86833.1		Akkermansia muciniphila	Verrucomicrobia	652	2	Mrr	0	
1303	WP_215609154.1		Leptothoe spongobia	Cyanobacteria	653	1	Mrr	1	1043
1304	WP_065711813.1		Synechococcus sp. PCC 7117	Cyanobacteria	654	1	Mrr	1	1107
1305	RJO61764.1		candidate division WS5 bacterium	Unclassified	655	1	Mrr	1	1110

1306	MBO12230 26.1		Candidatus Scalindua sediminis	Planctomyce tes	656	1	Mrr	1	1107
1307	MBC27063 66.1		ANME-2 cluster archaeon	Euryarchaeot a	657	1	Mrr	1	1107
1308	WP_081150 028.1		Cognaticolwe llia beringensis	Proteobacteri a	658	1	Mrr	1	1051
1309	WP_089368 382.1		Pseudoaltero monas nigrifaciens	Proteobacteri a	659	1	Mrr	1	1043
1310	WP_121637 683.1		Pseudoaltero monas agarivorans	Proteobacteri a	660	1	Mrr	1	1044
1311	MBK89721 12.1		Hahellaceae bacterium	Proteobacteri a	661	1	Mrr	1	1064
1312	MXX59718. 1		Rhodotherma ceae bacterium	Bacteroidete s	662	3	Mrr	1	1065
1313	MXW81897 .1		Rhodotherma ceae bacterium	Bacteroidete s	662	3	Mrr	0	
1314	MYD20220. 1		Rhodotherma ceae bacterium	Bacteroidete s	662	3	Mrr	0	
1315	WP_027156 741.1		Methylobacter luteus	Proteobacteri a	663	1	Mrr	1	1060
1316	WP_053980 725.1		Marinagarivor ans algicola	Proteobacteri a	664	1	Mrr	1	1052
1317	WP_139449 036.1		Desulfobotulu s mexicanus	Proteobacteri a	665	1	Mrr	1	1051
1318	WP_010368 030.1		Pseudoaltero monas citrea	Proteobacteri a	666	2	Mrr	1	1053
1319	WP_039492 837.1		Pseudoaltero monas flavipulchra	Proteobacteri a	666	2	Mrr	0	
1320	WP_140266 215.1		Vibrio parahaemolyti cus	Proteobacteri a	667	5	Mrr	1	1044
1321	WP_045490 215.1		Vibrio harveyi	Proteobacteri a	667	5	Mrr	0	
1322	WP_029788 671.1		Vibrio harveyi	Proteobacteri a	667	5	Mrr	0	
1323	WP_140366 686.1		Vibrio parahaemolyti cus	Proteobacteri a	667	5	Mrr	0	
1324	WP_130250 003.1		Vibrio vulnificus	Proteobacteri a	667	5	Mrr	0	
1325	WP_101905 915.1		Vibrio	Proteobacteri a	668	2	Mrr	1	1044
1326	WP_119978 803.1		Shewanella algidipiscicol a	Proteobacteri a	668	2	Mrr	0	
1327	WP_171750 988.1		Vibrio sp. T3Y01	Proteobacteri a	669	1	Mrr	1	1044
1328	HBC352944 3.1		Vibrio parahaemolyti cus	Proteobacteri a	670	1	Mrr	1	1045
1329	EGR134643 3.1		Vibrio cholerae	Proteobacteri a	671	3	Mrr	1	1043
1330	EGR047878 4.1		Vibrio cholerae	Proteobacteri a	671	3	Mrr	0	
1331	EGR500370 3.1		Vibrio parahaemolyti cus	Proteobacteri a	671	3	Mrr	0	
1332	WP_029805 918.1		Vibrio parahaemolyti cus	Proteobacteri a	672	1	Mrr	1	1043

1333	GGI68536.1		Deinococcus wulumuqiensis	Deinococcus-Thermus	673	2	Mrr	1	1093
1334	WP_152423827.1		Deinococcus wulumuqiensis	Deinococcus-Thermus	673	2	Mrr	0	
1335	OGP56887.1		Deltaproteobacteria bacterium RBG_13_61_14	Proteobacteria	674	1	Mrr	1	1103
1336	MYC37125.1		Chloroflexi bacterium	Chloroflexi	675	1	Mrr	1	1047
1337	WP_212286554.1		Bradyrhizobium liaoningense	Proteobacteria	676	1	Mrr	1	1078
1338	WP_135557753.1		Paenibacillus cymbidii	Firmicutes	677	1	Mrr	1	1031
1339	WP_136672923.1		Marinifilum sp. JC120	Bacteroidetes	678	1	Mrr	1	1031
1340	WP_174559821.1		Acinetobacter bouvetii	Proteobacteria	679	1	Mrr	1	1017
1341	WP_057048882.1		Acinetobacter baumannii	Proteobacteria	680	1	Mrr	1	1015
1342	WP_000080421.1		Acinetobacter baumannii	Proteobacteria	681	6	Mrr	1	1015
1343	WP_128202560.1		Acinetobacter baumannii	Proteobacteria	681	6	Mrr	0	
1344	HAV421999 6.1		Acinetobacter baumannii	Proteobacteria	681	6	Mrr	0	
1345	WP_200027913.1		Acinetobacter seifertii	Proteobacteria	681	6	Mrr	0	
1346	WP_119025394.1		Acinetobacter soli	Proteobacteria	681	6	Mrr	0	
1347	WP_149930128.1		Acinetobacter soli	Proteobacteria	681	6	Mrr	0	
1348	EGY635727 2.1		Acinetobacter baumannii	Proteobacteria	682	1	Mrr	1	1015
1349	WP_087662164.1		Acinetobacter indicus	Proteobacteria	683	1	Mrr	1	1015
1350	WP_130073940.1		Acinetobacter piscicola	Proteobacteria	684	3	Mrr	1	1015
1351	WP_180126178.1		unclassified Acinetobacter	Proteobacteria	684	3	Mrr	0	
1352	WP_195728199.1		Acinetobacter johnsonii	Proteobacteria	684	3	Mrr	0	
1353	WP_122901978.1		Acinetobacter sp. B51(2017)	Proteobacteria	685	1	Mrr	1	1015
1354	WP_137926299.1		Cupriavidus sp. 2SB	Proteobacteria	686	1	Mrr	1	1028
1355	WP_077519320.1		Rhodanobacter sp. C03	Proteobacteria	687	2	Mrr	1	1027
1356	WP_056388800.1		Rhodanobacter sp. Soil772	Proteobacteria	687	2	Mrr	0	
1357	WP_049469219.1		Stenotrophomonas maltophilia	Proteobacteria	688	2	Mrr	1	1040
1358	VUP15463.1		Stenotrophomonas maltophilia	Proteobacteria	688	2	Mrr	0	
1359	WP_134734864.1		Stenotrophomonas maltophilia	Proteobacteria	689	1	Mrr	1	1040
1360	WP_045840334.1		Comamonas thiooxydans	Proteobacteria	690	2	Mrr	1	1063
1361	WP_209800461.1		Delftia lacustris	Proteobacteria	690	2	Mrr	0	
1362	WP_110134431.1		Achromobacter sp. RW408	Proteobacteria	691	2	Mrr	1	1029

1363	WP_053500925.1		Achromobacter xylosoxidans	Proteobacteria	691	2	Mrr	0	
1364	WP_088166803.1		Delftia sp. K82	Proteobacteria	692	1	Mrr	1	1029
1365	HAT260958 1.1		Kluyvera intermedia	Proteobacteria	693	1	Mrr	1	1036
1366	WP_063117472.1		Escherichia coli	Proteobacteria	694	34	Mrr	1	1036
1367	WP_177334520.1		Klebsiella variicola	Proteobacteria	694	34	Mrr	0	
1368	WP_159220346.1		Klebsiella pneumoniae	Proteobacteria	694	34	Mrr	0	
1369	WP_117036878.1		Klebsiella pneumoniae	Proteobacteria	694	34	Mrr	0	
1370	WP_060175600.1		Klebsiella pneumoniae	Proteobacteria	694	34	Mrr	0	
1371	WP_102097097.1		Klebsiella pneumoniae	Proteobacteria	694	34	Mrr	0	
1372	WP_040165753.1		Klebsiella pneumoniae	Proteobacteria	694	34	Mrr	0	
1373	WP_032445595.1		Klebsiella pneumoniae	Proteobacteria	694	34	Mrr	0	
1374	EFB447546 6.1		Escherichia coli	Proteobacteria	694	34	Mrr	0	
1375	EFM233723 4.1		Escherichia coli	Proteobacteria	694	34	Mrr	0	
1376	WP_097748412.1		Escherichia coli	Proteobacteria	694	34	Mrr	0	
1377	WP_116298561.1		Escherichia coli	Proteobacteria	694	34	Mrr	0	
1378	WP_085326284.1		Escherichia coli	Proteobacteria	694	34	Mrr	0	
1379	WP_181548915.1		Citrobacter freundii	Proteobacteria	694	34	Mrr	0	
1380	WP_119834394.1		Klebsiella michiganensis	Proteobacteria	694	34	Mrr	0	
1381	MBA31761 46.1		Salmonella enterica	Proteobacteria	694	34	Mrr	0	
1382	WP_025421271.1		Sodalis praecaptivus	Proteobacteria	694	34	Mrr	0	
1383	HBC053006 3.1		Escherichia coli	Proteobacteria	694	34	Mrr	0	
1384	WP_137445441.1		Escherichia coli	Proteobacteria	694	34	Mrr	0	
1385	WP_029570507.1		Pantoea ananatis	Proteobacteria	694	34	Mrr	0	
1386	WP_206311478.1		Escherichia coli	Proteobacteria	694	34	Mrr	0	
1387	WP_208002871.1		Pantoea agglomerans	Proteobacteria	694	34	Mrr	0	
1388	WP_114191453.1		Enterobacteriaceae	Proteobacteria	694	34	Mrr	0	
1389	QKJ88288.1		Erwiniaceae bacterium PD-1	Proteobacteria	694	34	Mrr	0	
1390	WP_108367840.1		Enterobacter hormaechei	Proteobacteria	694	34	Mrr	0	
1391	WP_009654281.1		Klebsiella	Proteobacteria	694	34	Mrr	0	
1392	EGT576015 2.1		Cronobacter sakazakii	Proteobacteria	694	34	Mrr	0	
1393	EBQ315597 5.1		Salmonella enterica	Proteobacteria	694	34	Mrr	0	
1394	WP_080791576.1		Klebsiella pneumoniae	Proteobacteria	694	34	Mrr	0	
1395	WP_101868127.1		Klebsiella michiganensis	Proteobacteria	694	34	Mrr	0	
1396	EBF951682 7.1		Salmonella enterica	Proteobacteria	694	34	Mrr	0	

			subsp. enterica						
1397	MBH2682675.1		Serratia marcescens	Proteobacteria	694	34	Mrr	0	
1398	WP_149571749.1		Serratia marcescens	Proteobacteria	694	34	Mrr	0	
1399	ECK7274223.1		Salmonella enterica subsp. enterica	Proteobacteria	694	34	Mrr	0	
1400	WP_145585158.1		Yersinia intermedia	Proteobacteria	695	1	Mrr	1	1035
1401	WP_103860968.1		Pectobacterium versatile	Proteobacteria	696	19	Mrr	1	1036
1402	WP_201489512.1		Escherichia coli	Proteobacteria	696	19	Mrr	0	
1403	MBA8397036.1		Escherichia coli	Proteobacteria	696	19	Mrr	0	
1404	MBL1007972.1		Escherichia coli	Proteobacteria	696	19	Mrr	0	
1405	WP_126723113.1		Escherichia coli	Proteobacteria	696	19	Mrr	0	
1406	EFC1481673.1		Escherichia coli	Proteobacteria	696	19	Mrr	0	
1407	WP_097464621.1		Escherichia coli	Proteobacteria	696	19	Mrr	0	
1408	WP_113401561.1		Escherichia coli	Proteobacteria	696	19	Mrr	0	
1409	HAW1516613.1		Escherichia coli	Proteobacteria	696	19	Mrr	0	
1410	EFK4234508.1		Escherichia coli	Proteobacteria	696	19	Mrr	0	
1411	WP_127678941.1		Escherichia coli	Proteobacteria	696	19	Mrr	0	
1412	WP_103941323.1		Pectobacterium versatile	Proteobacteria	696	19	Mrr	0	
1413	ELG93155.1		Escherichia coli KTE146	Proteobacteria	696	19	Mrr	0	
1414	WP_127032142.1		Pectobacterium versatile	Proteobacteria	696	19	Mrr	0	
1415	WP_205713530.1		Escherichia	Proteobacteria	696	19	Mrr	0	
1416	WP_039270025.1		Enterobacteriaceae	Proteobacteria	696	19	Mrr	0	
1417	MBT1840195.1		Enterobacter hormaechei subsp. xiangfangensis	Proteobacteria	696	19	Mrr	0	
1418	WP_098363265.1		Klebsiella sp. KG9	Proteobacteria	696	19	Mrr	0	
1419	WP_095784585.1		Escherichia coli	Proteobacteria	696	19	Mrr	0	
1420	WP_122581035.1		Pseudomonas viridiflava	Proteobacteria	697	2	Mrr	1	1025
1421	WP_122593180.1		Pseudomonas viridiflava	Proteobacteria	697	2	Mrr	0	
1422	WP_076226250.1		Chromobacterium violaceum	Proteobacteria	698	2	Mrr	1	1031
1423	MBM2886400.1		Chromobacterium phragmitis	Proteobacteria	698	2	Mrr	0	
1424	WP_167567395.1		Aeromonas veronii	Proteobacteria	699	1	Mrr	1	1024
1425	WP_100645444.1		Aeromonas veronii	Proteobacteria	700	1	Mrr	1	1028
1426	WP_139424474.1		Aeromonas sobria	Proteobacteria	701	1	Mrr	1	1028
1427	WP_016349989.1		Aeromonas	Proteobacteria	702	4	Mrr	1	1027

1428	WP_108614109.1		Aeromonas sp. HMWF015	Proteobacteria	702	4	Mrr	0	
1429	WP_201987958.1		Aeromonas veronii	Proteobacteria	702	4	Mrr	0	
1430	WP_075115762.1		Aeromonas veronii	Proteobacteria	702	4	Mrr	0	
1431	WP_144716086.1		Aeromonas veronii	Proteobacteria	703	1	Mrr	1	1027
1432	WP_024779387.1		Pseudomonas corrugata	Proteobacteria	704	1	Mrr	1	1032
1433	WP_096650554.1		Halomonas nigrificans	Proteobacteria	705	1	Mrr	1	1027
1434	WP_133732835.1		Halomonas alkaliantarctica	Proteobacteria	706	1	Mrr	1	1026
1435	WP_199284173.1		Halomonas sp. A40-4	Proteobacteria	707	1	Mrr	1	1027
1436	WP_046735329.1		Xanthomonas	Proteobacteria	708	5	Mrr	1	1028
1437	WP_201061517.1		unclassified Pseudomonas	Proteobacteria	708	5	Mrr	0	
1438	WP_127446580.1		Xanthomonas	Proteobacteria	708	5	Mrr	0	
1439	WP_078589616.1		[Pseudomonas] cissicola	Proteobacteria	708	5	Mrr	0	
1440	WP_095574779.1		Xanthomonas	Proteobacteria	708	5	Mrr	0	
1441	WP_208665204.1		Pseudomonas thivervalensis	Proteobacteria	709	1	Mrr	1	1024
1442	WP_192485541.1		Pseudomonas savastanoi	Proteobacteria	710	2	Mrr	1	1023
1443	WP_117140574.1		Pseudomonas syringae group	Proteobacteria	710	2	Mrr	0	
1444	WP_200601120.1		Pseudomonas fluorescens	Proteobacteria	711	1	Mrr	1	1023
1445	WP_122098504.1		Pseudomonas songnenensis	Proteobacteria	712	1	Mrr	1	1028
1446	WP_115063812.1		Oligella ureolytica	Proteobacteria	713	1	Mrr	1	1036
1447	WP_085688200.1		unclassified Pseudomonas	Proteobacteria	714	1	Mrr	1	1028
1448	WP_123252034.1		Dickeya undicola	Proteobacteria	715	1	Mrr	1	1023
1449	HAT1608142.1		Raoultella planticola	Proteobacteria	716	1	Mrr	1	1036
1450	WP_039503414.1		Pectobacterium	Proteobacteria	717	1	Mrr	1	1037
1451	WP_189760768.1		Xenorhabdus griffiniae	Proteobacteria	718	2	Mrr	1	1037
1452	MBN5402352.1		Serratia marcescens	Proteobacteria	718	2	Mrr	0	
1453	WP_186062394.1		Burkholderia gladioli	Proteobacteria	719	1	Mrr	1	1028
1454	WP_093184735.1		Thiocapsa sp. KS1	Proteobacteria	720	1	Mrr	1	1027
1455	WP_171434995.1		Pseudomonas aeruginosa	Proteobacteria	721	4	Mrr	1	1028
1456	WP_030047295.1		Pseudomonas aeruginosa	Proteobacteria	721	4	Mrr	0	
1457	MBG4631804.1		Pseudomonas aeruginosa	Proteobacteria	721	4	Mrr	0	
1458	WP_174713098.1		Pseudomonas aeruginosa	Proteobacteria	721	4	Mrr	0	
1459	TQV64419.1		Halothiobacillaceae bacterium	Proteobacteria	722	1	Mrr	1	1028
1460	MBS4095874.1		Sulfuricella sp.	Proteobacteria	723	1	Mrr	1	1028

1461	WP_191487193.1		Pseudomonas sp. FEN	Proteobacteria	724	1	Mrr	1	1028
1462	WP_175025944.1		Burkholderia	Proteobacteria	725	1	Mrr	1	1028
1463	WP_099387245.1		Janthinobacterium sp. BJB301	Proteobacteria	726	1	Mrr	1	1028
1464	WP_204496001.1		Burkholderia	Proteobacteria	727	1	Mrr	1	1028
1465	RMM38886.1		Pseudomonas syringae pv. aptata	Proteobacteria	728	2	Mrr	1	1028
1466	WP_122267426.1		Pseudomonas syringae	Proteobacteria	728	2	Mrr	0	
1467	WP_119429236.1		Pseudomonas sp. 91RF	Proteobacteria	729	1	Mrr	1	1028
1468	MBP5786684.1		Kiritimatiellae bacterium	Kiritimatiellaeota	730	1	Mrr	1	1039
1469	MBN69236.1		Gimesia sp.	Planctomycetes	731	1	Mrr	1	1026
1470	MBL8874077.1		Planctomycetaceae bacterium	Planctomycetes	732	1	Mrr	1	1026
1471	MXX47636.1		Chloroflexi bacterium	Chloroflexi	733	1	Mrr	1	1028
1472	WP_146065609.1		Arthrobacter pityocampae	Actinobacteria	734	1	Mrr	1	1026
1473	WP_138605381.1		Pseudarthrobacter sp. NamE5	Actinobacteria	735	1	Mrr	1	1025
1474	WP_067896180.1		Actinomadura chibensis	Actinobacteria	736	1	Mrr	1	1039
1475	WP_104786653.1		Streptomyces	Actinobacteria	737	4	Mrr	1	1022
1476	WP_123543911.1		Streptomyces ossamyceticus	Actinobacteria	737	4	Mrr	0	
1477	WP_097287493.1		Streptomyces sp. OK228	Actinobacteria	737	4	Mrr	0	
1478	WP_119103341.1		Streptomyces sporangiiformans	Actinobacteria	737	4	Mrr	0	
1479	WP_131627332.1		Frankia sp. BMG5.11	Actinobacteria	738	1	Mrr	1	1034
1480	WP_139573972.1		Microbispora catharanthi	Actinobacteria	739	1	Mrr	1	1024
1481	SNT62830.1		Streptosporangium subroseum	Actinobacteria	740	2	Mrr	1	1022
1482	WP_143653629.1		Streptosporangium subroseum	Actinobacteria	740	2	Mrr	0	
1483	WP_158984337.1		Streptomyces sp. QHH-9511	Actinobacteria	741	2	Mrr	1	1022
1484	QGZ47421.1		Streptomyces sp. QHH-9511	Actinobacteria	741	2	Mrr	0	
1485	WP_097243097.1		Streptomyces	Actinobacteria	742	1	Mrr	1	1022
1486	WP_086754806.1		Streptomyces scabiei	Actinobacteria	743	1	Mrr	1	1020
1487	WP_055566890.1		Streptomyces atriruber	Actinobacteria	744	1	Mrr	1	1024
1488	WP_165343494.1		Streptomyces ureilyticus	Actinobacteria	745	1	Mrr	1	1020
1489	WP_093910335.1		Streptomyces sp. cf386	Actinobacteria	746	1	Mrr	1	1023
1490	WP_121720394.1		Streptomyces sp. E2N171	Actinobacteria	747	4	Mrr	1	1023
1491	GHC14845.1		Streptomyces albogriseolus	Actinobacteria	747	4	Mrr	0	

1492	WP_215188415.1		Streptomyces sp. McG8	Actinobacteria	747	4	Mrr	0	
1493	WP_215048734.1		Streptomyces sp. McG7	Actinobacteria	747	4	Mrr	0	
1494	WP_185906768.1		Streptomyces	Actinobacteria	748	2	Mrr	1	1023
1495	MBJ6622143.1		Streptomyces sp. DHE17-7	Actinobacteria	748	2	Mrr	0	
1496	WP_193500735.1		Streptomyces bellus	Actinobacteria	749	1	Mrr	1	1023
1497	WP_037920938.1		Streptomyces sp. PCS3-D2	Actinobacteria	750	1	Mrr	1	1022
1498	WP_129841551.1		Streptomyces sp. RFCAC02	Actinobacteria	751	1	Mrr	1	1023
1499	OQR64821.1		Streptomyces sp. B9173	Actinobacteria	752	1	Mrr	1	1023
1500	WP_184905372.1		Streptomyces scabiei	Actinobacteria	753	4	Mrr	1	1023
1501	WP_103509206.1		Streptomyces sp. SM13	Actinobacteria	753	4	Mrr	0	
1502	WP_210881745.1		Streptomyces sp. BH-SS-21	Actinobacteria	753	4	Mrr	0	
1503	WP_066954789.1		Streptomyces lushanensis	Actinobacteria	753	4	Mrr	0	
1504	WP_043727622.1		Kutzneria sp. 744	Actinobacteria	754	1	Mrr	1	1017
1505	WP_037336552.1		Amycolatopsis sp. MJM2582	Actinobacteria	755	1	Mrr	1	1017
1506	WP_116043498.1		Amycolatopsis palatopharyngis	Actinobacteria	756	1	Mrr	1	1017
1507	MBI5068377.1		Deltaproteobacteria bacterium	Proteobacteria	757	1	Mrr	1	1046
1508	MBU2829755.1		Acidithiobacillus ferriphilus	Proteobacteria	758	1	Mrr	1	1031
1509	KYF63561.1		Sorangium cellulosum	Proteobacteria	759	1	Mrr	1	1056
1510	WP_167230681.1		Pelagibius litoralis	Proteobacteria	760	1	Mrr	1	1053
1511	WP_163560422.1		Halomonas sp. NO4	Proteobacteria	761	1	Mrr	1	1008
1512	MBD1853165.1		Cyanobacteria bacterium FACHB-502	Cyanobacteria	762	1	Mrr	1	1020
1513	MBI2571104.1		Candidatus Schekmanbacteria bacterium	Unclassified	763	1	Mrr	1	1054
1514	WP_102653316.1		Halomonas endophytica	Proteobacteria	764	1	Mrr	1	1025
1515	NCC26742.1		Gammaproteobacteria bacterium	Proteobacteria	765	1	Mrr	1	1024
1516	MBS0960391.1		Acetobacter thailandicus	Proteobacteria	766	2	Mrr	1	1020
1517	WP_212341573.1		Acetobacter thailandicus	Proteobacteria	766	2	Mrr	0	
1518	RON73007.1		Pseudomonas fluorescens	Proteobacteria	767	1	Mrr	1	1057
1519	WP_147443123.1		Corallococcus sp. AB011P	Proteobacteria	768	2	Mrr	1	1030
1520	RKG60459.1		Corallococcus sp. AB011P	Proteobacteria	768	2	Mrr	0	
1521	WP_169825153.1		Corallococcus exiguus	Proteobacteria	769	1	Mrr	1	1030
1522	MAX25690.1		Phycisphaeraceae bacterium	Planctomycetes	770	1	Mrr	1	1053
1523	NLX95227.1		Rhodopirellula sp.	Planctomycetes	771	1	Mrr	1	1023

1524	TXH91661.1		Thauera aminoaromatica	Proteobacteria	772	1	Mrr	1	1013
1525	WP_124778846.1		Micromonospora	Actinobacteria	773	2	Mrr	1	1024
1526	ODB75071.1		Micromonospora sp. II	Actinobacteria	773	2	Mrr	0	
1527	WP_047893588.1		Micromonospora sp. RV43	Actinobacteria	774	1	Mrr	1	1016
1528	GGJ90402.1		Pilimelia anulata	Actinobacteria	775	2	Mrr	1	1018
1529	WP_189169794.1		Pilimelia anulata	Actinobacteria	775	2	Mrr	0	
1530	WP_149768182.1		Mumia	Actinobacteria	776	1	Mrr	1	1095
1531	WP_024801657.1		Nocardia sp. BMG51109	Actinobacteria	777	1	Mrr	1	1076
1532	WP_171673201.1		Kribbella sandramycini	Actinobacteria	778	1	Mrr	1	1064
1533	WP_157526394.1		Microbacterium oleivorans	Actinobacteria	779	1	Mrr	1	1103
1534	WP_094765105.1		Luteococcus japonicus	Actinobacteria	780	1	Mrr	1	1119
1535	WP_206473964.1		Dietzia sp. KRD202	Actinobacteria	781	1	Mrr	1	1122
1536	WP_198671575.1		Desertihabitans aurantiacus	Actinobacteria	782	1	Mrr	1	1116
1537	WP_171584013.1		Arthrobacter sp. 260	Actinobacteria	783	1	Mrr	1	1121
1538	WP_125778404.1		Antribacter gilvus	Actinobacteria	784	1	Mrr	1	1120
1539	EUA27791.1		Mycobacterium intracellulare	Actinobacteria	785	1	Mrr	1	998
1540	RYE39962.1		Hyphomicrobiales bacterium	Proteobacteria	786	1	Mrr	1	1006
1541	WP_143660831.1		Streptomyces sp. st140	Actinobacteria	787	1	Mrr	1	1005
1542	WP_165230118.1		Streptomyces sp. A7024	Actinobacteria	788	1	Mrr	1	1004
1543	WP_186582000.1		Streptomyces sp. BK340	Actinobacteria	789	1	Mrr	1	1002
1544	WP_164432263.1		Streptomyces sp. SID13726	Actinobacteria	790	1	Mrr	1	1002
1545	WP_165954929.1		Streptomyces sp. BK335	Actinobacteria	791	1	Mrr	1	1002
1546	WP_032769585.1		Streptomyces sp. CNS654	Actinobacteria	792	1	Mrr	1	1002
1547	WP_194274901.1		Streptomyces olivaceus	Actinobacteria	793	1	Mrr	1	1003
1548	WP_103492588.1		unclassified Streptomyces	Actinobacteria	794	1	Mrr	1	1008
1549	SMD24961.1		Lentzea albidocapillata	Actinobacteria	795	2	Short	1	1169
1550	WP_144065777.1		Lentzea albidocapillata	Actinobacteria	795	2	Short	0	
1551	WP_167975259.1		Lentzea indica	Actinobacteria	796	1	Cap4	1	1156
1552	WP_157495469.1		Kutzneria sp. 744	Actinobacteria	797	1	Mrr	1	1241
1553	WP_089014414.1		Micromonospora inositola	Actinobacteria	798	1	Mrr	1	1192
1554	SCG45733.1		Micromonospora zamorensis	Actinobacteria	799	2	Mrr	1	1192
1555	WP_088947554.1		Micromonospora zamorensis	Actinobacteria	799	2	Mrr	0	

1556	MBJ752017 4.1		Solirubrobact eraceae bacterium	Actinobacter ia	800	1	Mrr	1	1098
1557	MXX38720. 1		Gemmatimon adetes bacterium	Gemmatimo nadetes	801	1	Mrr	1	1098
1558	WP_139191 303.1		Rhodococcus tukisamuensis	Actinobacter ia	802	1	Mrr	1	1130
1559	RZL82155.1		Rhodococcus sp.	Actinobacter ia	803	1	Mrr	1	1128
1560	MBA23122 92.1		Actinobacteri a bacterium	Actinobacter ia	804	1	Mrr	1	1183
1561	WP_055346 437.1		Propionibacte rium freudenreichii	Actinobacter ia	805	1	Mrr	1	1181
1562	WP_062104 296.1		Cellulomonas sp. B6	Actinobacter ia	806	1	Mrr	1	1175
1563	WP_144081 713.1		Micromonosp ora wenchangensi s	Actinobacter ia	807	2	Mrr	1	1169
1564	OWV04714. 1		Micromonosp ora wenchangensi s	Actinobacter ia	807	2	Mrr	0	
1565	WP_143133 865.1		Actinoplanes philippinensis	Actinobacter ia	808	2	Mrr	1	1169
1566	GIE81715.1		Actinoplanes philippinensis	Actinobacter ia	808	2	Mrr	0	
1567	WP_204029 783.1		Micromonosp ora trujilloniae	Actinobacter ia	809	2	Mrr	1	1169
1568	GIJ42927.1		Micromonosp ora trujilloniae	Actinobacter ia	809	2	Mrr	0	
1569	WP_146102 188.1		Nocardia nova	Actinobacter ia	810	1	Mrr	1	1187
1570	WP_084750 964.1		Kitasatospora albolonga	Actinobacter ia	811	1	Mrr	1	1173
1571	WP_215060 624.1		Streptomyces sp. ISL-86	Actinobacter ia	812	1	Mrr	1	1169
1572	WP_148901 648.1		Streptomyces parvus	Actinobacter ia	813	2	Mrr	1	1170
1573	WP_097983 675.1		Streptomyces sp. f150	Actinobacter ia	813	2	Mrr	0	
1574	WP_136101 494.1		Streptomyces sp. S816	Actinobacter ia	814	1	Mrr	1	1171
1575	WP_189191 879.1		Streptomyces albiflavescens	Actinobacter ia	815	2	Mrr	1	1174
1576	GGN92036. 1		Streptomyces albiflavescens	Actinobacter ia	815	2	Mrr	0	
1577	WP_215455 082.1		Streptomyces sp. ATCC 21386	Actinobacter ia	816	1	Mrr	1	1170
1578	WP_141754 758.1		Corynebacteri um sp. HMSC069E0 4	Actinobacter ia	817	1	Mrr	1	1099
1579	WP_014741 707.1		Modestobacte r marinus	Actinobacter ia	818	1	Mrr	1	1070
1580	WP_167582 610.1		Kineococcus rubinsiae	Actinobacter ia	819	1	Mrr	1	1084
1581	WP_145942 400.1		Corynebacteri um glyciniphilum	Actinobacter ia	820	1	Mrr	1	1066
1582	WP_194706 557.1		Nocardioides islandensis	Actinobacter ia	821	1	Mrr	1	1104
1583	WP_060700 552.1		Glutamicibact er halophytocola	Actinobacter ia	822	1	Mrr	1	1081

1584	WP_131817136.1		Nocardia salmonicida	Actinobacteria	823	1	Mrr	1	1105
1585	MBM4523045.1		Rhodococcus hoagii	Actinobacteria	824	1	Mrr	1	1076
1586	WP_134402295.1		Cryobacterium sp. TMT4-10	Actinobacteria	825	1	Mrr	1	1224
1587	RSM80919.1		Kibdelosporangium aridum	Actinobacteria	826	2	Mrr	1	1303
1588	WP_125727664.1		Kibdelosporangium aridum	Actinobacteria	826	2	Mrr	0	
1589	WP_016828401.1		Streptomyces viridosporus	Actinobacteria	827	1	Mrr	1	1297
1590	WP_167986622.1		Streptomyces sp. PRB2-1	Actinobacteria	828	1	Mrr	1	1296
1591	WP_161322877.1		Streptomyces sp. SID5614	Actinobacteria	829	1	Mrr	1	1289
1592	WP_097955005.1		unclassified Streptomyces	Actinobacteria	830	1	Mrr	1	1290
1593	NEE23013.1		Streptomyces sp. SID7499	Actinobacteria	831	1	Mrr	1	1319
1594	SFM55451.1		Streptomyces sp. cf124	Actinobacteria	832	2	Mrr	1	1301
1595	WP_093788890.1		Streptomyces sp. cf124	Actinobacteria	832	2	Mrr	0	
1596	WP_128846456.1		Streptomyces lavendulae	Actinobacteria	833	1	Trypco 2 + serine proteas e	1	1300
1597	ALO05647.1		Streptomyces venezuelae	Actinobacteria	834	2	Mrr	1	1307
1598	WP_079037822.1		Streptomyces gardneri	Actinobacteria	834	2	Mrr	0	
1599	AEN07982.1		Streptomyces sp. SirexAA-E	Actinobacteria	835	2	Mrr	1	1301
1600	WP_106707985.1		unclassified Streptomyces	Actinobacteria	835	2	Mrr	0	
1601	WP_097958015.1		Streptomyces sp. wa1002	Actinobacteria	836	2	Mrr	1	1305
1602	WP_097923353.1		Streptomyces sp. wa1063	Actinobacteria	836	2	Mrr	0	
1603	WP_067229856.1		Streptomyces longwoodensis	Actinobacteria	837	2	Mrr	1	1296
1604	MYU31131.1		Streptomyces sp. SID7810	Actinobacteria	837	2	Mrr	0	
1605	WP_124445852.1		Streptomyces sp. NL15-2K	Actinobacteria	838	2	Mrr	1	1296
1606	WP_141580792.1		Streptomyces ipomoeae	Actinobacteria	838	2	Mrr	0	
1607	WP_127892364.1		Streptomyces sp. S10(2018)	Actinobacteria	839	1	Mrr	1	1296
1608	WP_091447969.1		Geodermatophilus poikilotrophus	Actinobacteria	840	1	Mrr	1	1222
1609	WP_153362237.1		Modestobacter roseus	Actinobacteria	841	1	Mrr	1	1201
1610	WP_179556978.1		Pseudoclavibacter sp. JAI123	Actinobacteria	842	1	Mrr	1	1218
1611	WP_162321452.1		Nesterenkonia haasae	Actinobacteria	843	1	Mrr	1	1214
1612	WP_156368218.1		Aeromicrobium sp. Leaf289	Actinobacteria	844	1	Mrr	1	1214
1613	WP_178010233.1		Microbacterium oleivorans	Actinobacteria	845	1	Mrr	1	1056
1614	WP_090820965.1		Arthrobacter sp. yr096	Actinobacteria	846	1	Mrr	1	1062

1615	WP_163511179.1		Fodinicola acaciae	Actinobacteria	847	1	Mrr	1	1063
1616	WP_184846918.1		Allocatelliglobosispora scoriae	Actinobacteria	848	1	Mrr	1	1060
1617	WP_147432938.1		Catellatospora citrea	Actinobacteria	849	1	Mrr	1	1062
1618	WP_184814741.1		Actinophytocola algeriensis	Actinobacteria	850	1	Mrr	1	1060
1619	WP_148311466.1		Amycolatopsis japonica	Actinobacteria	851	1	Mrr	1	1059
1620	WP_148344011.1		Actinomadura chibensis	Actinobacteria	852	1	Mrr	1	1060
1621	WP_143120957.1		Streptosporangium canum	Actinobacteria	853	2	Mrr	1	1062
1622	SFJ52788.1		Streptosporangium canum	Actinobacteria	853	2	Mrr	0	
1623	SEL14208.1		Nonomuraea pusilla	Actinobacteria	854	2	Mrr	1	1062
1624	WP_143078604.1		Nonomuraea pusilla	Actinobacteria	854	2	Mrr	0	
1625	MXW76923.1		Acidimicrobiaceae bacterium	Actinobacteria	855	2	Mrr	1	1045
1626	MXW61276.1		Acidimicrobiaceae bacterium	Actinobacteria	855	2	Mrr	0	
1627	MXW62031.1		Acidimicrobiaceae bacterium	Actinobacteria	856	1	Mrr	1	1051
1628	WP_141280135.1		Pseudonocardia hydrocarbonoxydans	Actinobacteria	857	1	Mrr	1	1043
1629	WP_156051616.1		Allokutzneria albata	Actinobacteria	858	1	Mrr	1	1044
1630	WP_150989764.1		Bacillus cereus	Firmicutes	859	1	Mrr	1	1074
1631	WP_038370686.1		Brachybacterium phenoliresistens	Actinobacteria	860	1	Mrr	1	1055
1632	WP_156968531.1		Cellulomonas bogoriensis	Actinobacteria	861	2	Mrr	1	1055
1633	KGM09736.1		Cellulomonas bogoriensis 69B4 = DSM 16987	Actinobacteria	861	2	Mrr	0	
1634	WP_029458575.1		Dietzia alimentaria	Actinobacteria	862	1	Mrr	1	1051
1635	WP_133833309.1		Arthrobacter sp. JUb115	Actinobacteria	863	1	Mrr	1	1047
1636	WP_192046462.1		Frigoribacterium sp. CFBP 8754	Actinobacteria	864	1	Mrr	1	1041
1637	WP_173153836.1		unclassified Frigoribacterium	Actinobacteria	865	1	Mrr	1	1039
1638	WP_162722408.1		Microbacterium sp. PM5	Actinobacteria	866	2	Mrr	1	1040
1639	AXA95852.1		Microbacterium sp. PM5	Actinobacteria	866	2	Mrr	0	
1640	WP_152642149.1		Microbacterium azadirachtae	Actinobacteria	867	1	Mrr	1	1043
1641	WP_156228638.1		Corynebacterium sp. 2019	Actinobacteria	868	2	Mrr	1	1043
1642	QGU05159.1		Corynebacterium sp. 2019	Actinobacteria	868	2	Mrr	0	

1643	WP_043056071.1		Enteractinococcus helveticum	Actinobacteria	869	1	Mrr	1	1142
1644	WP_094657849.1		Rhodococcus sp. 06-418-5	Actinobacteria	870	1	Mrr	1	1042
1645	WP_094732913.1		Rhodococcus sp. 15-1154-1	Actinobacteria	871	1	Mrr	1	1042
1646	WP_003849268.1		Corynebacterium ammoniagenes	Actinobacteria	872	1	Mrr	1	1080
1647	WP_156489204.1		Corynebacterium stationis	Actinobacteria	873	2	Mrr	1	1045
1648	OAH26516.1		Corynebacterium stationis	Actinobacteria	873	2	Mrr	0	
1649	WP_144228655.1		Rariglobus hedericola	Verrucomicrobia	874	1	Mrr	1	1116
1650	WP_043582220.1		Geminisphaera colitermitum	Verrucomicrobia	875	1	Mrr	1	1145
1651	MSM38673.1		Geobacter sp.	Proteobacteria	876	1	CMP hydrolase	1	1108
1652	WP_164009894.1		Pyxidicoccus trucidator	Proteobacteria	877	1	CMP hydrolase	1	1102
1653	WP_151126618.1		Geobacter sp. Red96	Proteobacteria	878	1	CMP hydrolase	1	1108
1654	OZA14959.1		Hydrogenophilales bacterium 17-62-8	Proteobacteria	879	1	CMP hydrolase	1	1123
1655	ACV37622.1		Candidatus Accumulibacter phosphatis clade IIA str. UW-1	Proteobacteria	880	2	CMP hydrolase	1	1108
1656	WP_199931525.1		Candidatus Accumulibacter phosphatis	Proteobacteria	880	2	CMP hydrolase	0	
1657	RFC34541.1		Candidatus Nitrotoga sp. MKT	Proteobacteria	881	1	CMP hydrolase	1	1108
1658	NDU85873.1		Ferrovum sp.	Proteobacteria	882	1	CMP hydrolase	1	1106
1659	NOT19915.1		Sideroxydans sp.	Proteobacteria	883	1	CMP hydrolase	1	1101
1660	THG79505.1		Pseudomonas sp. A-1	Proteobacteria	884	2	CMP hydrolase	1	1112
1661	WP_136490705.1		Pseudomonas sp. A-1	Proteobacteria	884	2	CMP hydrolase	0	
1662	TXI73979.1		Dokdonella sp.	Proteobacteria	885	1	CMP hydrolase	1	1113
1663	MBU1618255.1		Gammaproteobacteria bacterium	Proteobacteria	886	1	CMP hydrolase	1	1109
1664	KFB75426.1		Candidatus Accumulibacter sp. SK-02	Proteobacteria	887	1	CMP hydrolase	1	1108
1665	WP_201019106.1		Pseudomonas cichorii	Proteobacteria	888	1	CMP hydrolase	1	1111
1666	WP_134480320.1		unclassified Cryobacterium	Actinobacteria	889	1	Mrr	1	936

1667	WP_012867610.1		Sanguibacter keddieii	Actinobacteria	890	1	Mrr	1	935
1668	WP_157236008.1		Brachybacterium sp. P6-10-X1	Actinobacteria	891	2	Mrr	1	924
1669	APX33997.1		Brachybacterium sp. P6-10-X1	Actinobacteria	891	2	Mrr	0	
1670	WP_171050250.1		Nocardia cyriacigeorgica	Actinobacteria	892	1	Mrr	1	966
1671	WP_107659339.1		Nocardia suismassiliense	Actinobacteria	893	1	Mrr	1	944
1672	WP_148287086.1		Mycobacteroides abscessus	Actinobacteria	894	2	Mrr	1	933
1673	RIT93153.1		Mycobacteroides abscessus	Actinobacteria	894	2	Mrr	0	
1674	WP_074329052.1		Mycobacteroides abscessus	Actinobacteria	895	1	Mrr	1	911
1675	WP_200996284.1		Mycobacteroides chelonae	Actinobacteria	896	1	Mrr	1	928
1676	OBA50915.1		Nocardia sp. 852002-20019_SCH5090214	Actinobacteria	897	2	REase	1	991
1677	WP_156769632.1		Nocardia sp. 852002-20019_SCH5090214	Actinobacteria	897	2	REase	0	
1678	WP_062966389.1		Nocardia africana	Actinobacteria	898	5	REase	1	991
1679	WP_128145494.1		Nocardia africana	Actinobacteria	898	5	REase	0	
1680	SUA47424.1		Nocardia africana	Actinobacteria	898	5	REase	0	
1681	OBB52915.1		Nocardia sp. 852002-51244_SCH5132740	Actinobacteria	898	5	REase	0	
1682	WP_156747561.1		Nocardia	Actinobacteria	898	5	REase	0	
1683	KPI17653.1		Actinobacteria bacterium OK074	Actinobacteria	899	1	Mrr	1	1040
1684	WP_159424791.1		Streptomyces sp. TLI_053	Actinobacteria	900	2	Short	1	1013
1685	SDT03839.1		Streptomyces sp. TLI_053	Actinobacteria	900	2	Short	0	
1686	WP_053725496.1		Streptomyces sp. WM6378	Actinobacteria	901	1	Short	1	1011
1687	WP_162688589.1		unclassified Streptomyces	Actinobacteria	902	3	Short	1	1013
1688	AWZ12887.1		Streptomyces sp. ICC1	Actinobacteria	902	3	Short	0	
1689	AWZ05789.1		Streptomyces sp. ICC4	Actinobacteria	902	3	Short	0	
1690	WP_188301562.1		Streptomyces sp. CBMA156	Actinobacteria	903	1	Mrr	1	1056
1691	WP_158716977.1		Streptomyces rimosus	Actinobacteria	904	1	Mrr	1	1023
1692	ALC26399.1		Streptomyces sp. CFMR 7	Actinobacteria	905	3	Short	1	1015
1693	WP_159041063.1		Streptomyces sp. CFMR 7	Actinobacteria	905	3	Short	0	
1694	WP_125620524.1		Streptomyces sp. WAC04770	Actinobacteria	905	3	Short	0	
1695	WP_161221015.1		Streptomyces sp. SID6137	Actinobacteria	906	1	Short	1	1020

1696	ROQ62660.1		Streptomyces sp. 840.1	Actinobacteria	907	3	Short	1	1010
1697	WP_148083907.1		Streptomyces sp. 840.1	Actinobacteria	907	3	Short	0	
1698	NED89973.1		Streptomyces sp. SID11233	Actinobacteria	907	3	Short	0	
1699	WP_019553321.1		Propionispira raffinosivorans	Firmicutes	908	2	MBL + serine protease	1	1024
1700	WP_048515563.1		Megasphaera cerevisiae	Firmicutes	908	2	MBL + serine protease	0	
1701	WP_032846239.1		Fusobacterium sp. CM22	Fusobacteria	909	1	MBL + serine protease	1	1007
1702	WP_099972116.1		Fusobacterium pseudoperiodonticum	Fusobacteria	910	2	MBL + serine protease	1	1007
1703	WP_088387751.1		Fusobacterium nucleatum	Fusobacteria	910	2	MBL + serine protease	0	
1704	RKJ32227.1		Ruminococcus sp. 1xD21-23	Firmicutes	911	1	MBL + serine protease	1	1023
1705	WP_055169243.1		Roseburia inulinivorans	Firmicutes	912	2	MBL + serine protease	1	1013
1706	WP_118092614.1		Roseburia inulinivorans	Firmicutes	912	2	MBL + serine protease	0	
1707	WP_096515734.1		Clostridium perfringens	Firmicutes	913	1	MBL + serine protease	1	1013
1708	RHP49885.1		Clostridiaceae bacterium AF31-3BH	Firmicutes	914	1	MBL + serine protease	1	1010
1709	WP_173871528.1		[Clostridium] symbiosum	Firmicutes	915	2	MBL + serine protease	1	1010
1710	MBE5920952.1		Pseudobutyrivibrio ruminis	Firmicutes	915	2	MBL + serine protease	0	
1711	CDE01433.1		Roseburia sp. CAG:471	Firmicutes	916	1	MBL + serine protease	1	1010
1712	OLA48344.1		Firmicutes bacterium CAG:552_3919	Firmicutes	917	1	MBL + serine protease	1	1010
1713	NSC77111.1		[Eubacterium] rectale	Firmicutes	918	1	MBL + serine protease	1	1010
1714	WP_159765070.1		Streptomyces sp. HM190	Actinobacteria	919	1	Mrr	1	1038
1715	WP_198504301.1		Streptomyces alfalfae	Actinobacteria	920	1	Mrr	1	1049

1716	WP_190622455.1		Streptomyces venezuelae	Actinobacteria	921	2	Mrr	1	1048
1717	QES18641.1		Streptomyces venezuelae	Actinobacteria	921	2	Mrr	0	
1718	WP_143416701.1		Georgenia sp. Z446	Actinobacteria	922	1	Mrr	1	1040
1719	WP_194903723.1		Catenulispora rubra	Actinobacteria	923	1	Mrr	1	1041
1720	WP_181722936.1		Nocardia gipuzkoensis	Actinobacteria	924	1	Mrr	1	1042
1721	WP_093887864.1		Streptosporangium canum	Actinobacteria	925	1	Mrr	1	1038
1722	WP_184879921.1		Sphaerisporangium siamense	Actinobacteria	926	1	Mrr	1	1040
1723	WP_091099610.1		Nonomuraea pusilla	Actinobacteria	927	1	Mrr	1	1040
1724	WP_131757994.1		Actinomadura fibrosa	Actinobacteria	928	1	Mrr	1	1039
1725	WP_030917270.1		Streptosporangium amethystogenes	Actinobacteria	929	1	Mrr	1	1040
1726	WP_117197773.1		Nocardiopsis sp. TNDT3	Actinobacteria	930	1	Mrr	1	1039
1727	MAS53336.1		Pimelobacter sp.	Actinobacteria	931	1	Mrr	1	1043
1728	WP_128644251.1		Rhodococcus opacus	Actinobacteria	932	1	Mrr	1	1039
1729	WP_134420034.1		Cryobacterium sp. TMT2-17-1	Actinobacteria	933	1	Mrr	1	1040
1730	WP_133520983.1		Curtobacterium flaccumfaciens	Actinobacteria	934	1	Mrr	1	1040
1731	WP_166998617.1		Microbacterium sp. HY89	Actinobacteria	935	1	Mrr	1	1040
1732	NUL45367.1		Cellulosimicrobium funkei	Actinobacteria	936	1	Mrr	1	1040
1733	WP_074045371.1		Dermacoccus nishinomiyaensis	Actinobacteria	937	3	Mrr	1	1040
1734	WP_144707700.1		Dermacoccus nishinomiyaensis	Actinobacteria	937	3	Mrr	0	
1735	SDF52659.1		Cellulosimicrobium cellulans	Actinobacteria	937	3	Mrr	0	
1736	WP_079727825.1		Okibacterium fritillariae	Actinobacteria	938	1	Mrr	1	1040
1737	WP_183622568.1		Pseudoclavibacter helvolus	Actinobacteria	939	1	Mrr	1	1040
1738	WP_159878989.1		Plantibacter sp. T3	Actinobacteria	940	1	Mrr	1	1040
1739	MBI5173507.1		Candidatus Melainabacteria bacterium	Candidatus Melainabacteria	941	1	Mrr	1	1057
1740	MBI5021140.1		Ignavibacteriales bacterium	Ignavibacteriae	942	1	Mrr	1	1042
1741	WP_198128668.1		Amazonocrinis nigriterrae	Cyanobacteria	943	2	Caspase + EAD10	1	1054
1742	BBD60844.1		Nostoc sp. HK-01	Cyanobacteria	943	2	Caspase + EAD10	0	
1743	WP_014042857.1		Caldicellulosiruptor lactoaceticus	Firmicutes	944	1	Mrr	1	1049

1744	OQX52235.1		Candidatus Cloacimonas sp. 4484_209	Candidatus Cloacimonetes	945	1	Mrr	1	1053
1745	RLF27144.1		Thermoplasmata archaeon	Candidatus Thermoplasmatota	946	1	Mrr	1	1052
1746	MBS1706743.1		Armatimonadetes bacterium	Armatimonadetes	947	1	Mrr	1	1035
1747	MBK6938577.1		Planctomycetes bacterium	Planctomycetes	948	1	Mrr	1	1052
1748	MBN2285729.1		Tissierellales bacterium	Firmicutes	949	1	Mrr	1	1050
1749	NOS68312.1		Verrucomicrobia bacterium	Verrucomicrobia	950	1	Mrr	1	1051
1750	WP_006930471.1		Calditrix abyssi	Calditrichaeota	951	1	Mrr	1	1053
1751	MAT40103.1		Ectothiorhodospiraceae bacterium	Proteobacteria	952	1	Mrr	1	1051
1752	MBI3249539.1		Deltaproteobacteria bacterium	Proteobacteria	953	1	Mrr	1	1076
1753	RLB92429.1		Deltaproteobacteria bacterium	Proteobacteria	954	1	Mrr	1	1054
1754	WP_155257485.1		Rhizobium leguminosarum	Proteobacteria	955	1	Mrr	1	1052
1755	WP_184259386.1		Rhodopseudomonas rhenobacensis	Proteobacteria	956	1	Mrr	1	1055
1756	WP_174073178.1		Agrobacterium tumefaciens	Proteobacteria	957	1	Mrr	1	1030
1757	RWE81329.1		Mesorhizobium sp.	Proteobacteria	958	3	Mrr	1	1033
1758	PWB88532.1		Methylocystis sp. MitZ-2018	Proteobacteria	958	3	Mrr	0	
1759	WP_146190142.1		Methylocystaceae	Proteobacteria	958	3	Mrr	0	
1760	WP_130728760.1		Rhizobium leguminosarum	Proteobacteria	959	2	Mrr	1	1047
1761	WP_130748748.1		Rhizobium leguminosarum	Proteobacteria	959	2	Mrr	0	
1762	WP_182084430.1		Aureimonas sp. ME7	Proteobacteria	960	1	Mrr	1	1023
1763	WP_097541488.1		Rhizobium phaseoli	Proteobacteria	961	1	Mrr	1	1033
1764	WP_174278319.1		Sphingomonas bacterium	Proteobacteria	962	1	Mrr	1	1028
1765	WP_126053119.1		Sphingomonas sp. S-NIH.Pt1_0416	Proteobacteria	963	5	Mrr	1	1032
1766	RSU68431.1		Sphingomonas sp. S-NIH.Pt1_0416	Proteobacteria	963	5	Mrr	0	
1767	KIU25869.1		Sphingomonas melonis	Proteobacteria	963	5	Mrr	0	
1768	WP_131588984.1		Sphingomonas melonis	Proteobacteria	963	5	Mrr	0	
1769	RYY26349.1		Sphingomonadales bacterium	Proteobacteria	963	5	Mrr	0	
1770	WP_086896758.1		Acetobacter senegalensis	Proteobacteria	964	1	Mrr	1	1024
1771	WP_062027155.1		Gluconobacter albidus	Proteobacteria	965	1	Mrr	1	1030

1772	WP_211497349.1		Gluconobacter cerinus	Proteobacteria	966	1	Mrr	1	1028
1773	WP_048119863.1		Methanosarcina	Euryarchaeota	967	2	Mrr	1	1061
1774	WP_196297267.1		Methanosarcina barkeri	Euryarchaeota	967	2	Mrr	0	
1775	WP_129441789.1		Paenibacillus protaetiae	Firmicutes	968	1	Mrr	1	1048
1776	WP_075601003.1		Leptolyngbya sp. hensonii	Cyanobacteria	969	1	Mrr	1	1090
1777	NET28923.1		Okeania sp. SIO1I7	Cyanobacteria	970	1	Mrr	1	1137
1778	NET41796.1		Okeania sp. SIO2B3	Cyanobacteria	971	1	Mrr	1	1130
1779	MBE9046965.1		Pleurocapsales cyanobacterium LEGE 10410	Cyanobacteria	972	1	Mrr	1	1062
1780	WP_168653298.1		unclassified Dolichospermum	Cyanobacteria	973	1	Mrr	1	1066
1781	WP_190604384.1		Richelia sinica	Cyanobacteria	974	1	Mrr	1	1060
1782	MBP3366068.1		Treponema sp.	Spirochaetes	975	1	Mrr	1	1024
1783	WP_192854687.1		Vibrio lentus	Proteobacteria	976	1	Mrr	1	1027
1784	WP_153006684.1		Aeromonas schubertii	Proteobacteria	977	1	Mrr	1	1027
1785	WP_110091197.1		Photorhabdus luminescens	Proteobacteria	978	1	Mrr	1	1020
1786	WP_186848270.1		Morganella morganii	Proteobacteria	979	1	Mrr	1	1020
1787	WP_180183726.1		Acinetobacter sp. YH01020	Proteobacteria	980	1	Mrr	1	1024
1788	WP_130638992.1		Pectobacterium versatile	Proteobacteria	981	1	Mrr	1	1027
1789	WP_058797210.1		Pseudoaltero monas rubra	Proteobacteria	982	1	Mrr	1	1031
1790	WP_060615557.1		Enterobacteriaceae	Proteobacteria	983	12	Mrr	1	1027
1791	HAS6778723.1		Vibrio parahaemolyticus	Proteobacteria	983	12	Mrr	0	
1792	EGQ8259148.1		Vibrio parahaemolyticus	Proteobacteria	983	12	Mrr	0	
1793	MBE3855970.1		Vibrio parahaemolyticus	Proteobacteria	983	12	Mrr	0	
1794	WP_154115049.1		Vibrio cincinnatiensis	Proteobacteria	983	12	Mrr	0	
1795	WP_162113464.1		Escherichia coli	Proteobacteria	983	12	Mrr	0	
1796	EAS8591635.1		Salmonella enterica	Proteobacteria	983	12	Mrr	0	
1797	WP_113604026.1		Vibrio sp. 2017V-1124	Proteobacteria	983	12	Mrr	0	
1798	WP_179231273.1		Escherichia coli	Proteobacteria	983	12	Mrr	0	
1799	WP_167433323.1		Pantoea sp. EKM20T	Proteobacteria	983	12	Mrr	0	
1800	WP_162626533.1		Acinetobacter sp. CFCC 10889	Proteobacteria	983	12	Mrr	0	
1801	WP_115370610.1		Proteus vulgaris	Proteobacteria	983	12	Mrr	0	

1802	MBK83614.1		Gammaproteobacteria bacterium	Proteobacteria	984	1	Mrr	1	1028
1803	WP_169420242.1		Ramlibacter sp. G-1-2-2	Proteobacteria	985	1	Mrr	1	1027
1804	WP_157779545.1		Cupriavidus pauculus	Proteobacteria	986	1	Mrr	1	1026
1805	WP_031364364.1		Caballeronia sordidicola	Proteobacteria	987	1	Mrr	1	1024
1806	WP_062125334.1		Paraburkholderia monticola	Proteobacteria	988	1	Mrr	1	1025
1807	WP_175850769.1		Burkholderia cepacia	Proteobacteria	989	1	Mrr	1	1023
1808	WP_174947400.1		Burkholderia contaminans	Proteobacteria	990	1	Mrr	1	1023
1809	WP_197416652.1		Herbaspirillum rubrisubalbicans	Proteobacteria	991	1	Mrr	1	1025
1810	WP_013236014.1		Herbaspirillum seropedicae	Proteobacteria	992	1	Mrr	1	1032
1811	WP_001017806.1	Ec2Avs2	Escherichia coli	Proteobacteria	993	114	Mrr	1	1024
1812	WP_048023577.1		Raoultella ornithinolytica	Proteobacteria	993	114	Mrr	0	
1813	WP_071446405.1		Citrobacter portucalensis	Proteobacteria	993	114	Mrr	0	
1814	WP_061077500.1		Citrobacter amalonaticus	Proteobacteria	993	114	Mrr	0	
1815	WP_087050556.1		Citrobacter europaeus	Proteobacteria	993	114	Mrr	0	
1816	WP_153688336.1		Enterobacter cancerogenus	Proteobacteria	993	114	Mrr	0	
1817	WP_109080789.1		Enterobacter cloacae	Proteobacteria	993	114	Mrr	0	
1818	WP_045349955.1		Enterobacter cloacae complex	Proteobacteria	993	114	Mrr	0	
1819	WP_050312176.1		Yersinia intermedia	Proteobacteria	993	114	Mrr	0	
1820	WP_130574277.1		Klebsiella pneumoniae	Proteobacteria	993	114	Mrr	0	
1821	VGF49680.1		Klebsiella quasipneumoniae	Proteobacteria	993	114	Mrr	0	
1822	WP_063923755.1		Enterobacter hormaechei	Proteobacteria	993	114	Mrr	0	
1823	WP_089617536.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1824	VGD86322.1		Klebsiella quasipneumoniae	Proteobacteria	993	114	Mrr	0	
1825	SAU74098.1		Klebsiella pneumoniae	Proteobacteria	993	114	Mrr	0	
1826	WP_160470773.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1827	WP_086581273.1		Enterobacter hormaechei	Proteobacteria	993	114	Mrr	0	
1828	WP_101473388.1		Klebsiella	Proteobacteria	993	114	Mrr	0	
1829	EFT713251 3.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1830	WP_103798344.1		Citrobacter freundii complex sp. CFNIH11	Proteobacteria	993	114	Mrr	0	
1831	WP_135683178.1		Klebsiella quasipneumoniae	Proteobacteria	993	114	Mrr	0	

1832	WP_149462508.1		Pseudocitrobacter sp. 73	Proteobacteria	993	114	Mrr	0	
1833	WP_064141429.1		Klebsiella	Proteobacteria	993	114	Mrr	0	
1834	WP_115455983.1		Enterobacteriaceae	Proteobacteria	993	114	Mrr	0	
1835	WP_142590812.1		Serratia marcescens	Proteobacteria	993	114	Mrr	0	
1836	HAV1824689.1		Enterobacter hormaechei subsp. steigerwaltii	Proteobacteria	993	114	Mrr	0	
1837	WP_187228224.1		Klebsiella michiganensis	Proteobacteria	993	114	Mrr	0	
1838	WP_096960858.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1839	HAX7763913.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1840	EFG3934126.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1841	WP_186702101.1		Morganella morganii	Proteobacteria	993	114	Mrr	0	
1842	WP_023277119.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1843	WP_086869063.1		Kosakonia pseudosaccharI	Proteobacteria	993	114	Mrr	0	
1844	WP_134240797.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1845	EHL1197784.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1846	EGO8560632.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1847	WP_201476380.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1848	WP_097513253.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1849	WP_129711228.1		Pectobacterium zantedeschiae	Proteobacteria	993	114	Mrr	0	
1850	WP_126656619.1		Proteus mirabilis	Proteobacteria	993	114	Mrr	0	
1851	WP_100684190.1		Raoultella sp. T31	Proteobacteria	993	114	Mrr	0	
1852	HAV9761509.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1853	MBT9746199.1		Enterobacteriaceae bacterium MCC505	Proteobacteria	993	114	Mrr	0	
1854	WP_094322234.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1855	WP_147690613.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1856	WP_101417333.1		Klebsiella variicola	Proteobacteria	993	114	Mrr	0	
1857	WP_113977293.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1858	QMC40388.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1859	EFK6265877.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1860	WP_042201197.1		Enterobacteriaceae	Proteobacteria	993	114	Mrr	0	
1861	EFK4014607.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1862	MBJ9284380.1		Citrobacter freundii	Proteobacteria	993	114	Mrr	0	
1863	EHK0763066.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	

1864	WP_025108342.1		Enterobacteriaceae	Proteobacteria	993	114	Mrr	0	
1865	EFK1517522.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1866	WP_196344907.1		Enterobacter hormaechei	Proteobacteria	993	114	Mrr	0	
1867	WP_023063451.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1868	WP_128317410.1		Citrobacter freundii	Proteobacteria	993	114	Mrr	0	
1869	EFJ1800872.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1870	WP_096963950.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1871	HAN9276147.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1872	HAX8578568.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1873	HAG7598885.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1874	EFO7598874.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1875	HAH6667921.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1876	WP_113078630.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1877	WP_171995051.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1878	EHK4100973.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1879	WP_095512840.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1880	HBB3109134.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1881	WP_010275047.1		Pectobacterium brasiliense	Proteobacteria	993	114	Mrr	0	
1882	WP_158118081.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1883	WP_055321747.1		Enterobacter roggenkampii	Proteobacteria	993	114	Mrr	0	
1884	WP_061353966.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1885	EFM6037267.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1886	EFE8033915.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1887	WP_087907486.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1888	EFI5691907.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1889	WP_046402012.1		Citrobacter amalonaticus	Proteobacteria	993	114	Mrr	0	
1890	WP_164538422.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1891	EEQ2516251.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1892	WP_048231708.1		Citrobacter freundii complex	Proteobacteria	993	114	Mrr	0	
1893	QMC40405.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1894	HBB3077693.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1895	HBD3809431.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1896	EFB5478974.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1897	WP_032222811.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	

1898	EEW1469992.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1899	HAU5797907.1		Citrobacter amalonaticus	Proteobacteria	993	114	Mrr	0	
1900	WP_096247029.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1901	EFU9186819.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1902	WP_181508541.1		Citrobacter freundii	Proteobacteria	993	114	Mrr	0	
1903	EFH1106528.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1904	WP_191236290.1		Citrobacter freundii	Proteobacteria	993	114	Mrr	0	
1905	HBB3151880.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1906	WP_057072625.1		Enterobacter cloacae	Proteobacteria	993	114	Mrr	0	
1907	WP_049097748.1		Klebsiella	Proteobacteria	993	114	Mrr	0	
1908	HAU6430700.1		Klebsiella oxytoca	Proteobacteria	993	114	Mrr	0	
1909	EFH7880409.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1910	WP_048247357.1		Enterobacter sp. MGH120	Proteobacteria	993	114	Mrr	0	
1911	WP_063116721.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1912	HBE5127878.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1913	WP_089592439.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1914	WP_016245985.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1915	HAL5912523.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1916	WP_064670493.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1917	EGH6640931.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1918	WP_096942253.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1919	EFH5692818.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1920	WP_208310682.1		Klebsiella quasipneumoniae	Proteobacteria	993	114	Mrr	0	
1921	WP_121581526.1		Citrobacter freundii	Proteobacteria	993	114	Mrr	0	
1922	WP_161590333.1		Cronobacter dublinensis	Proteobacteria	993	114	Mrr	0	
1923	WP_086529234.1		Citrobacter freundii	Proteobacteria	993	114	Mrr	0	
1924	HBB3017363.1		Escherichia coli	Proteobacteria	993	114	Mrr	0	
1925	WP_195432750.1		Morganella morganii	Proteobacteria	994	2	Mrr	1	1018
1926	WP_186724001.1		Morganella morganii	Proteobacteria	994	2	Mrr	0	
1927	WP_148714727.1		Chitinolyticbacter meiyuanensis	Proteobacteria	995	1	Mrr	1	1031
1928	WP_179480835.1		Rhodanobacter sp. K2T2	Proteobacteria	996	1	Mrr	1	1031
1929	MBI5899123.1		Rhodocyclales bacterium	Proteobacteria	997	1	Mrr	1	1031
1930	WP_197020102.1		Desulfovibrio magneticus	Proteobacteria	998	1	Mrr	1	1054
1931	EGR0596037.1		Vibrio cholerae	Proteobacteria	999	2	Mrr	1	1046

1932	WP_110189615.1		Pokkaliibacter plantistimulans	Proteobacteria	999	2	Mrr	0	
1933	TRZ99068.1		Rhodocyclaceae bacterium	Proteobacteria	1000	1	Mrr	1	1032
1934	WP_019717795.1		Ralstonia solanacearum	Proteobacteria	1001	1	Mrr	1	1033
1935	WP_175038241.1		Burkholderia contaminans	Proteobacteria	1002	3	Mrr	1	1034
1936	WP_175223292.1		Burkholderia	Proteobacteria	1002	3	Mrr	0	
1937	WP_176100960.1		Burkholderia cepacia	Proteobacteria	1002	3	Mrr	0	
1938	ABZ00055.1		Pseudomonas putida GB-1	Proteobacteria	1003	2	Mrr	1	1030
1939	WP_041166568.1		Pseudomonas	Proteobacteria	1003	2	Mrr	0	
1940	WP_186719825.1		Pseudomonas sp. RW3S2	Proteobacteria	1004	1	Mrr	1	1030
1941	WP_122587945.1		Pseudomonas viridiflava	Proteobacteria	1005	1	Mrr	1	1030
1942	WP_008069367.1		Pseudomonas sp. GM79	Proteobacteria	1006	1	Mrr	1	1030
1943	WP_095974783.1		Pseudomonas kilonensis	Proteobacteria	1007	1	Mrr	1	1029
1944	WP_008079090.1		Pseudomonas sp. GM80	Proteobacteria	1008	1	Mrr	1	1030
1945	WP_063341207.1		Pseudomonas	Proteobacteria	1009	2	Mrr	1	1028
1946	WP_064595881.1		Pseudomonas sp. DR 5-09	Proteobacteria	1009	2	Mrr	0	
1947	WP_122527990.1		Pseudomonas viridiflava	Proteobacteria	1010	1	Mrr	1	1028
1948	WP_058445254.1		Pseudomonas syringae	Proteobacteria	1011	1	Mrr	1	1028
1949	WP_103204151.1		Pectobacterium odoriferum	Proteobacteria	1012	1	Mrr	1	1028
1950	WP_139657397.1		Pseudomonas shirazica	Proteobacteria	1013	1	Mrr	1	1028
1951	WP_210482331.1		Pseudomonas protegens	Proteobacteria	1014	1	Mrr	1	1028
1952	WP_145288581.1		Rosistilla oblonga	Planctomycetes	1015	1	Mrr	1	1149
1953	WP_146601953.1		Novipirellula aureliae	Planctomycetes	1016	1	Mrr	1	1196
1954	AMV35283.1		Pirellula sp. SH-Sr6A	Planctomycetes	1017	2	Mrr	1	1188
1955	WP_146679107.1		Pirellula sp. SH-Sr6A	Planctomycetes	1017	2	Mrr	0	
1956	WP_201572928.1		Psychrobacter immobilis	Proteobacteria	1018	1	Mrr	1	1162
1957	WP_191691173.1		Psychrobacter sp. Sa4CVA2	Proteobacteria	1019	1	Mrr	1	1162
1958	WP_212387117.1		Acinetobacter junii	Proteobacteria	1020	1	Mrr	1	1178
1959	WP_005314388.1		Acinetobacter sp. ANC 3880	Proteobacteria	1021	2	Mrr	1	1197
1960	WP_018677043.1		Acinetobacter tjernbergiae	Proteobacteria	1021	2	Mrr	0	
1961	WP_087548154.1		Acinetobacter sp. WCHA39	Proteobacteria	1022	1	Mrr	1	1195
1962	WP_007146436.1		Methylophaga aminisulfidivorans	Proteobacteria	1023	1	Mrr	1	1176
1963	SNC74478.1		Marinobacter sp. es.048	Proteobacteria	1024	2	Mrr	1	1163
1964	WP_088829540.1		Marinobacter sp. es.048	Proteobacteria	1024	2	Mrr	0	

1965	WP_193494019.1		Nitrincola alkalisedimins	Proteobacteria	1025	1	Mrr	1	1167
1966	WP_198411346.1		Marinimicrobium alkaliphilum	Proteobacteria	1026	1	Mrr	1	1167
1967	MYF97815.1		Candidatus Poribacteria bacterium	Unclassified	1027	1	Mrr	1	1164
1968	MBT2988330.1		Candidatus Thiodiazotropha sp. (ex Ctena orbiculata)	Proteobacteria	1028	1	Mrr	1	1167
1969	WP_167815437.1		Pseudoaltero monas sp. S4498	Proteobacteria	1029	1	Mrr	1	1166
1970	WP_197744396.1		Candidatus Nitrosacidococcus tergens	Proteobacteria	1030	1	Mrr	1	1168
1971	WP_077638929.1		Salinivibrio sp. ML198	Proteobacteria	1031	2	Mrr	1	1169
1972	WP_077676150.1		Salinivibrio proteolyticus	Proteobacteria	1031	2	Mrr	0	
1973	WP_100709784.1		Leptospira saintgironsiae	Spirochaetes	1032	1	Mrr	1	1166
1974	WP_016758471.1		Leptospira weilii	Spirochaetes	1033	1	Mrr	1	1166
1975	WP_061298454.1		Leptospira borgpetersenii	Spirochaetes	1034	1	Mrr	1	1166
1976	WP_004461983.1		Leptospira santarosai	Spirochaetes	1035	4	Mrr	1	1166
1977	WP_004482375.1		Leptospira santarosai	Spirochaetes	1035	4	Mrr	0	
1978	WP_004477602.1		Leptospira santarosai	Spirochaetes	1035	4	Mrr	0	
1979	WP_004485855.1		Leptospira santarosai	Spirochaetes	1035	4	Mrr	0	
1980	WP_139941215.1		Emcibacter nanhaiensis	Proteobacteria	1036	1	Mrr	1	1155
1981	MAF77944.1		Halobacteriovoraceae bacterium	Proteobacteria	1037	1	Mrr	1	1179
1982	WP_035267753.1		Desulfitibacter alkalitolerans	Firmicutes	1038	1	Mrr	1	1178
1983	WP_161841706.1		Enterococcus sp. JM4C	Firmicutes	1039	2	Mrr	1	1184
1984	KAF1298164.1		Enterococcus sp. JM4C	Firmicutes	1039	2	Mrr	0	
1985	WP_057732511.1		Levilactobacillus hammesii	Firmicutes	1040	1	Mrr	1	1139
1986	WP_049176562.1		Lacticaseibacillus paracasei	Firmicutes	1041	5	Mrr	1	1167
1987	WP_156656887.1		Lacticaseibacillus paracasei	Firmicutes	1041	5	Mrr	0	
1988	WP_096772675.1		Lacticaseibacillus paracasei	Firmicutes	1041	5	Mrr	0	
1989	WP_016387833.1		Lacticaseibacillus paracasei	Firmicutes	1041	5	Mrr	0	
1990	EPD03444.1		Lactobacillus paracasei subsp. paracasei CNCM I-2877	Firmicutes	1041	5	Mrr	0	
1991	WP_125688319.1		Levilactobacillus fujinensis	Firmicutes	1042	1	Mrr	1	1149
1992	WP_205464731.1		Clostridium botulinum	Firmicutes	1043	1	Mrr	1	1166

1993	WP_077364242.1		Clostridium saccharoperbutylacetonicum	Firmicutes	1044	1	Mrr	1	1160
1994	WP_185258788.1		Anaerocolumna sp. CTTW	Firmicutes	1045	1	Mrr	1	1184
1995	MBS5627299.1		Clostridiales bacterium	Firmicutes	1046	1	Mrr	1	1183
1996	WP_176345203.1		Priestia aryabhattai	Firmicutes	1047	1	Mrr	1	1165
1997	WP_204202019.1		Bacillus sp. RD4P76	Firmicutes	1048	1	Mrr	1	1166
1998	KXY60383.1		Bacillus cereus	Firmicutes	1049	1	Mrr	1	1193
1999	WP_103653032.1		Bacillus thuringiensis	Firmicutes	1050	1	Mrr	1	1194
2000	WP_061183612.1		Bacillus cereus	Firmicutes	1051	1	Mrr	1	1187
2001	WP_098649951.1		Bacillus wiedmannii	Firmicutes	1052	1	Mrr	1	1191
2002	WP_192598612.1		Filibacter limicola	Firmicutes	1053	1	Mrr	1	1155
2003	WP_107839881.1		Lysinibacillus meyeri	Firmicutes	1054	1	Mrr	1	1168
2004	OZI13336.1		Bacillaceae bacterium SAS-127	Firmicutes	1055	1	Mrr	1	1163
2005	WP_200750474.1		Viridibacillus sp. YIM B01967	Firmicutes	1056	1	Mrr	1	1171
2006	WP_103132291.1		Bacillus stratosphericus	Firmicutes	1057	1	Mrr	1	1167
2007	WP_214850670.1		Exiguobacterium sp. s193	Firmicutes	1058	1	Mrr	1	1162
2008	WP_144487452.1		Bacillus pumilus	Firmicutes	1059	2	Mrr	1	1162
2009	MBP1075813.1		Bacillus pumilus	Firmicutes	1059	2	Mrr	0	
2010	WP_095950902.1		Variovorax boronicumulans	Proteobacteria	1060	1	Mrr	1	1173
2011	WP_157979292.1		Rhodoferax sp. OTU1	Proteobacteria	1061	1	Mrr	1	1170
2012	WP_157754910.1		Variovorax boronicumulans	Proteobacteria	1062	2	Mrr	1	1163
2013	ATA52358.1		Variovorax boronicumulans	Proteobacteria	1062	2	Mrr	0	
2014	WP_131299226.1		Azotobacter chroococcum	Proteobacteria	1063	1	Mrr	1	1163
2015	WP_090312599.1		Pseudomonas linyingensis	Proteobacteria	1064	1	Mrr	1	1164
2016	WP_096236235.1		Pseudomonas aeruginosa	Proteobacteria	1065	15	Mrr	1	1176
2017	WP_152029940.1		Pseudomonas sp. LPH1	Proteobacteria	1065	15	Mrr	0	
2018	AQZ32756.1		Pseudomonas sp. LPH1	Proteobacteria	1065	15	Mrr	0	
2019	WP_088145119.1		Pseudomonas aeruginosa	Proteobacteria	1065	15	Mrr	0	
2020	WP_124715014.1		Pseudomonas aeruginosa	Proteobacteria	1065	15	Mrr	0	
2021	WP_077563678.1		Pseudomonas aeruginosa	Proteobacteria	1065	15	Mrr	0	
2022	WP_180983549.1		Pseudomonas stutzeri	Proteobacteria	1065	15	Mrr	0	
2023	MBG4198052.1		Pseudomonas aeruginosa	Proteobacteria	1065	15	Mrr	0	
2024	WP_140420045.1		Pseudomonas aeruginosa	Proteobacteria	1065	15	Mrr	0	

2025	WP_147316385.1		Pseudomonas aeruginosa	Proteobacteria	1065	15	Mrr	0	
2026	PNF86121.1		Pseudomonas stutzeri	Proteobacteria	1065	15	Mrr	0	
2027	OWJ52753.1		Pseudomonas aeruginosa	Proteobacteria	1065	15	Mrr	0	
2028	WP_143177002.1		Pseudomonas aeruginosa	Proteobacteria	1065	15	Mrr	0	
2029	MBG4161408.1		Pseudomonas aeruginosa	Proteobacteria	1065	15	Mrr	0	
2030	MBG6708154.1		Pseudomonas aeruginosa	Proteobacteria	1065	15	Mrr	0	
2031	WP_141252323.1		Pseudomonas aeruginosa	Proteobacteria	1066	4	Mrr	1	1175
2032	PBV51122.1		Pseudomonas aeruginosa	Proteobacteria	1066	4	Mrr	0	
2033	WP_196331926.1		Pseudomonas aeruginosa	Proteobacteria	1066	4	Mrr	0	
2034	MBI7277798.1		Pseudomonas aeruginosa	Proteobacteria	1066	4	Mrr	0	
2035	WP_055137896.1		Burkholderia plantarii	Proteobacteria	1067	2	Mrr	1	1170
2036	ALK29028.1		Burkholderia plantarii	Proteobacteria	1067	2	Mrr	0	
2037	MBS0339627.1		Proteobacteria bacterium	Proteobacteria	1068	1	Mrr	1	1166
2038	WP_212682615.1		Undibacterium sp. BYS107W	Proteobacteria	1069	1	Mrr	1	1165
2039	MBS4018226.1		Dechloromonas sp.	Proteobacteria	1070	1	Mrr	1	1165
2040	WP_130188031.1		Massilia lutea	Proteobacteria	1071	1	Mrr	1	1165
2041	WP_160947704.1		Xanthomonas	Proteobacteria	1072	1	Mrr	1	1165
2042	TXT29799.1		Rhodocyclaceae bacterium	Proteobacteria	1073	1	Mrr	1	1165
2043	MBN8442193.1		Thauera sp.	Proteobacteria	1074	1	Mrr	1	1165
2044	SDN22741.1		Janthinobacterium sp. OK676	Proteobacteria	1075	5	Mrr	1	1165
2045	WP_166450025.1		Janthinobacterium lividum	Proteobacteria	1075	5	Mrr	0	
2046	WP_133990934.1		Janthinobacterium sp. 75	Proteobacteria	1075	5	Mrr	0	
2047	TDY35344.1		Janthinobacterium sp. 75	Proteobacteria	1075	5	Mrr	0	
2048	WP_175539352.1		Janthinobacterium sp. OK676	Proteobacteria	1075	5	Mrr	0	
2049	WP_161125370.1		Malikia spinosa	Proteobacteria	1076	1	Mrr	1	1167
2050	WP_214939620.1		Rhodoferax sp. U11-2br	Proteobacteria	1077	1	Mrr	1	1166
2051	WP_157461930.1		Chitinimonas koreensis	Proteobacteria	1078	1	Mrr	1	1166
2052	OSZ41879.1		Alcaligenes faecalis	Proteobacteria	1079	2	Mrr	1	1166
2053	WP_143243308.1		Alcaligenes faecalis	Proteobacteria	1079	2	Mrr	0	
2054	WP_157197974.1		Methylomonas koyamae	Proteobacteria	1080	2	Mrr	1	1178
2055	OAI13920.1		Methylomonas koyamae	Proteobacteria	1080	2	Mrr	0	
2056	WP_190303240.1		Methylomonas rhizoryzae	Proteobacteria	1081	1	Mrr	1	1178
2057	WP_210800751.1		Ideonella sp. 4Y11	Proteobacteria	1082	1	Mrr	1	1168

2058	WP_156380867.1		Pseudolabrys sp. Root1462	Proteobacteria	1083	2	Mrr	1	1258
2059	KQZ02014.1		Pseudolabrys sp. Root1462	Proteobacteria	1083	2	Mrr	0	
2060	WP_135918576.1		unclassified Mesorhizobium	Proteobacteria	1084	1	Mrr	1	1187
2061	WP_037187045.1		Rhizobium sp. YR519	Proteobacteria	1085	1	Mrr	1	1193
2062	WP_071017235.1		Ensifer sp. LCM 4579	Proteobacteria	1086	1	Mrr	1	1194
2063	WP_150192096.1		Agrobacterium sp. ICMP 7243	Proteobacteria	1087	1	Mrr	1	1197
2064	WP_051643515.1		Mesorhizobium sp. URHC0008	Proteobacteria	1088	1	Mrr	1	1236
2065	WP_126070239.1		Mesorhizobium sp. M7D.F.Ca.US.005.01.1.1	Proteobacteria	1089	1	Mrr	1	1235
2066	WP_051377883.1		Bradyrhizobium	Proteobacteria	1090	1	Mrr	1	1246
2067	WP_007596448.1		Bradyrhizobium sp. WSM1253	Proteobacteria	1091	1	Mrr	1	1275
2068	WP_130212430.1		Bradyrhizobium genosp. SA-3	Proteobacteria	1092	1	Mrr	1	1271
2069	WP_051311001.1		Bradyrhizobium sp. Cp5.3	Proteobacteria	1093	1	Mrr	1	1271
2070	WP_147283852.1		Bosea caraganae	Proteobacteria	1094	2	Mrr	1	1181
2071	RDJ19650.1		Bosea caraganae	Proteobacteria	1094	2	Mrr	0	
2072	WP_030538494.1		Sphingobium sp. DC-2	Proteobacteria	1095	1	Mrr	1	1178
2073	WP_133066446.1		Mameliella	Proteobacteria	1096	2	Mrr	1	1178
2074	OWV54504.1		Mameliella alba	Proteobacteria	1096	2	Mrr	0	
2075	WP_176267621.1		Altererythrobacter sp. HHUK3-1	Proteobacteria	1097	1	Mrr	1	1178
2076	WP_184380431.1		Xanthomonas arboricola	Proteobacteria	1098	3	Mrr	1	1173
2077	WP_104542905.1		Xanthomonas arboricola	Proteobacteria	1098	3	Mrr	0	
2078	WP_184551861.1		Xanthomonas arboricola	Proteobacteria	1098	3	Mrr	0	
2079	WP_199284724.1		Halomonas sp. A40-4	Proteobacteria	1099	1	Mrr	1	1178
2080	WP_169140926.1		Aromatoleum toluolicum	Proteobacteria	1100	1	Mrr	1	1175
2081	WP_122704593.1		Pseudomonas viridiflava	Proteobacteria	1101	1	Mrr	1	1179
2082	WP_073454674.1		Pseudomonas	Proteobacteria	1102	13	Mrr	1	1179
2083	WP_069552391.1		Pseudomonas sp. AP19	Proteobacteria	1102	13	Mrr	0	
2084	WP_130925992.1		Pseudomonas cedrina	Proteobacteria	1102	13	Mrr	0	
2085	OEC70234.1		Pseudomonas sp. AP19	Proteobacteria	1102	13	Mrr	0	
2086	WP_047710812.1		Pseudomonas lactis	Proteobacteria	1102	13	Mrr	0	
2087	WP_141543032.1		Pseudomonas poae	Proteobacteria	1102	13	Mrr	0	
2088	ONH50879.1		Pseudomonas cedrina subsp. cedrina	Proteobacteria	1102	13	Mrr	0	

2089	WP_122279320.1		Pseudomonas syringae group genomosp. 3	Proteobacteria	1102	13	Mrr	0	
2090	WP_163019079.1		Pseudomonas viridiflava	Proteobacteria	1102	13	Mrr	0	
2091	WP_147303871.1		Pseudomonas sp. 1033	Proteobacteria	1102	13	Mrr	0	
2092	WP_095068683.1		Pseudomonas sp. Irchel 3A18	Proteobacteria	1102	13	Mrr	0	
2093	WP_124417413.1		Pseudomonas sp. LBUM920	Proteobacteria	1102	13	Mrr	0	
2094	WP_141562842.1		Pseudomonas sp. ICMP 8385	Proteobacteria	1102	13	Mrr	0	
2095	WP_146036018.1		Pseudomonas sp. LFM046	Proteobacteria	1103	1	Mrr	1	1186
2096	WP_153574109.1		Pseudomonas aeruginosa	Proteobacteria	1104	2	Mrr	1	1180
2097	WP_179573130.1		unclassified Pseudomonas	Proteobacteria	1104	2	Mrr	0	
2098	WP_114104649.1		Hyphomonas sp. CACIAM 19H1	Proteobacteria	1105	1	Mrr	1	1165
2099	WP_144254176.1		Glycocaulis profundi	Proteobacteria	1106	1	Mrr	1	1192
2100	MBN8765613.1		Thiobacillus sp.	Proteobacteria	1107	1	Mrr	1	1182
2101	MBU1324353.1		Alphaproteobacteria bacterium	Proteobacteria	1108	1	Mrr	1	1205
2102	WP_165051903.1		Methylocystis sp. MJC1	Proteobacteria	1109	2	Mrr	1	1220
2103	KAF2990930.1		Methylocystis sp. MJC1	Proteobacteria	1109	2	Mrr	0	
2104	WP_160756629.1		Pseudopontixanthobacter sediminis	Proteobacteria	1110	1	Mrr	1	1180
2105	WP_157776537.1		Nitratireductor aquibiodomus	Proteobacteria	1111	4	Mrr	1	1174
2106	VVS95920.1		Rhizobium sp. EC-SD404	Proteobacteria	1111	4	Mrr	0	
2107	WP_151983210.1		Rhizobium sp. EC-SD404	Proteobacteria	1111	4	Mrr	0	
2108	EIM73504.1		Nitratireductor aquibiodomus RA22	Proteobacteria	1111	4	Mrr	0	
2109	WP_025374068.1		Advenella mimigardefordensis	Proteobacteria	1112	1	Mrr	1	1169
2110	WP_063689135.1		Alcaligenes faecalis	Proteobacteria	1113	2	Mrr	1	1174
2111	WP_109088935.1		Alcaligenes faecalis	Proteobacteria	1113	2	Mrr	0	
2112	MBL4803248.1		Alphaproteobacteria bacterium	Proteobacteria	1114	1	Mrr	1	1198
2113	WP_018862276.1		Thioalkalivibrio sp. ALJ3	Proteobacteria	1115	1	Mrr	1	1174
2114	WP_011391500.1		Rhodospirillum rubrum	Proteobacteria	1116	1	Mrr	1	1169
2115	TVR78598.1		Rhodospirillales bacterium	Proteobacteria	1117	1	Mrr	1	1185
2116	WP_100557531.1		Brucella pituitosa	Proteobacteria	1118	1	Mrr	1	1142
2117	WP_140630232.1		unclassified Mesorhizobium	Proteobacteria	1119	1	Mrr	1	1154

2118	WP_191075918.1		Sulfitobacter sp. TSTF-M16	Proteobacteria	1120	1	Mrr	1	1159
2119	WP_215265788.1		Epibacterium ulvae	Proteobacteria	1121	1	Mrr	1	1159
2120	PTR33238.1		Mameliella alba	Proteobacteria	1122	3	Mrr	1	1162
2121	WP_188868921.1		Mameliella alba	Proteobacteria	1122	3	Mrr	0	
2122	WP_133066206.1		Mameliella alba	Proteobacteria	1122	3	Mrr	0	
2123	WP_160383884.1		Pseudooceanicola pacificus	Proteobacteria	1123	1	Mrr	1	1159
2124	PID35443.1		Rhodobacteraleses bacterium	Proteobacteria	1124	1	Mrr	1	1163
2125	WP_203155294.1		Methylobacterium aquaticum	Proteobacteria	1125	1	Mrr	1	1150
2126	SBW00780.1		uncultured Alphaproteobacteria bacterium	Proteobacteria	1126	1	Mrr	1	1156
2127	WP_103040587.1		Azospirillum brasilense	Proteobacteria	1127	1	Mrr	1	1156
2128	WP_088375230.1		Stenotrophomonas	Proteobacteria	1128	7	Mrr	1	1170
2129	WP_049455276.1		Stenotrophomonas maltophilia	Proteobacteria	1128	7	Mrr	0	
2130	WP_180840142.1		Stenotrophomonas maltophilia	Proteobacteria	1128	7	Mrr	0	
2131	QNG88790.1		Stenotrophomonas maltophilia	Proteobacteria	1128	7	Mrr	0	
2132	WP_208044118.1		Stenotrophomonas maltophilia	Proteobacteria	1128	7	Mrr	0	
2133	MBH1574377.1		Stenotrophomonas maltophilia	Proteobacteria	1128	7	Mrr	0	
2134	VUP18801.1		Stenotrophomonas maltophilia	Proteobacteria	1128	7	Mrr	0	
2135	WP_139053926.1		Pseudomonas jessenii	Proteobacteria	1129	1	Mrr	1	1156
2136	WP_194690756.1		Pseudomonas sp. p50(2008)	Proteobacteria	1130	1	Mrr	1	1151
2137	WP_122846515.1		Pseudomonas viridiflava	Proteobacteria	1131	1	Mrr	1	1153
2138	WP_013078834.1		Caulobacter segnis	Proteobacteria	1132	4	Mrr	1	1163
2139	WP_082656490.1		Caulobacter vibrioides	Proteobacteria	1132	4	Mrr	0	
2140	KSB91140.1		Caulobacter vibrioides	Proteobacteria	1132	4	Mrr	0	
2141	AVQ01923.1		Caulobacter segnis	Proteobacteria	1132	4	Mrr	0	
2142	WP_038630572.1		Pseudomonas chlororaphis	Proteobacteria	1133	1	Mrr	1	1158
2143	WP_198834233.1		Pseudomonas carnis	Proteobacteria	1134	2	Mrr	1	1157
2144	WP_124374181.1		Pseudomonas chlororaphis	Proteobacteria	1134	2	Mrr	0	
2145	QOZ84708.1		Chromobacterium sp. Rain0013	Proteobacteria	1135	2	Mrr	1	1157
2146	WP_196783784.1		Chromobacterium sp. Rain0013	Proteobacteria	1135	2	Mrr	0	

2147	WP_201302888.1		Aeromonas sp. 9A	Proteobacteria	1136	2	Mrr	1	1149
2148	VXA80166.1		Aeromonas salmonicida	Proteobacteria	1136	2	Mrr	0	
2149	WP_204787245.1		Aeromonas media	Proteobacteria	1137	6	Mrr	1	1149
2150	BBQ29933.1		Aeromonas caviae	Proteobacteria	1137	6	Mrr	0	
2151	WP_197971241.1		Aeromonas caviae	Proteobacteria	1137	6	Mrr	0	
2152	WP_204375548.1		Aeromonas allosaccharophila	Proteobacteria	1137	6	Mrr	0	
2153	KRW63061.1		Aeromonas allosaccharophila	Proteobacteria	1137	6	Mrr	0	
2154	TNI73176.1		Aeromonas media	Proteobacteria	1137	6	Mrr	0	
2155	WP_184416969.1		Rhodanobacter sp. MP7CTX1	Proteobacteria	1138	1	Mrr	1	1170
2156	WP_023707082.1		Mesorhizobium sp. LSJC255A00	Proteobacteria	1139	1	Mrr	1	1164
2157	WP_197423582.1		Novosphingobium sp. CCH12-A3	Proteobacteria	1140	1	Mrr	1	1218
2158	WP_011389337.1		Rhodospirillum rubrum	Proteobacteria	1141	1	Mrr	1	1174
2159	WP_085857160.1		Methylorubrum zatmanii	Proteobacteria	1142	1	Mrr	1	1164
2160	WP_153457869.1		Sinorhizobium fredii	Proteobacteria	1143	1	Mrr	1	1165
2161	WP_184033129.1		Sphingomonas endophytica	Proteobacteria	1144	1	Mrr	1	1165
2162	WP_070325427.1		Acetobacter ascendens	Proteobacteria	1145	3	Mrr	1	1165
2163	WP_070324331.1		Acetobacter ascendens	Proteobacteria	1145	3	Mrr	0	
2164	MBE0580024.1		Devosia sp.	Proteobacteria	1145	3	Mrr	0	
2165	WP_182573600.1		Aminobacter ciceronei	Proteobacteria	1146	1	Mrr	1	1158
2166	WP_176062752.1		Serratia marcescens	Proteobacteria	1147	3	Mrr	1	1153
2167	WP_153648023.1		Enterobacter cloacae complex	Proteobacteria	1147	3	Mrr	0	
2168	WP_132464672.1		Pantoea sp. BK028	Proteobacteria	1147	3	Mrr	0	
2169	HAY450463 1.1		Escherichia coli	Proteobacteria	1148	1	Mrr	1	1153
2170	WP_076945547.1		Raoultella terrigena	Proteobacteria	1149	1	Mrr	1	1155
2171	WP_101858464.1		Chimaeribacter arupi	Proteobacteria	1150	1	Mrr	1	1155
2172	WP_019831137.1		Sphingomonas sp. PR090111-T3T-6A	Proteobacteria	1151	1	Mrr	1	1165
2173	WP_175544633.1		Ralstonia solanacearum	Proteobacteria	1152	2	Mrr	1	1197
2174	OIT10171.1		Ralstonia solanacearum	Proteobacteria	1152	2	Mrr	0	
2175	WP_207866116.1		Burkholderia pseudomallei	Proteobacteria	1153	1	Mrr	1	1197
2176	WP_006053167.1		Paraburkholderia graminis	Proteobacteria	1154	1	Mrr	1	1175
2177	WP_012732732.1		Burkholderia glumae	Proteobacteria	1155	2	Mrr	1	1162

2178	WP_208455638.1		Burkholderia glumae	Proteobacteria	1155	2	Mrr	0	
2179	WP_186095272.1		Burkholderia gladioli	Proteobacteria	1156	1	Mrr	1	1162
2180	WP_163233921.1		Caulobacter rhizosphaerae	Proteobacteria	1157	2	Mrr	1	1169
2181	GGL35615.1		Bacillus thuringiensis	Firmicutes	1157	2	Mrr	0	
2182	WP_157076463.1		Brevundimonas vesicularis	Proteobacteria	1158	1	Mrr	1	1166
2183	WP_191554995.1		Brevundimonas aurantiaca	Proteobacteria	1159	1	Mrr	1	1178
2184	WP_201010217.1		Agrobacterium rhizogenes	Proteobacteria	1160	1	Mrr	1	1199
2185	WP_131705868.1		Rhizobium leguminosarum	Proteobacteria	1161	2	Mrr	1	1208
2186	WP_130710805.1		Rhizobium leguminosarum	Proteobacteria	1161	2	Mrr	0	
2187	WP_140767423.1		Mesorhizobium sp. B4-1-4	Proteobacteria	1162	1	Mrr	1	1208
2188	WP_156618020.1		Agrobacterium vitis	Proteobacteria	1163	2	Mrr	1	1212
2189	WP_156599079.1		Agrobacterium vitis	Proteobacteria	1163	2	Mrr	0	
2190	MBM7322394.1		Agrobacterium sp. S2	Proteobacteria	1164	1	Mrr	1	1207
2191	WP_067652620.1		Agrobacterium bohemicum	Proteobacteria	1165	1	Mrr	1	1180
2192	WP_121556610.1		Agrobacterium tumefaciens	Proteobacteria	1166	1	Mrr	1	1180
2193	WP_155924225.1		Mesorhizobium sp. L2C066B000	Proteobacteria	1167	2	Mrr	1	1175
2194	ESZ38195.1		Mesorhizobium sp. L2C066B000	Proteobacteria	1167	2	Mrr	0	
2195	WP_128962935.1		Bradyrhizobium guangzhouense	Proteobacteria	1168	2	Mrr	1	1184
2196	RXH07280.1		Bradyrhizobium guangzhouense	Proteobacteria	1168	2	Mrr	0	
2197	WP_157785160.1		Bradyrhizobium yuanmingense	Proteobacteria	1169	1	Mrr	1	1184
2198	MBF0133164.1		Magnetococcales bacterium	Proteobacteria	1170	1	Mrr	1	1062
2199	MYF07034.1		Rhodospirillaceae bacterium	Proteobacteria	1171	1	Mrr	1	1076
2200	NVN98200.1		Geobacteraceae bacterium	Proteobacteria	1172	1	Mrr	1	1181
2201	WP_215494882.1		Lysobacter capsici	Proteobacteria	1173	1	Mrr	1	1161
2202	WP_152682155.1		Stenotrophomonas maltophilia	Proteobacteria	1174	2	Mrr	1	1161
2203	KMU63651.1		Stenotrophomonas maltophilia	Proteobacteria	1174	2	Mrr	0	
2204	WP_157764506.1		Rhodobacter sp. CZR27	Proteobacteria	1175	1	Mrr	1	1082
2205	WP_162801321.1		Azospirillum brasilense	Proteobacteria	1176	1	Mrr	1	1171
2206	WP_213978898.1		Sphingomonas sp. dw_22	Proteobacteria	1177	1	Mrr	1	1159

2207	WP_128438884.1		Rhizobium leguminosarum	Proteobacteria	1178	1	Mrr	1	1162
2208	EXS70068.1		Sphingobium sp. Ant17	Proteobacteria	1179	2	Mrr	1	1162
2209	WP_152542682.1		Sphingobium sp. Ant17	Proteobacteria	1179	2	Mrr	0	
2210	KUR80448.1		Novosphingobium sp. Fuku2-ISO-50	Proteobacteria	1180	2	Mrr	1	1162
2211	WP_156427203.1		Novosphingobium sp. Fuku2-ISO-50	Proteobacteria	1180	2	Mrr	0	
2212	WP_121119977.1		Croceibacterium ferulae	Proteobacteria	1181	1	Mrr	1	1164
2213	WP_132738001.1		Sphingomonas sp. PP-F2F-A104-K0414	Proteobacteria	1182	1	Mrr	1	1164
2214	AGI72664.1		Octadecabacter arcticus 238	Proteobacteria	1183	2	Mrr	1	1164
2215	WP_044036897.1		Octadecabacter arcticus	Proteobacteria	1183	2	Mrr	0	
2216	WP_043551110.1		Acetobacter malorum	Proteobacteria	1184	2	Mrr	1	1165
2217	WP_156102524.1		Acetobacter malorum	Proteobacteria	1184	2	Mrr	0	
2218	WP_197792979.1		Serratia ureilytica	Proteobacteria	1185	3	Mrr	1	1164
2219	WP_111330731.1		Hafnia alvei	Proteobacteria	1185	3	Mrr	0	
2220	MBH2763720.1		Serratia marcescens	Proteobacteria	1185	3	Mrr	0	
2221	WP_168012076.1		Halomonas sp. G5-11	Proteobacteria	1186	1	Mrr	1	1164
2222	MBF2755898.1		Gammaproteobacteria bacterium AqS3	Proteobacteria	1187	1	Mrr	1	1093
2223	MBF2755894.1		Gammaproteobacteria bacterium AqS3	Proteobacteria	1188	1	Mrr	1	1079
2224	MXV75398.1		Candidatus Poribacteria bacterium	Unclassified	1189	1	Mrr	1	1208
2225	MYJ87184.1		Rhodobacteraceae bacterium	Proteobacteria	1190	1	Mrr	1	1204
2226	MXZ50555.1		Rhodobacteraceae bacterium	Proteobacteria	1191	1	Mrr	1	1166
2227	WP_174148990.1		Leisingera sp. ANG59	Proteobacteria	1192	1	Mrr	1	1169
2228	WP_084353999.1		Pseudooceanicola flagellatus	Proteobacteria	1193	1	Mrr	1	1184
2229	WP_055684873.1		Loktanella sp. 5RATIMAR09	Proteobacteria	1194	1	Mrr	1	1200
2230	WP_171639014.1		Ruegeria sp. HKCCD6157	Proteobacteria	1195	1	Mrr	1	1200
2231	WP_182767985.1		unclassified Pseudoaltero monas	Proteobacteria	1196	1	Mrr	1	1157
2232	WP_115121837.1		Pseudoaltero monas nigrifaciens	Proteobacteria	1197	1	Mrr	1	1157

2233	WP_036566029.1		Oceanospirillum beijerinckii	Proteobacteria	1198	1	Mrr	1	1166
2234	WP_136853895.1		Ferrimonas sp. IMCC35001	Proteobacteria	1199	1	Mrr	1	1162
2235	WP_102474308.1		Vibrio splendidus	Proteobacteria	1200	2	Mrr	1	1159
2236	WP_108118228.1		Vibrio splendidus	Proteobacteria	1200	2	Mrr	0	
2237	WP_152470888.1		Vibrio sp. THAF190c	Proteobacteria	1201	1	Mrr	1	1159
2238	WP_110415135.1		Vibrio harveyi	Proteobacteria	1202	3	Mrr	1	1161
2239	EHD2239922.1		Vibrio vulnificus	Proteobacteria	1202	3	Mrr	0	
2240	WP_114321963.1		Vibrio harveyi	Proteobacteria	1202	3	Mrr	0	
2241	WP_102458499.1		Vibrio splendidus	Proteobacteria	1203	1	Mrr	1	1159
2242	WP_155656448.1		Aliivibrio fischeri	Proteobacteria	1204	1	Mrr	1	1163
2243	WP_102887379.1		Vibrio vulnificus	Proteobacteria	1205	1	Mrr	1	1162
2244	WP_042780655.1		Vibrio parahaemolyticus	Proteobacteria	1206	11	Mrr	1	1159
2245	EHH1282515.1		Vibrio parahaemolyticus	Proteobacteria	1206	11	Mrr	0	
2246	MBM4816099.1		Vibrio parahaemolyticus	Proteobacteria	1206	11	Mrr	0	
2247	MBE4513447.1		Vibrio parahaemolyticus	Proteobacteria	1206	11	Mrr	0	
2248	HAS6611456.1		Vibrio parahaemolyticus	Proteobacteria	1206	11	Mrr	0	
2249	EGQ8494630.1		Vibrio alginolyticus	Proteobacteria	1206	11	Mrr	0	
2250	MBE4247008.1		Vibrio parahaemolyticus	Proteobacteria	1206	11	Mrr	0	
2251	EGR5517598.1		Vibrio parahaemolyticus	Proteobacteria	1206	11	Mrr	0	
2252	WP_140324516.1		Vibrio parahaemolyticus	Proteobacteria	1206	11	Mrr	0	
2253	EGR3178023.1		Vibrio parahaemolyticus	Proteobacteria	1206	11	Mrr	0	
2254	EHH2548740.1		Vibrio parahaemolyticus	Proteobacteria	1206	11	Mrr	0	
2255	WP_215407345.1		Vibrio gigantis	Proteobacteria	1207	1	Mrr	1	1159
2256	WP_179877356.1		Vibrio cholerae	Proteobacteria	1208	3	Mrr	1	1161
2257	WP_213899630.1		Vibrio alginolyticus	Proteobacteria	1208	3	Mrr	0	
2258	WP_120029469.1		Vibrio sinensis	Proteobacteria	1208	3	Mrr	0	
2259	WP_107302181.1		Photobacterium sp. GB-56	Proteobacteria	1209	1	Mrr	1	1160
2260	EGR4251981.1		Vibrio cholerae	Proteobacteria	1210	1	Mrr	1	1158
2261	MBK6896187.1		Alphaproteobacteria bacterium	Proteobacteria	1211	1	Mrr	1	1162

2262	TAA75133.1		Candidatus Electronema sp. GS	Proteobacteria	1212	1	Mrr	1	1123
2263	WP_132715871.1		Rhizobium sp. BK315	Proteobacteria	1213	2	Mrr	1	1180
2264	TCR99238.1		Rhizobium sp. BK315	Proteobacteria	1213	2	Mrr	0	
2265	WP_050578583.1		Rhizobium leguminosarum	Proteobacteria	1214	2	Mrr	1	1176
2266	WP_131703283.1		Rhizobium leguminosarum	Proteobacteria	1214	2	Mrr	0	
2267	WP_155249404.1		Rhizobium etli	Proteobacteria	1215	2	Mrr	1	1173
2268	AGS20882.1		Rhizobium etli bv. mimosae str. Mim1	Proteobacteria	1215	2	Mrr	0	
2269	WP_052227468.1		Rhizobium	Proteobacteria	1216	1	Mrr	1	1172
2270	WP_069613802.1		Rhizobium sp. YK2	Proteobacteria	1217	1	Mrr	1	1172
2271	WP_005208163.1		Acinetobacter sp. NIPH 1867	Proteobacteria	1218	1	Mrr	1	1155
2272	WP_205666721.1		Acinetobacter johnsonii	Proteobacteria	1219	1	Mrr	1	1162
2273	WP_104075494.1		Acinetobacter pittii	Proteobacteria	1220	3	Mrr	1	1157
2274	WP_199943842.1		Acinetobacter oleivorans	Proteobacteria	1220	3	Mrr	0	
2275	WP_163049140.1		Acinetobacter baumannii	Proteobacteria	1220	3	Mrr	0	
2276	WP_120429292.1		Acinetobacter baylyi	Proteobacteria	1221	1	Mrr	1	1157
2277	WP_004945568.1		Acinetobacter soli	Proteobacteria	1222	3	Mrr	1	1162
2278	WP_158312102.1		Acinetobacter calcoaceticus	Proteobacteria	1222	3	Mrr	0	
2279	WP_065993694.1		Acinetobacter defluvii	Proteobacteria	1222	3	Mrr	0	
2280	OGQ97650.1		Deltaproteobacteria bacterium RIFOXYD12_FULL_57_12	Proteobacteria	1223	1	Mrr	1	1174
2281	WP_070267962.1		Duganella sp. HH101	Proteobacteria	1224	1	Mrr	1	1177
2282	WP_126471149.1		Variovorax gossypii	Proteobacteria	1225	1	Mrr	1	1178
2283	WP_155633635.1		Burkholderia cepacia	Proteobacteria	1226	3	Mrr	1	1173
2284	WP_146075242.1		Burkholderia anthina	Proteobacteria	1226	3	Mrr	0	
2285	KVF16971.1		Burkholderia cepacia	Proteobacteria	1226	3	Mrr	0	
2286	WP_175094249.1		Methylocystis sp. FS	Proteobacteria	1227	1	Mrr	1	1177
2287	PFG71805.1		Pseudomonas poae	Proteobacteria	1228	2	Mrr	1	1057
2288	WP_141543033.1		Pseudomonas poae	Proteobacteria	1228	2	Mrr	0	
2289	WP_150705868.1		Pseudomonas fluorescens	Proteobacteria	1229	1	Mrr	1	1069
2290	WP_186511903.1		Pseudomonas panacis	Proteobacteria	1230	3	Mrr	1	1188
2291	WP_183142061.1		Pseudomonas cichorii	Proteobacteria	1230	3	Mrr	0	

2292	WP_179291823.1		Pseudomonas lundensis	Proteobacteria	1230	3	Mrr	0	
2293	MBI8948689.1		Pseudomonas aeruginosa	Proteobacteria	1231	2	Mrr	1	1185
2294	WP_179544338.1		Pseudomonas sediminis	Proteobacteria	1231	2	Mrr	0	
2295	SQK96951.1		Pseudomonas aeruginosa	Proteobacteria	1232	5	Mrr	1	1176
2296	WP_199865607.1		Pseudomonas aeruginosa	Proteobacteria	1232	5	Mrr	0	
2297	MBI7354289.1		Pseudomonas aeruginosa	Proteobacteria	1232	5	Mrr	0	
2298	WP_212612199.1		Pseudomonas oleovorans	Proteobacteria	1232	5	Mrr	0	
2299	MBU0835681.1		Alphaproteobacteria bacterium	Proteobacteria	1232	5	Mrr	0	
2300	WP_079391350.1		Pseudomonas aeruginosa	Proteobacteria	1233	3	Mrr	1	1173
2301	WP_178126199.1		Pseudomonas sp. CAH-1	Proteobacteria	1233	3	Mrr	0	
2302	QPN44358.1		Priestia aryabhattai	Firmicutes	1233	3	Mrr	0	
2303	WP_180962732.1		Pseudomonas guariconensis	Proteobacteria	1234	1	Mrr	1	1175
2304	WP_150733008.1		Pseudomonas fluorescens	Proteobacteria	1235	1	Mrr	1	1163
2305	WP_177333906.1		Pseudomonas sp. NBRC 111128	Proteobacteria	1236	1	Mrr	1	1177
2306	WP_076406136.1		Pseudomonas sp. 7SR1	Proteobacteria	1237	3	Mrr	1	1177
2307	WP_177336595.1		Pseudomonas sp. Root562	Proteobacteria	1237	3	Mrr	0	
2308	WP_205580064.1		Pseudomonas aeruginosa	Proteobacteria	1237	3	Mrr	0	
2309	WP_181819115.1		Pseudomonas fluorescens	Proteobacteria	1238	2	Mrr	1	1180
2310	MBH3803099.1		Pseudomonas aeruginosa	Proteobacteria	1238	2	Mrr	0	
2311	WP_198744243.1		Pseudomonas putida	Proteobacteria	1239	1	Mrr	1	1161
2312	WP_105753682.1		Pseudomonas	Proteobacteria	1240	4	Mrr	1	1164
2313	WP_169918563.1		unclassified Pseudomonas	Proteobacteria	1240	4	Mrr	0	
2314	WP_200655964.1		Pseudomonas sp. MF7448	Proteobacteria	1240	4	Mrr	0	
2315	WP_185817807.1		Pseudomonas kielensis	Proteobacteria	1240	4	Mrr	0	
2316	WP_154951285.1		Pseudomonas fluorescens	Proteobacteria	1241	2	Mrr	1	1164
2317	SBW84898.1		Pseudomonas veronii 1YdBTEX2	Proteobacteria	1241	2	Mrr	0	
2318	WP_078479288.1		Pseudomonas parafulva	Proteobacteria	1242	2	Mrr	1	1164
2319	WP_012053983.1		Pseudomonas putida	Proteobacteria	1242	2	Mrr	0	
2320	MBI8147480.1		Pseudomonas aeruginosa	Proteobacteria	1243	1	Mrr	1	1164
2321	WP_093945300.1		Pseudomonas aeruginosa	Proteobacteria	1244	1	Mrr	1	1164
2322	ROL67932.1		Pseudomonas protegens	Proteobacteria	1245	1	Mrr	1	1167
2323	RMP36855.1		Pseudomonas syringae pv. antirrhini	Proteobacteria	1246	2	Mrr	1	1166

2324	KPW43329.1		Pseudomonas syringae pv. antirrhini	Proteobacteria	1246	2	Mrr	0	
2325	WP_179599295.1		Pseudomonas	Proteobacteria	1247	2	Mrr	1	1166
2326	WP_183784774.1		Pseudomonas sp. JAI115	Proteobacteria	1247	2	Mrr	0	
2327	KPZ11726.1		Pseudomonas syringae pv. viburni	Proteobacteria	1248	1	Mrr	1	1188
2328	WP_057429655.1		Pseudomonas syringae group genomosp. 3	Proteobacteria	1249	2	Mrr	1	1166
2329	EGH30479.1		Pseudomonas syringae pv. japonica str. M301072	Proteobacteria	1249	2	Mrr	0	
2330	WP_059003864.1		Pseudomonas syringae	Proteobacteria	1250	4	Mrr	1	1173
2331	NAT15102.1		Pseudomonas syringae pv. actinidifoliorum	Proteobacteria	1250	4	Mrr	0	
2332	WP_088236138.1		Pseudomonas viridiflava	Proteobacteria	1250	4	Mrr	0	
2333	WP_122677723.1		Pseudomonas viridiflava	Proteobacteria	1250	4	Mrr	0	
2334	WP_057723473.1		Pseudomonas orientalis	Proteobacteria	1251	2	Mrr	1	1166
2335	WP_122524424.1		Pseudomonas viridiflava	Proteobacteria	1251	2	Mrr	0	
2336	WP_128873570.1		Pseudomonas sp. VI4.1	Proteobacteria	1252	1	Mrr	1	1166
2337	WP_116817473.1		Pseudomonas syringae	Proteobacteria	1253	1	Mrr	1	1166
2338	SEC47346.1		Pseudomonas mohnii	Proteobacteria	1254	1	Mrr	1	1166
2339	WP_127649200.1		Pseudomonas koreensis	Proteobacteria	1255	4	Mrr	1	1166
2340	WP_122682036.1		Pseudomonas viridiflava	Proteobacteria	1255	4	Mrr	0	
2341	WP_186701163.1		Pseudomonas sp. SWRI56	Proteobacteria	1255	4	Mrr	0	
2342	BBN65662.1		Pseudomonas sp. KUIN-1	Proteobacteria	1255	4	Mrr	0	
See also Data S5 of Gao et al., Science 377, eabm4096 (2022), which is incorporated herein by reference as if expressed in its entirety herein.									

Table 5. Avs3 homologs and domain architectures.

#	Accession	Name	Organism	Phylum	Tree Order	Cluster Size	Architecture	Repres-entative	Length excluding NTD
1	WP_145562674.1		Yersinia canariae	Proteobacteria	1	1	Cap4_small ORF	1	1828
2	WP_050567085.1		Dickeya zeae	Proteobacteria	2	4	Cap4_small ORF	1	1828
3	WP_052439301.1		Pectobacterium brasiliense	Proteobacteria	2	4	Cap4_small ORF	0	
4	WP_171851056.1		Dickeya dadantii	Proteobacteria	2	4	Cap4_small ORF	0	
5	WP_125261322.1		Dickeya lacustris	Proteobacteria	2	4	Cap4_small ORF	0	
6	WP_199375998.1		Brenneria sp. L3-3C-1	Proteobacteria	3	1	Cap4_small ORF	1	1828

7	WP_179895378.1		Pantoea sp. WMus005	Proteobacteria	4	1	Cap4_small ORF	1	1828
8	WP_097512812.1		Escherichia coli	Proteobacteria	5	19	Cap4_small ORF	1	1828
9	WP_181506516.1		Citrobacter freundii	Proteobacteria	5	19	Cap4_small ORF	0	
10	WP_159102106.1		Franconibacter helveticus	Proteobacteria	5	19	Cap4_small ORF	0	
11	WP_161808549.1		Franconibacter helveticus	Proteobacteria	5	19	Cap4_small ORF	0	
12	WP_203259999.1		Klebsiella grimontii	Proteobacteria	5	19	Cap4_small ORF	0	
13	WP_200033517.1		Citrobacter farmeri	Proteobacteria	5	19	Cap4_small ORF	0	
14	WP_103766796.1		Citrobacter braakii	Proteobacteria	5	19	Cap4_small ORF	0	
15	WP_064381467.1		Klebsiella michiganensis	Proteobacteria	5	19	Cap4_small ORF	0	
16	WP_103847954.1		Citrobacter braakii	Proteobacteria	5	19	Cap4_small ORF	0	
17	WP_158650671.1		Yersinia ruckeri	Proteobacteria	5	19	Cap4_small ORF	0	
18	WP_185774272.1		Klebsiella variicola	Proteobacteria	5	19	Cap4_small ORF	0	
19	WP_180240540.1		Enterobacter hormaechei	Proteobacteria	5	19	Cap4_small ORF	0	
20	WP_016243030.1		Escherichia coli	Proteobacteria	5	19	Cap4_small ORF	0	
21	EFM8967077.1		Escherichia coli	Proteobacteria	5	19	Cap4_small ORF	0	
22	WP_175539456.1		Yersinia ruckeri	Proteobacteria	5	19	Cap4_small ORF	0	
23	MBN5261490.1		Serratia marcescens	Proteobacteria	5	19	Cap4_small ORF	0	
24	AUQ42233.1		Yersinia ruckeri	Proteobacteria	5	19	Cap4_small ORF	0	
25	OJB81915.1		Yersinia ruckeri	Proteobacteria	5	19	Cap4_small ORF	0	
26	WP_106107416.1		Citrobacter freundii	Proteobacteria	5	19	Cap4_small ORF	0	
27	WP_064392579.1		Klebsiella michiganensis	Proteobacteria	6	2	Cap4_small ORF	1	1826
28	WP_192469176.1		Klebsiella michiganensis	Proteobacteria	6	2	Cap4_small ORF	0	
29	WP_046077356.1		Escherichia fergusonii	Proteobacteria	7	3	Cap4_small ORF	1	1828
30	WP_125282642.1		Escherichia fergusonii	Proteobacteria	7	3	Cap4_small ORF	0	
31	EHG6160237.1		Escherichia fergusonii	Proteobacteria	7	3	Cap4_small ORF	0	
32	WP_143840194.1		Dickeya sp. ws52	Proteobacteria	8	1	Cap4_small ORF	1	1828
33	WP_072265208.1		Serratia marcescens	Proteobacteria	9	2	Cap4_small ORF	1	1828
34	WP_161764442.1		Serratia marcescens	Proteobacteria	9	2	Cap4_small ORF	0	

35	WP_126523998.1	SeAvs3	Salmonella	Proteobacteria	10	27	Cap4_small ORF	1	1828
36	WP_048290692.1		Klebsiella pneumoniae	Proteobacteria	10	27	Cap4_small ORF	0	
37	EAP7662839.1		Salmonella enterica	Proteobacteria	10	27	Cap4_small ORF	0	
38	EBV4669867.1		Salmonella enterica subsp. enterica serovar Tamberma	Proteobacteria	10	27	Cap4_small ORF	0	
39	HAK2950351.1		Salmonella enterica	Proteobacteria	10	27	Cap4_small ORF	0	
40	WP_094964792.1		Salmonella enterica	Proteobacteria	10	27	Cap4_small ORF	0	
41	EBF8310363.1		Salmonella enterica subsp. enterica	Proteobacteria	10	27	Cap4_small ORF	0	
42	WP_116773335.1		Klebsiella pneumoniae	Proteobacteria	10	27	Cap4_small ORF	0	
43	WP_181511746.1		Klebsiella grimontii	Proteobacteria	10	27	Cap4_small ORF	0	
44	WP_130969824.1		Klebsiella pneumoniae	Proteobacteria	10	27	Cap4_small ORF	0	
45	WP_102006634.1		Klebsiella quasipneumoniae	Proteobacteria	10	27	Cap4_small ORF	0	
46	WP_171996114.1		Klebsiella pneumoniae	Proteobacteria	10	27	Cap4_small ORF	0	
47	WP_064734311.1		Klebsiella pneumoniae	Proteobacteria	10	27	Cap4_small ORF	0	
48	WP_202688916.1		Klebsiella pneumoniae	Proteobacteria	10	27	Cap4_small ORF	0	
49	WP_135695804.1		Klebsiella pneumoniae	Proteobacteria	10	27	Cap4_small ORF	0	
50	WP_159207573.1		Klebsiella pneumoniae	Proteobacteria	10	27	Cap4_small ORF	0	
51	WP_115658447.1		Klebsiella pneumoniae	Proteobacteria	10	27	Cap4_small ORF	0	
52	WP_064146026.1		Klebsiella pneumoniae	Proteobacteria	10	27	Cap4_small ORF	0	
53	WP_080870228.1		Klebsiella pneumoniae	Proteobacteria	10	27	Cap4_small ORF	0	
54	EHO5102357.1		Salmonella enterica	Proteobacteria	10	27	Cap4_small ORF	0	
55	WP_064148100.1		Klebsiella pneumoniae	Proteobacteria	10	27	Cap4_small ORF	0	
56	WP_114272321.1		Klebsiella pneumoniae	Proteobacteria	10	27	Cap4_small ORF	0	
57	ECC1694591.1		Salmonella enterica subsp. salamae	Proteobacteria	10	27	Cap4_small ORF	0	
58	WP_141335416.1		Raoultella terrigena	Proteobacteria	10	27	Cap4_small ORF	0	

59	WP_118213404.1		Klebsiella pneumoniae	Proteobacteria	10	27	Cap4_small ORF	0	
60	EAR2770941.1		Salmonella enterica	Proteobacteria	10	27	Cap4_small ORF	0	
61	ECE6741993.1		Salmonella enterica subsp. salamae	Proteobacteria	10	27	Cap4_small ORF	0	
62	WP_101737373.1		Enterobacter cancerogenus	Proteobacteria	11	1	Cap4_small ORF	1	1828
63	WP_108637153.1		Cronobacter sakazakii	Proteobacteria	12	6	Cap4_small ORF	1	1828
64	WP_146110369.1		Cronobacter sakazakii	Proteobacteria	12	6	Cap4_small ORF	0	
65	EGT4369587.1		Cronobacter sakazakii	Proteobacteria	12	6	Cap4_small ORF	0	
66	HAU5489726.1		Cronobacter sakazakii	Proteobacteria	12	6	Cap4_small ORF	0	
67	WP_199974611.1		Citrobacter sp. FDAARGOS_156	Proteobacteria	12	6	Cap4_small ORF	0	
68	PQX71787.1		Cronobacter sakazakii	Proteobacteria	12	6	Cap4_small ORF	0	
69	WP_078310439.1		Enterobacter kobei	Proteobacteria	13	2	Cap4_small ORF	1	1828
70	WP_177336629.1		Enterobacter kobei	Proteobacteria	13	2	Cap4_small ORF	0	
71	HAJ7145939.1		Escherichia coli	Proteobacteria	14	8	Cap4_small ORF	1	1828
72	EFL4225230.1		Escherichia coli	Proteobacteria	14	8	Cap4_small ORF	0	
73	HAW2445584.1		Escherichia coli	Proteobacteria	14	8	Cap4_small ORF	0	
74	EAC0144068.1		Salmonella enterica subsp. enterica	Proteobacteria	14	8	Cap4_small ORF	0	
75	EED8424924.1		Salmonella enterica subsp. enterica serovar Losangeles	Proteobacteria	14	8	Cap4_small ORF	0	
76	EBZ5774223.1		Salmonella enterica subsp. enterica serovar Redlands	Proteobacteria	14	8	Cap4_small ORF	0	
77	ECH9281679.1		Salmonella enterica subsp. enterica	Proteobacteria	14	8	Cap4_small ORF	0	
78	WP_110227906.1		Klebsiella variicola	Proteobacteria	14	8	Cap4_small ORF	0	
79	EAO3108267.1		Salmonella enterica	Proteobacteria	15	2	Cap4_small ORF	1	1828
80	EDF7620921.1		Salmonella enterica	Proteobacteria	15	2	Cap4_small ORF	0	
81	EEF7941980.1		Salmonella enterica	Proteobacteria	16	8	Cap4_small ORF	1	1828

			subsp. diarizonae						
82	WP_128348843.1		Enterobacter cloacae	Proteobacteria	16	8	Cap4_small ORF	0	
83	EEF1804563.1		Salmonella enterica	Proteobacteria	16	8	Cap4_small ORF	0	
84	EAY1316436.1		Salmonella enterica	Proteobacteria	16	8	Cap4_small ORF	0	
85	EEP9809350.1		Salmonella enterica subsp. diarizonae	Proteobacteria	16	8	Cap4_small ORF	0	
86	EDG9922790.1		Salmonella enterica	Proteobacteria	16	8	Cap4_small ORF	0	
87	ECD9467711.1		Salmonella enterica subsp. diarizonae	Proteobacteria	16	8	Cap4_small ORF	0	
88	HAU3152496.1		Salmonella enterica subsp. diarizonae	Proteobacteria	16	8	Cap4_small ORF	0	
89	WP_135192963.1		Pseudomonas putida	Proteobacteria	17	1	Cap4_small ORF	1	1839
90	WP_060767305.1		Pseudomonas	Proteobacteria	18	6	Cap4_small ORF	1	1825
91	WP_205547940.1		Pseudomonas carnis	Proteobacteria	18	6	Cap4_small ORF	0	
92	WP_054896463.1		Pseudomonas sp. NBRC 111137	Proteobacteria	18	6	Cap4_small ORF	0	
93	WP_122561302.1		Pseudomonas viridiflava	Proteobacteria	18	6	Cap4_small ORF	0	
94	WP_054998274.1		Pseudomonas syringae group	Proteobacteria	18	6	Cap4_small ORF	0	
95	WP_209996350.1		Pseudomonas sp. PvP028	Proteobacteria	18	6	Cap4_small ORF	0	
96	WP_150692522.1		Pseudomonas fluorescens	Proteobacteria	19	2	Cap4_small ORF	1	1826
97	MBH1966914.1		Pseudomonadales bacterium	Proteobacteria	19	2	Cap4_small ORF	0	
98	WP_060765484.1		Pseudomonas	Proteobacteria	20	1	Cap4_small ORF	1	1827
99	WP_186751566.1		Pseudomonas poae	Proteobacteria	21	1	Cap4_small ORF	1	1827
100	WP_044313039.1		Pseudomonas syringae	Proteobacteria	22	4	Cap4_small ORF	1	1827
101	WP_122255596.1		Pseudomonas syringae	Proteobacteria	22	4	Cap4_small ORF	0	
102	POD55958.1		Pseudomonas syringae pv. syringae	Proteobacteria	22	4	Cap4_small ORF	0	
103	WP_099314565.1		Pseudomonas syringae	Proteobacteria	22	4	Cap4_small ORF	0	
104	WP_144200047.1		Cupriavidus	Proteobacteria	23	1	Cap4_small ORF	1	1827

			campinensis						
105	WP_092136803.1		Cupriavidus sp. YR651	Proteobacteria	24	1	Cap4_small ORF	1	1829
106	WP_022701126.1		Oceanicaulis alexandrii	Proteobacteria	25	1	Cap4_small ORF	1	1827
107	PJI42197.1		Ferrovibrio sp.	Proteobacteria	26	1	Cap4_small ORF	1	1824
108	WP_026015995.1		Methylocystis parvus	Proteobacteria	27	1	Cap4_small ORF	1	1831
109	WP_197055083.1		Alcanivorax nanhaiticus	Proteobacteria	28	2	Cap4_small ORF	1	1830
110	MBM68712.1		Haliea sp.	Proteobacteria	28	2	Cap4_small ORF	0	
111	WP_116810026.1		Steroidobacter cummioxidans	Proteobacteria	29	1	Cap4_small ORF	1	1828
112	WP_176632926.1		Desulfolutivibrio sulfoxidireducens	Proteobacteria	30	1	Cap4_small ORF	1	1823
113	CCJ07338.1		Methylocystis sp. SC2	Proteobacteria	31	2	Cap4_small ORF	1	1827
114	WP_148277244.1		Methylocystis sp. SC2	Proteobacteria	31	2	Cap4_small ORF	0	
115	KZK75237.1		Pelodictyon luteolum	Chlorobi	32	1	Cap4_small ORF	1	1827
116	WP_145081765.1		Aureliella helgolandensis	Planctomycetes	33	1	Cap4_small ORF	1	1829
117	WP_167862151.1		Rhizobium	Proteobacteria	34	1	Cap4_small ORF	1	1827
118	WP_130784993.1		Rhizobium leguminosarum	Proteobacteria	35	2	Cap4_small ORF	1	1827
119	WP_130801219.1		Rhizobium leguminosarum	Proteobacteria	35	2	Cap4_small ORF	0	
120	WP_108003630.1		Mycoplana dimorpha	Proteobacteria	36	1	Cap4_small ORF	1	1827
121	WP_149743272.1		Rhizobium sp. NFACC06-2	Proteobacteria	37	1	Cap4_small ORF	1	1826
122	WP_075639957.1		Rhizobium oryziradicis	Proteobacteria	38	1	Cap4_small ORF	1	1826
123	WP_038085120.1		Thioalkalivibrio sp. ALM2T	Proteobacteria	39	1	Cap4_small ORF	1	1834
124	NKC15031.1		Gammaproteobacteria bacterium	Proteobacteria	40	1	Cap4_small ORF	1	1834
125	WP_127040394.1		Halomonas sp. PC	Proteobacteria	41	1	Cap4_small ORF	1	1833

126	WP_199475775.1		Marinomonas spartinae	Proteobacteria	42	1	Cap4_small ORF	1	1845
127	WP_158601342.1		Marinomonas rhizomae	Proteobacteria	43	2	Cap4_small ORF	1	1839
128	RNF75092.1		Marinomonas rhizomae	Proteobacteria	43	2	Cap4_small ORF	0	
129	WP_108126722.1		Saccharospirillum mangrovi	Proteobacteria	44	1	Cap4_small ORF	1	1828
130	RFC31657.1		Candidatus Nitrotoga sp. SPKER	Proteobacteria	45	1	Cap4_small ORF	1	1835
131	WP_019555396.1		Propionispira raffinosivorans	Firmicutes	46	1	Cap4_small ORF	1	1835
132	WP_196408715.1		Methylosinus sp. H3A	Proteobacteria	47	2	Cap4_small ORF	1	1837
133	WP_161914165.1		Methylosinus sp. C49	Proteobacteria	47	2	Cap4_small ORF	0	
134	WP_041954487.1		Bradyrhizobium japonicum	Proteobacteria	48	1	Cap4_small ORF	1	1831
135	WP_076827852.1		Bradyrhizobium sp. UFLA 03-321	Proteobacteria	49	1	Cap4_small ORF	1	1827
136	WP_023828174.1		Mesorhizobium sp. L103C120A0	Proteobacteria	50	1	Cap4_small ORF	1	1831
137	RWL80907.1		Mesorhizobium sp.	Proteobacteria	51	1	Cap4_small ORF	1	1832
138	RWP54695.1		Mesorhizobium sp.	Proteobacteria	52	2	Cap4_small ORF	1	1831
139	RWN69291.1		Mesorhizobium sp.	Proteobacteria	52	2	Cap4_small ORF	0	
140	WP_064984182.1		Mesorhizobium	Proteobacteria	53	1	Cap4_small ORF	1	1829
141	WP_023683415.1		Mesorhizobium sp. LSJC277A00	Proteobacteria	54	4	Cap4_small ORF	1	1830
142	WP_023823355.1		Mesorhizobium sp. L103C565B0	Proteobacteria	54	4	Cap4_small ORF	0	
143	WP_210211939.1		Mesorhizobium sp. M1B.F.Ca.ET.045.04.1.1	Proteobacteria	54	4	Cap4_small ORF	0	
144	AZO29781.1		Mesorhizobium sp. M1B.F.Ca.ET.045.04.1.1	Proteobacteria	54	4	Cap4_small ORF	0	
145	WP_127302580.1		Mesorhizobium sp. M2A.F.Ca.ET.042.01.1.1	Proteobacteria	55	1	Cap4_small ORF	1	1831
146	WP_128346628.1		Mesorhizobium sp.	Proteobacteria	56	3	Cap4_small ORF	1	1830

			M4B.F.Ca.ET.013.02.1.1						
147	WP_135913948.1		unclassified Mesorhizobium	Proteobacteria	56	3	Cap4_small ORF	0	
148	TIX00791.1		Mesorhizobium sp.	Proteobacteria	56	3	Cap4_small ORF	0	
149	MBT9444555.1		Hyphomonadaceae bacterium	Proteobacteria	57	1	Cap4_small ORF	1	1821
150	WP_205504627.1		Rufibacter sp. SYSUD00308	Bacteroidetes	58	1	Cap4_small ORF	1	1830
151	WP_040433433.1		Chlorobium ferrooxidans	Chlorobi	59	2	Cap4_small ORF	1	1832
152	EAT58839.1		Chlorobium ferrooxidans DSM 13031	Chlorobi	59	2	Cap4_small ORF	0	
153	NTU49843.1		Desulfobulbaceae bacterium	Proteobacteria	60	1	Cap4_small ORF	1	1832
154	NTW15855.1		Syntrophaceae bacterium	Proteobacteria	61	1	Cap4_small ORF	1	1830
155	WP_012465202.1		Chlorobium limicola	Chlorobi	62	1	Cap4_small ORF	1	1833
156	MYD62050.1		Gemmatimonadetes bacterium	Gemmatimonadetes	63	1	Cap4_small ORF	1	1820
157	WP_214187430.1		Geobacter hydrogenophilus	Proteobacteria	64	1	Cap4_small ORF	1	1825
158	RYG89901.1		Loktanella sp. IMCC34160	Proteobacteria	65	2	Cap4_small ORF	1	1824
159	WP_207209213.1		Loktanella sp. IMCC34160	Proteobacteria	65	2	Cap4_small ORF	0	
160	WP_118069184.1		Phaeobacter sp. LSS9	Proteobacteria	66	2	Cap4_small ORF	1	1824
161	MAY32158.1		Rhodovulum sp.	Proteobacteria	66	2	Cap4_small ORF	0	
162	WP_209194659.1		unclassified Ruegeria	Proteobacteria	67	1	Cap4_small ORF	1	1824
163	WP_153584838.1		Caenimonas koreensis	Proteobacteria	68	1	Cap4_small ORF	1	1823
164	WP_137747173.1		Acidovorax sp. NB1	Proteobacteria	69	1	Cap4_small ORF	1	1840
165	WP_156916843.1		Leisingera aquimarina	Proteobacteria	70	1	Cap4_small ORF	1	1804
166	WP_209843395.1		Rubellimicrobium aerolatum	Proteobacteria	71	1	Cap4_small ORF	1	1807
167	WP_114832840.1		Bosea caraganae	Proteobacteria	72	1	Cap4_small ORF	1	1834

168	WP_053193837.1		Chelatococcus sp. CO-6	Proteobacteria	73	2	Cap4_small ORF	1	1840
169	WP_019404581.1		Chelatococcus sp. GW1	Proteobacteria	73	2	Cap4_small ORF	0	
170	WP_183119922.1		Gluconacetobacter azotocaptans	Proteobacteria	74	1	Cap4_small ORF	1	1845
171	WP_147222915.1		Novosphingobium barchaimii	Proteobacteria	75	1	Cap4_small ORF	1	1854
172	WP_056048979.1		Sphingomonas sp. Leaf198	Proteobacteria	76	1	Cap4_small ORF	1	1845
173	WP_107125578.1		unclassified Sphingobium	Proteobacteria	77	1	Cap4_small ORF	1	1845
174	WP_157783562.1		Sphingomonas wittichii	Proteobacteria	78	1	Cap4_small ORF	1	1812
175	WP_184379455.1		Xanthomonas arboricola	Proteobacteria	79	1	Cap4_small ORF	1	1831
176	WP_162410148.1		Pseudoxanthomonas daejeonensis	Proteobacteria	80	1	Cap4_small ORF	1	1830
177	VUO36982.1		Stenotrophomonas maltophilia	Proteobacteria	81	1	Cap4_small ORF	1	1831
178	WP_122335103.1		Pseudomonas coronafaciens	Proteobacteria	82	2	Cap4_small ORF	1	1829
179	RMS04945.1		Pseudomonas coronafaciens pv. garcae	Proteobacteria	82	2	Cap4_small ORF	0	
180	WP_184643578.1		Xanthomonas arboricola	Proteobacteria	83	1	Cap4_small ORF	1	1831
181	WP_101378182.1		Xanthomonas prunicola	Proteobacteria	84	2	Cap4_small ORF	1	1826
182	PKV11009.1		Xanthomonas prunicola	Proteobacteria	84	2	Cap4_small ORF	0	
183	WP_119131591.1		Xanthomonas	Proteobacteria	85	3	Cap4_small ORF	1	1831
184	WP_126943006.1		Xanthomonas sp. BRIP62409	Proteobacteria	85	3	Cap4_small ORF	0	
185	WP_128418127.1		Xanthomonas populi	Proteobacteria	85	3	Cap4_small ORF	0	
186	WP_090898878.1		Azotobacter beijerinckii	Proteobacteria	86	1	Cap4_small ORF	1	1832
187	WP_122350126.1		Pseudomonas coronafaciens	Proteobacteria	87	2	Cap4_small ORF	1	1832

188	WP_159389519.1		Pseudomonas sp. S34	Proteobacteria	87	2	Cap4_small ORF	0	
189	WP_129398940.1		Pseudomonas syringae	Proteobacteria	88	1	Cap4_small ORF	1	1831
190	WP_193452839.1		Pseudomonas nitritireducens	Proteobacteria	89	1	Cap4_small ORF	1	1830
191	WP_003154773.1		Pseudomonas aeruginosa	Proteobacteria	90	7	Cap4_small ORF	1	1830
192	WP_121431117.1		Pseudomonas aeruginosa	Proteobacteria	90	7	Cap4_small ORF	0	
193	MBH8769402.1		Pseudomonas aeruginosa	Proteobacteria	90	7	Cap4_small ORF	0	
194	MBG6503339.1		Pseudomonas aeruginosa	Proteobacteria	90	7	Cap4_small ORF	0	
195	MBG5227047.1		Pseudomonas aeruginosa	Proteobacteria	90	7	Cap4_small ORF	0	
196	WP_034063567.1		Pseudomonas aeruginosa	Proteobacteria	90	7	Cap4_small ORF	0	
197	MBG4707713.1		Pseudomonas aeruginosa	Proteobacteria	90	7	Cap4_small ORF	0	
198	WP_101212722.1		Psychrobacter sp. Sarcosine-3u-12	Proteobacteria	91	2	Cap4_small ORF	1	1816
199	WP_102078010.1		unclassified Psychrobacter	Proteobacteria	91	2	Cap4_small ORF	0	
200	WP_201544112.1		Psychrobacter sp. H7-1	Proteobacteria	92	1	Cap4_small ORF	1	1816
201	WP_086376279.1		Acinetobacter pittii	Proteobacteria	93	2	Cap4_small ORF	1	1817
202	WP_199937077.1		Acinetobacter pittii	Proteobacteria	93	2	Cap4_small ORF	0	
203	WP_065726022.1		Acinetobacter baumannii	Proteobacteria	94	5	Cap4_small ORF	1	1817
204	WP_098729666.1		Acinetobacter baumannii	Proteobacteria	94	5	Cap4_small ORF	0	
205	HAV5509017.1		Acinetobacter baumannii	Proteobacteria	94	5	Cap4_small ORF	0	
206	WP_207272616.1		Acinetobacter baumannii	Proteobacteria	94	5	Cap4_small ORF	0	
207	WP_069140493.1		Acinetobacter pittii	Proteobacteria	94	5	Cap4_small ORF	0	
208	WP_002064786.1		Acinetobacter baumannii	Proteobacteria	95	5	Cap4_small ORF	1	1816

209	WP_199112574.1		unclassified Acinetobacter	Proteobacteria	95	5	Cap4_small ORF	0	
210	WP_194301982.1		Acinetobacter baumannii	Proteobacteria	95	5	Cap4_small ORF	0	
211	WP_057045576.1		Acinetobacter baumannii	Proteobacteria	95	5	Cap4_small ORF	0	
212	WP_130120739.1		Acinetobacter pittii	Proteobacteria	95	5	Cap4_small ORF	0	
213	WP_007481002.1		Acinetobacter sp. NBRC 100985	Proteobacteria	96	1	Cap4_small ORF	1	1817
214	WP_151780342.1		Acinetobacter bereziniae	Proteobacteria	97	1	Cap4_small ORF	1	1818
215	WP_166782102.1		Acinetobacter sp. BIGb0196	Proteobacteria	98	1	Cap4_small ORF	1	1816
216	WP_180022926.1		Acinetobacter sp. YH16044	Proteobacteria	99	1	Cap4_small ORF	1	1816
217	WP_086185859.1		Acinetobacter sp. ANC 4973	Proteobacteria	100	1	Cap4_small ORF	1	1817
218	WP_180159071.1		Acinetobacter sp. YH01026	Proteobacteria	101	1	Cap4_small ORF	1	1817
219	WP_075175523.1		Acinetobacter indicus	Proteobacteria	102	2	Cap4_small ORF	1	1817
220	WP_067731792.1		Acinetobacter sp. NCu2D-2	Proteobacteria	102	2	Cap4_small ORF	0	
221	WP_104795294.1		Acinetobacter ursingii	Proteobacteria	103	1	Cap4_small ORF	1	1817
222	WP_194558073.1		Acinetobacter towneri	Proteobacteria	104	1	Cap4_small ORF	1	1817
223	WP_157267973.1		Azohydromonas aeria	Proteobacteria	105	1	Cap4_small ORF	1	1827
224	WP_105749592.1		Malikia granosa	Proteobacteria	106	1	Cap4_small ORF	1	1826
225	NNM63481.1		Burkholderiales bacterium	Proteobacteria	107	1	Cap4_small ORF	1	1825
226	WP_110118898.1		unclassified Caulobacter	Proteobacteria	108	1	Cap4_small ORF	1	1836
227	WP_183961794.1		Chiayiivirga flava	Proteobacteria	109	1	Cap4_small ORF	1	1812
228	WP_074529308.1		Pseudomonas psychrotolerans	Proteobacteria	110	2	Cap4_small ORF	1	1828
229	WP_169839430.1		Pseudomonas lundensis	Proteobacteria	110	2	Cap4_small ORF	0	
230	WP_065834073.1		Pseudomonas syringae	Proteobacteria	111	1	Cap4_small ORF	1	1820
231	WP_110639959.1		unclassified	Proteobacteria	112	1	Cap4_small ORF	1	1820

			Pseudomonas						
232	WP_005735238.1		Pseudomonas syringae group	Proteobacteria	113	1	Cap4_small ORF	1	1820
233	WP_024681286.1		Pseudomonas syringae	Proteobacteria	114	4	Cap4_small ORF	1	1820
234	WP_060691152.1		Pseudomonas versuta	Proteobacteria	114	4	Cap4_small ORF	0	
235	WP_201491400.1		Pseudomonas paracarnis	Proteobacteria	114	4	Cap4_small ORF	0	
236	WP_198705661.1		Pseudomonas syringae	Proteobacteria	114	4	Cap4_small ORF	0	
237	WP_122685665.1		Pseudomonas viridiflava	Proteobacteria	115	1	Cap4_small ORF	1	1820
238	ALU44170.1		Pseudoalteromonas rubra	Proteobacteria	116	2	Cap4_small ORF	1	1817
239	WP_208856741.1		Pseudoalteromonas rubra	Proteobacteria	116	2	Cap4_small ORF	0	
240	WP_156020613.1		Shewanella algae	Proteobacteria	117	1	Cap4_small ORF	1	1818
241	VUQ29546.1		Stenotrophomonas maltophilia	Proteobacteria	118	1	Cap4_small ORF	1	1831
242	RKH95705.1		Coralloccus sp. AB038B	Proteobacteria	119	2	Cap4_small ORF	1	1823
243	WP_199749767.1		Coralloccus sp. AB038B	Proteobacteria	119	2	Cap4_small ORF	0	
244	WP_205525541.1		Pyxidicoccus trucidator	Proteobacteria	120	1	Cap4_small ORF	1	1828
245	WP_200384856.1		Rhodocyclus tenuis	Proteobacteria	121	1	Cap4_small ORF	1	1819
246	WP_121206485.1		Marinobacter hydrocarbonoclasticus	Proteobacteria	122	1	Cap4_small ORF	1	1824
247	WP_041158298.1		Halomonas sp. KHS3	Proteobacteria	123	2	Cap4_small ORF	1	1823
248	WP_027958588.1		unclassified Halomonas	Proteobacteria	123	2	Cap4_small ORF	0	
249	WP_052096709.1		Achromobacter sp. RTa	Proteobacteria	124	5	Cap4_small ORF	1	1845
250	WP_169269814.1		Achromobacter sp. Bel	Proteobacteria	124	5	Cap4_small ORF	0	
251	KGE01755.1		Achromobacter sp. RTa	Proteobacteria	124	5	Cap4_small ORF	0	
252	ATE61953.1		Thauera sp. K11	Proteobacteria	124	5	Cap4_small ORF	0	
253	WP_198363205.1		Thauera sp. K11	Proteobacteria	124	5	Cap4_small ORF	0	
254	WP_018151191.1		Leeia oryzae	Proteobacteria	125	1	Cap4_small ORF	1	1846

255	WP_161018932.1		Duganella sp. FT50W	Proteobacteria	126	1	Cap4_small ORF	1	1845
256	WP_176386491.1		Janthinobacterium lividum	Proteobacteria	127	1	Cap4_small ORF	1	1845
257	WP_198797636.1		Pseudomonas	Proteobacteria	128	1	Cap4_small ORF	1	1815
258	WP_026036035.1		Cupriavidus sp. BIS7	Proteobacteria	129	1	Cap4_small ORF	1	1825
259	AMV48212.1		Paraburkholderia caribensis	Proteobacteria	130	2	Cap4_small ORF	1	1816
260	WP_082845565.1		Paraburkholderia caribensis	Proteobacteria	130	2	Cap4_small ORF	0	
261	WP_147236899.1		Paraburkholderia azotifigens	Proteobacteria	131	1	Cap4_small ORF	1	1822
262	WP_134465996.1		Paraburkholderia dipogonis	Proteobacteria	132	1	Cap4_small ORF	1	1822
263	WP_109483812.1		Paraburkholderia sp. C35	Proteobacteria	133	1	Cap4_small ORF	1	1822
264	WP_162015164.1		Acidovorax sp. 210-6	Proteobacteria	134	1	Cap4_small ORF	1	1823
265	WP_003054383.1		Comamonas testosteroni	Proteobacteria	135	2	Cap4_small ORF	1	1823
266	WP_182283021.1		Comamonas testosteroni	Proteobacteria	135	2	Cap4_small ORF	0	
267	MBG6833261.1		Pseudomonas aeruginosa	Proteobacteria	136	1	Cap4_small ORF	1	1823
268	WP_102073478.1		Pusillimonas sp. JR1/69-3-13	Proteobacteria	137	1	Cap4_small ORF	1	1822
269	WP_136179726.1		Hydrogenophaga sp. PAMC20947	Proteobacteria	138	1	Cap4_small ORF	1	1822
270	MBL8436298.1		Zoogloea sp.	Proteobacteria	139	1	Cap4_small ORF	1	1817
271	MBH2018478.1		Burkholderiales bacterium	Proteobacteria	140	1	Cap4_small ORF	1	1823
272	HIM97130.1		Gammaproteobacteria bacterium	Proteobacteria	141	1	Cap4_small ORF	1	1820
273	WP_042126867.1		Pseudomonas japonica	Proteobacteria	142	1	Cap4_small ORF	1	1820
274	MBG5462220.1		Pseudomonas aeruginosa	Proteobacteria	143	1	Cap4_small ORF	1	1823
275	WP_124082542.1		Pseudomonas aeruginosa	Proteobacteria	144	1	Cap4_small ORF	1	1823

276	WP_031687412.1		Pseudomonas aeruginosa	Proteobacteria	145	1	Cap4_small ORF	1	1823
277	WP_161864276.1		Pseudomonas	Proteobacteria	146	1	Cap4_small ORF	1	1822
278	WP_183894544.1		Pseudomonas mendocina	Proteobacteria	147	2	Cap4_small ORF	1	1823
279	WP_204915775.1		Pseudomonas sp. UL073	Proteobacteria	147	2	Cap4_small ORF	0	
280	WP_083728566.1		Pseudomonas pachastrellae	Proteobacteria	148	1	Cap4_small ORF	1	1822
281	WP_110674237.1		Salinicola sp. RZ23	Proteobacteria	149	1	Cap4_small ORF	1	1823
282	WP_026083746.1		Pseudomonas sp. 313	Proteobacteria	150	1	Cap4_small ORF	1	1822
283	WP_110747839.1		Pseudomonas sp. LB-090624	Proteobacteria	151	1	Cap4_small ORF	1	1821
284	WP_198747357.1		Pseudomonas putida	Proteobacteria	152	2	Cap4_small ORF	1	1823
285	WP_115273831.1		Pseudomonas putida	Proteobacteria	152	2	Cap4_small ORF	0	
286	WP_096052415.1		Caulobacter vibrioides	Proteobacteria	153	1	Cap4_small ORF	1	1817
287	WP_146069194.1		Arthrobacter sp. ZGTC412	Actinobacteria	154	1	Cap4_small ORF	1	1799
288	WP_159852480.1		Rathayibacter sp. VKM Ac-2754	Actinobacteria	155	1	Cap4_small ORF	1	1795
289	WP_209514676.1		Frigoribacterium sp. PvP121	Actinobacteria	156	1	Cap4_small ORF	1	1797
290	WP_116647513.1		Microbacterium sp. Gd 4-13	Actinobacteria	157	1	Cap4_small ORF	1	1784
291	OZD38633.1		Rhodococcus sp. 06-1477-1B	Actinobacteria	158	1	Cap4_small ORF	1	1803
292	WP_058595048.1		Microbacterium testaceum	Actinobacteria	159	2	Cap4_small ORF	1	1803
293	KTS07559.1		Microbacterium testaceum	Actinobacteria	159	2	Cap4_small ORF	0	
294	WP_144018563.1		Demequina sp. NBRC 110051	Actinobacteria	160	1	Cap4_small ORF	1	1800
295	MBA3020037.1		Propionicimonas sp.	Actinobacteria	161	1	Cap4_small ORF	1	1807
296	WP_040358575.1		Corynebacterium durum	Actinobacteria	162	2	Cap4_small ORF	1	1790
297	EKX92431.1		Corynebacterium durum F0235	Actinobacteria	162	2	Cap4_small ORF	0	

298	WP_102034948.1		Rothia dentocariosa	Actinobacteria	163	1	Cap4_small ORF	1	1783
299	WP_198412067.1		Cryobacterium sp. Hz9	Actinobacteria	164	2	Cap4_small ORF	1	1803
300	TFB68068.1		Cryobacterium sp. Hz9	Actinobacteria	164	2	Cap4_small ORF	0	
301	WP_200953175.1		Arthrobacter sp. Soil761	Actinobacteria	165	2	Cap4_small ORF	1	1803
302	KRE68306.1		Arthrobacter sp. Soil761	Actinobacteria	165	2	Cap4_small ORF	0	
303	WP_081630541.1		Rhodococcus sp. 29MFTsu3.1	Actinobacteria	166	1	Cap4_small ORF	1	1803
304	RLJ73322.1		Kocuria sp. OV113	Actinobacteria	167	4	Cap4_small ORF	1	1798
305	WP_141564929.1		Kocuria	Actinobacteria	167	4	Cap4_small ORF	0	
306	WP_145000462.1		Kocuria rhizophila	Actinobacteria	167	4	Cap4_small ORF	0	
307	WP_194607797.1		unclassified Rathayibacter	Actinobacteria	167	4	Cap4_small ORF	0	
308	WP_196482691.1		Propionibacterium freudenreichii	Actinobacteria	168	5	Cap4_small ORF	1	1787
309	WP_196488275.1		Propionibacterium freudenreichii	Actinobacteria	168	5	Cap4_small ORF	0	
310	CEH06303.1		Propionibacterium freudenreichii	Actinobacteria	168	5	Cap4_small ORF	0	
311	CEG86733.1		Propionibacterium freudenreichii	Actinobacteria	168	5	Cap4_small ORF	0	
312	WP_055345341.1		Propionibacterium freudenreichii	Actinobacteria	168	5	Cap4_small ORF	0	
313	WP_205881305.1		Leucobacter viscericola	Actinobacteria	169	2	Cap4_small ORF	1	1802
314	QIK64893.1		Leucobacter viscericola	Actinobacteria	169	2	Cap4_small ORF	0	
315	WP_074495317.1		Corynebacterium glutamicum	Actinobacteria	170	3	Cap4_small ORF	1	1797
316	AKF26116.1		[Brevibacterium] flavum	Actinobacteria	170	3	Cap4_small ORF	0	
317	WP_155268202.1		Corynebacterium	Actinobacteria	170	3	Cap4_small ORF	0	
318	PZP38120.1		Kocuria rhizophila	Actinobacteria	171	1	Cap4_small ORF	1	1798

319	WP_200936125.1		Microbacterium sp. Leaf151	Actinobacteria	172	1	Cap4_small ORF	1	1798
320	WP_083403330.1		Curtobacterium sp. MCBA15_016	Actinobacteria	173	1	Cap4_small ORF	1	1798
321	WP_145010976.1		Kocuria salsicia	Actinobacteria	174	1	Cap4_small ORF	1	1797
322	TQO24165.1		Microbacterium lindanitolerans	Actinobacteria	175	2	Cap4_small ORF	1	1798
323	WP_211346093.1		Microbacterium lindanitolerans	Actinobacteria	175	2	Cap4_small ORF	0	
324	WP_206541989.1		Leucobacter celer	Actinobacteria	176	1	Cap4_small ORF	1	1798
325	WP_206398880.1		Kocuria indica	Actinobacteria	177	1	Cap4_small ORF	1	1803
326	WP_057478157.1		Rhodococcus sp. Leaf278	Actinobacteria	178	1	Cap4_small ORF	1	1821
327	RZL21839.1		Rhodococcus sp.	Actinobacteria	179	1	Cap4_small ORF	1	1826
328	WP_173054954.1		Phytohabitans houttuyneae	Actinobacteria	180	1	Cap4_small ORF	1	1827
329	WP_143234482.1		Actinoplanes atraurantiacus	Actinobacteria	181	1	Cap4_small ORF	1	1822
330	WP_134490247.1		Cryobacterium sp. Sr47	Actinobacteria	182	1	Cap4_small ORF	1	1802
331	PVZ58357.1		Arthrobacter sp. H-02-3	Actinobacteria	183	2	Cap4_small ORF	1	1824
332	WP_133247487.1		Arthrobacter sp. H-02-3	Actinobacteria	183	2	Cap4_small ORF	0	
333	WP_211880084.1		Pseudarthrobacter sp. NJ-Z5	Actinobacteria	184	1	Cap4_small ORF	1	1823
334	WP_185276281.1		Leifsonia shinshuensis	Actinobacteria	185	1	Cap4_small ORF	1	1787
335	WP_087196144.1		Clavibacter michiganensis	Actinobacteria	186	1	Cap4_small ORF	1	1791
336	WP_134368297.1		Cryobacterium sp. TMT2-10	Actinobacteria	187	1	Cap4_small ORF	1	1798
337	WP_157508769.1		Luteipulveratus halotolerans	Actinobacteria	188	1	Cap4_small ORF	1	1774
338	WP_110485701.1		Arthrobacter psychrolactophilus	Actinobacteria	189	1	Cap4_small ORF	1	1777
339	WP_184794645.1		Kribbella italica	Actinobacteria	190	1	Cap4_small ORF	1	1780
340	TXH25686.1		Mycobacterium sp.	Actinobacteria	191	1	Cap4_small ORF	1	1782
341	WP_189254728.1		Lentzea flava	Actinobacteria	192	1	Cap4_small ORF	1	1780

342	WP_182883913.1		Microbispora sp. H10949	Actinobacteria	193	1	Cap4_small ORF	1	1788
343	WP_005245526.1		Rhodococcus opacus	Actinobacteria	194	2	Cap4_small ORF	1	1785
344	AHK36007.1		Rhodococcus opacus PD630	Actinobacteria	194	2	Cap4_small ORF	0	
345	WP_154606729.1		Rhodococcus sp. AQ5-07	Actinobacteria	195	2	Cap4_small ORF	1	1788
346	RAL31640.1		Rhodococcus sp. AQ5-07	Actinobacteria	195	2	Cap4_small ORF	0	
347	WP_192713647.1		Rhodococcus sp. OAS809	Actinobacteria	196	1	Cap4_small ORF	1	1832
348	WP_181785492.1		Streptomyces sp. PIP175	Actinobacteria	197	1	Cap4_small ORF	1	1842
349	WP_127891584.1		Streptomyces sp. S10(2018)	Actinobacteria	198	1	Cap4_small ORF	1	1834
350	WP_018491709.1		unclassified Streptomyces	Actinobacteria	199	1	Cap4_small ORF	1	1834
351	WP_098755023.1		Streptomyces sp. Ru87	Actinobacteria	200	1	Cap4_small ORF	1	1835
352	WP_189134488.1		Streptomyces tyrosinilyticus	Actinobacteria	201	6	Cap4_small ORF	1	1835
353	WP_093896577.1		Streptomyces sp. Ncost-T10-10d	Actinobacteria	201	6	Cap4_small ORF	0	
354	GGO96081.1		Streptomyces tyrosinilyticus	Actinobacteria	201	6	Cap4_small ORF	0	
355	WP_030835043.1		Streptomyces sp. NRRL S-475	Actinobacteria	201	6	Cap4_small ORF	0	
356	WP_073749969.1		Streptomyces sp. CB02058	Actinobacteria	201	6	Cap4_small ORF	0	
357	WP_037889127.1		Streptomyces sp. Tu 6176	Actinobacteria	201	6	Cap4_small ORF	0	
358	WP_175255073.1		Streptomyces lunaelactis	Actinobacteria	202	1	Cap4_small ORF	1	1835
359	WP_118898966.1		Streptomyces globisporus	Actinobacteria	203	5	Cap4_small ORF	1	1836
360	GGX07999.1		Streptomyces chryseus	Actinobacteria	203	5	Cap4_small ORF	0	
361	WP_138907574.1		Streptomyces chryseus	Actinobacteria	203	5	Cap4_small ORF	0	
362	NUV51789.1		Streptomyces coelicolor	Actinobacteria	203	5	Cap4_small ORF	0	

363	NED75608.1		Streptomyces sp. SID9944	Actinobacteria	203	5	Cap4_small ORF	0	
364	WP_157430493.1		Actinomadura macra	Actinobacteria	204	1	Cap4_small ORF	1	1833
365	WP_176574478.1		Streptomyces chartreusis	Actinobacteria	205	1	Cap4_small ORF	1	1833
366	WP_143645187.1		Streptomyces antioxidans	Actinobacteria	206	2	Cap4_small ORF	1	1833
367	OPF82340.1		Streptomyces antioxidans	Actinobacteria	206	2	Cap4_small ORF	0	
368	WP_085288617.1		Mycobacterium vulneris	Actinobacteria	207	1	Cap4_small ORF	1	1861
369	WP_111509610.1		Mycobacterium kyogaense	Actinobacteria	208	3	Cap4_small ORF	1	1855
370	WP_065071141.1		Mycolicibacterium fortuitum	Actinobacteria	208	3	Cap4_small ORF	0	
371	WP_210393762.1		Mycolicibacterium fortuitum	Actinobacteria	208	3	Cap4_small ORF	0	
372	WP_005099516.1		Mycobacteroides abscessus	Actinobacteria	209	2	Cap4_small ORF	1	1858
373	WP_079670236.1		Mycobacteroides abscessus	Actinobacteria	209	2	Cap4_small ORF	0	
374	WP_085264060.1		Mycolicibacter longobardus	Actinobacteria	210	1	Cap4_small ORF	1	1852
375	GIJ01210.1		Spirilliplanes yamanashiensis	Actinobacteria	211	2	Cap4_small ORF	1	1874
376	WP_203936542.1		Spirilliplanes yamanashiensis	Actinobacteria	211	2	Cap4_small ORF	0	
377	TWA52200.1		Micromonospora sp. CNZ322	Actinobacteria	212	2	Cap4_small ORF	1	1873
378	WP_145794418.1		Micromonospora sp. CNZ322	Actinobacteria	212	2	Cap4_small ORF	0	
379	WP_210502800.1		Nocardioides sp. SYSU D00778	Actinobacteria	213	1	Cap4_small ORF	1	1845
380	WP_156365978.1		Nocardiopsis sp. NRRL B-16309	Actinobacteria	214	1	Cap4_small ORF	1	1880
381	WP_131987623.1		Jiangella ureilytica	Actinobacteria	215	1	Cap4_small ORF	1	1884
382	WP_147915720.1		unclassified Ruania	Actinobacteria	216	1	Cap4_small ORF	1	1866
383	WP_157720217.1		Friedmanniella luteola	Actinobacteria	217	1	Cap4_small ORF	1	1882

384	WP_173156932.1		Brevibacterium permense	Actinobacteria	218	1	Cap4_small ORF	1	1865
385	WP_204839548.1		Nocardioides nitrophenolicus	Actinobacteria	219	1	Cap4_small ORF	1	1872
386	WP_143445429.1		Janibacter indicus	Actinobacteria	220	2	Cap4_small ORF	1	1866
387	SMC44981.1		Janibacter indicus	Actinobacteria	220	2	Cap4_small ORF	0	
388	WP_210220577.1		Thiothrix unzii	Proteobacteria	221	1	Cap4_small ORF	1	1846
389	WP_014615442.1		Shewanella baltica	Proteobacteria	222	1	Transmembrane (5' split)_small ORF	1	1823
390	WP_210456023.1		Vibrio crassostreae	Proteobacteria	223	1	Transmembrane (5' split)_small ORF	1	1825
391	WP_006960618.1		Vibrio coralliilyticus	Proteobacteria	224	1	Transmembrane (5' split)_small ORF	1	1820
392	WP_017048849.1		Vibrio anguillarum	Proteobacteria	225	2	Transmembrane (5' split)_small ORF	1	1822
393	WP_170906941.1		Vibrio sp. B4-6	Proteobacteria	225	2	Transmembrane (5' split)_small ORF	0	
394	WP_102576397.1		Vibrio splendidus	Proteobacteria	226	1	Transmembrane (5' split)_small ORF	1	1818
395	WP_028834416.1		unclassified Pseudoalteromonas	Proteobacteria	227	1	Transmembrane (5' split)_small ORF	1	1818
396	WP_129049322.1		Oleihareniccola lentus	Verrucomicrobia	228	1	Cap4_small ORF	1	1840
397	WP_165204585.1		Roseimicrobium sp. ORNL1	Verrucomicrobia	229	1	Cap4_small ORF	1	1845
398	WP_146849627.1		Brevifollis gellanilyticus	Verrucomicrobia	230	2	Cap4_small ORF	1	1852
399	GEP42020.1		Brevifollis gellanilyticus	Verrucomicrobia	230	2	Cap4_small ORF	0	
400	EHH1179034.1		Vibrio vulnificus	Proteobacteria	231	2	Cap4_small ORF	1	1850
401	EHK9117281.1		Vibrio vulnificus	Proteobacteria	231	2	Cap4_small ORF	0	
402	WP_025524892.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	1	1849
403	HAS6493532.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
404	EGR5665740.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
405	WP_059678416.1		Vibrio harveyi	Proteobacteria	232	29	Cap4_small ORF	0	

406	EGR4447106.1		Vibrio cholerae	Proteobacteria	232	29	Cap4_small ORF	0	
407	WP_199352693.1		Vibrio cholerae	Proteobacteria	232	29	Cap4_small ORF	0	
408	MBM4964991.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
409	WP_142577637.1		Vibrio cholerae	Proteobacteria	232	29	Cap4_small ORF	0	
410	HAS6900126.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
411	MBM4919251.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
412	EGR4770911.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
413	WP_074545412.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
414	EHD0129869.1		Vibrio alginolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
415	WP_089204840.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
416	WP_069488063.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
417	HAS8571182.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
418	EHH1033375.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
419	WP_206975022.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
420	WP_154218601.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
421	HBC3374621.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
422	EGR1973961.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
423	EGR4822567.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
424	EGQ9880045.1		Vibrio vulnificus	Proteobacteria	232	29	Cap4_small ORF	0	
425	NMU55101.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
426	WP_169601013.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
427	WP_025612510.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
428	MBM5097193.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
429	MBM5179100.1		Vibrio parahaemolyticus	Proteobacteria	232	29	Cap4_small ORF	0	
430	WP_138940246.1		Vibrio alginolyticus	Proteobacteria	232	29	Cap4_small ORF	0	

431	WP_140126602.1		Vibrio parahaemolyticus	Proteobacteria	233	5	Cap4_small ORF	1	1851
432	WP_171357311.1		Vibrio rotiferianus	Proteobacteria	233	5	Cap4_small ORF	0	
433	WP_213904488.1		Vibrio alginolyticus	Proteobacteria	233	5	Cap4_small ORF	0	
434	WP_147724129.1		Vibrio parahaemolyticus	Proteobacteria	233	5	Cap4_small ORF	0	
435	MBE4090851.1		Vibrio parahaemolyticus	Proteobacteria	233	5	Cap4_small ORF	0	
436	WP_094766432.1		Vibrio parahaemolyticus	Proteobacteria	234	2	Cap4_small ORF	1	1851
437	MBE4166305.1		Vibrio parahaemolyticus	Proteobacteria	234	2	Cap4_small ORF	0	
438	WP_057758602.1		Psychrobacter sp. P11G3	Proteobacteria	235	1	Cap4_small ORF	1	1851
439	WP_021813675.1		Psychrobacter aquaticus	Proteobacteria	236	2	Cap4_small ORF	1	1852
440	WP_201580476.1		Psychrobacter sp. Pi2-52	Proteobacteria	236	2	Cap4_small ORF	0	
441	WP_067334455.1		Moraxella atlantae	Proteobacteria	237	1	Cap4_small ORF	1	1855
442	WP_201585664.1		Psychrobacter sp. HII-4	Proteobacteria	238	1	Cap4_small ORF	1	1871
443	WP_029913053.1		Hydrogenovibrio marinus	Proteobacteria	239	1	Cap4_small ORF	1	1843
444	WP_101154329.1		Colwellia sp. Bg11-28	Proteobacteria	240	1	Cap4_small ORF	1	1847
445	WP_182760531.1		Pseudoalteromonas sp. SG44-17	Proteobacteria	241	1	Cap4_small ORF	1	1841
446	WP_101082493.1		Psychromonas sp. Urea-02u-13	Proteobacteria	242	1	Cap4_small ORF	1	1843
447	WP_017091607.1		Vibrio splendidus	Proteobacteria	243	9	Cap4_small ORF	1	1845
448	WP_135382739.1		Vibrio tasmaniensis	Proteobacteria	243	9	Cap4_small ORF	0	
449	WP_022610029.1		Vibrio nigripulchritudo	Proteobacteria	243	9	Cap4_small ORF	0	
450	WP_107342107.1		Photobacterium angustum	Proteobacteria	243	9	Cap4_small ORF	0	
451	WP_057622974.1		Vibrio crassostreae	Proteobacteria	243	9	Cap4_small ORF	0	
452	WP_123539622.1		Vibrio crassostreae	Proteobacteria	243	9	Cap4_small ORF	0	
453	WP_017099029.1		Vibrio tasmaniensis	Proteobacteria	243	9	Cap4_small ORF	0	

454	WP_187993704.1		Vibrio harveyi	Proteobacteria	243	9	Cap4_small ORF	0	
455	WP_146489654.1		Vibrio sp. T20	Proteobacteria	243	9	Cap4_small ORF	0	
456	WP_084204731.1		Aeromonas popoffii	Proteobacteria	244	1	Cap4_small ORF	1	1842
457	OCW44989.1		Aeromonas caviae	Proteobacteria	245	1	Cap4_small ORF	1	1842
458	WP_164959900.1		Clostridium esterthecticum	Firmicutes	246	1	Cap4_small ORF	1	1840
459	WP_206459099.1		Anaerovorax sp. IOR16	Firmicutes	247	1	Cap4_small ORF	1	1811
460	MBO5055396.1		Lachnospiraceae bacterium	Firmicutes	248	1	Cap4_small ORF	1	1744
461	MBK9099491.1		bacterium	Unclassified	249	1	Cap4_small ORF	1	1847
462	KUO73511.1		Desulfosporosinus sp. BRH_c37	Firmicutes	250	1	Cap4_small ORF	1	1840
463	MBI5849358.1		Nitrospirae bacterium	Nitrospirae	251	1	Cap4_small ORF	1	1845
464	MBL8078945.1		Anaerolineales bacterium	Chloroflexi	252	1	Cap4_small ORF	1	1859
465	MBK9926314.1		Anaerolineales bacterium	Chloroflexi	253	1	Cap4_small ORF	1	1853
466	MBE7554033.1		Anaerolineales bacterium	Chloroflexi	254	1	Cap4_small ORF	1	1842
467	WP_133161910.1		Deinococcus koreensis	Deinococcus-Thermus	255	2	Cap4_small ORF	1	1856
468	PNY79182.1		Deinococcus koreensis	Deinococcus-Thermus	255	2	Cap4_small ORF	0	
469	WP_073644223.1		Nostoc calcicola	Cyanobacteria	256	1	Cap4_small ORF	1	1880
470	WP_190649414.1		Leptolyngbya sp. FACHB-1624	Cyanobacteria	257	1	Cap4_small ORF	1	1847
471	WP_105218067.1		Gloeocapsopsis sp. AAB1 = 1H9	Cyanobacteria	258	1	Cap4_small ORF	1	1843
472	WP_041054963.1		Jeotgalibacillus campisalis	Firmicutes	259	1	Cap4_small ORF	1	1837
473	WP_204499300.1		Aquibacillus albus	Firmicutes	260	1	Cap4_small ORF	1	1838
474	WP_076119697.1		Paenibacillus odorifer	Firmicutes	261	1	Cap4_small ORF	1	1840
475	WP_064449679.1		Bacillus wiedmannii	Firmicutes	262	1	Cap4_small ORF	1	1847
476	WP_209369553.1		Priestia megaterium	Firmicutes	263	1	Cap4_small ORF	1	1849
477	WP_214803739.1		Exiguobacterium sp. s194	Firmicutes	264	2	Cap4_small ORF	1	1835

478	WP_214876906.1		Exiguobacterium sp. CH10	Firmicutes	264	2	Cap4_small ORF	0	
479	WP_036694840.1		Paenibacillus sp. FSL R7-269	Firmicutes	265	1	Cap4_small ORF	1	1849
480	WP_113050511.1		Paenibacillus polymyxa	Firmicutes	266	1	Cap4_small ORF	1	1848
481	WP_166246131.1		Paenibacillus turpanensis	Firmicutes	267	1	Cap4_small ORF	1	1842
482	WP_009671434.1		unclassified Paenibacillus	Firmicutes	268	1	Cap4_small ORF	1	1849
483	WP_094018646.1		Paenibacillus rigui	Firmicutes	269	1	Cap4_small ORF	1	1846
484	WP_216478305.1		Paenibacillus sp. MSJ-6	Firmicutes	270	1	Cap4_small ORF	1	1849
485	WP_039276246.1		Paenibacillus polymyxa	Firmicutes	271	1	Cap4_small ORF	1	1847
486	WP_175394225.1		Paenibacillus xylanilyticus	Firmicutes	272	1	Cap4_small ORF	1	1848
487	WP_125012890.1		Flavobacterium macacae	Bacteroidetes	273	1	Cap4_small ORF	1	1838
488	MBC7410085.1		Arcicella sp.	Bacteroidetes	274	1	Cap4_small ORF	1	1833
489	RZJ36562.1		Flavobacterium sp.	Bacteroidetes	275	1	Cap4_small ORF	1	1835
490	WP_072944882.1		Flavobacterium granuli	Bacteroidetes	276	1	Cap4_small ORF	1	1835
491	OYX85552.1		Flavobacteriales bacterium 32-34-25	Bacteroidetes	277	1	Cap4_small ORF	1	1835
492	WP_135525443.1		Flavobacterium humi	Bacteroidetes	278	2	Cap4_small ORF	1	1836
493	MBA4154833.1		Flavobacterium sp.	Bacteroidetes	278	2	Cap4_small ORF	0	
494	THD32247.1		Flavobacterium johnsoniae	Bacteroidetes	279	1	Cap4_small ORF	1	1834
495	WP_058877893.1		Elizabethkingia anophelis	Bacteroidetes	280	1	Cap4_small ORF	1	1834
496	WP_109928798.1		Pedobacter paludis	Bacteroidetes	281	2	Cap4_small ORF	1	1830
497	PWS32626.1		Pedobacter paludis	Bacteroidetes	281	2	Cap4_small ORF	0	
498	WP_138819158.1		Pedobacter sp. KBS0701	Bacteroidetes	282	1	Cap4_small ORF	1	1831
499	WP_124587326.1		Pedobacter sp. KBW06	Bacteroidetes	283	1	Cap4_small ORF	1	1849
500	WP_205398410.1		Sphingobacterium siyangense	Bacteroidetes	284	1	Cap4_small ORF	1	1829

501	WP_073404449.1		Mucilaginibacter sp. OK098	Bacteroidetes	285	1	Cap4_small ORF	1	1840
502	RYY90843.1		Chitinophagaceae bacterium	Bacteroidetes	286	1	Cap4_small ORF	1	1833
503	WP_154491947.1		Elizabethkingia anophelis	Bacteroidetes	287	1	Cap4_small ORF	1	1833
504	WP_083729907.1		[Flexibacter] sp. ATCC 35208	Bacteroidetes	288	1	Cap4_small ORF	1	1815
505	MBL7684291.1		Flavipsychrobacter sp.	Bacteroidetes	289	1	Cap4_small ORF	1	1835
506	WP_091095677.1		Flavobacterium marinum	Bacteroidetes	290	1	Cap4_small ORF	1	1842
507	WP_044047851.1		Flavobacterium johnsoniae	Bacteroidetes	291	2	Cap4_small ORF	1	1842
508	ABQ06549.1		Flavobacterium johnsoniae UW101	Bacteroidetes	291	2	Cap4_small ORF	0	
509	WP_127337608.1		Flavobacterium cupreum	Bacteroidetes	292	2	Cap4_small ORF	1	1845
510	WP_132033485.1		Flavobacterium circumlabens	Bacteroidetes	292	2	Cap4_small ORF	0	
511	HAO08425.1		Chryseobacterium sp.	Bacteroidetes	293	1	Cap4_small ORF	1	1836
512	WP_065394170.1		Chryseobacterium artocarpi	Bacteroidetes	294	1	Cap4_small ORF	1	1836
513	WP_159478906.1		Chryseobacterium sp. 18068	Bacteroidetes	295	1	Cap4_small ORF	1	1838
514	WP_027374294.1		Chryseobacterium sp. UNC8MFCol	Bacteroidetes	296	1	Cap4_small ORF	1	1838
515	WP_204550454.1		Chitinophaga sancti	Bacteroidetes	297	1	Cap4_small ORF	1	1830
516	WP_157298060.1		Chitinophaga oryziterrae	Bacteroidetes	298	1	Cap4_small ORF	1	1832
517	WP_029034556.1		Salinimicrobium terrae	Bacteroidetes	299	1	Cap4_small ORF	1	1838
518	WP_170148130.1		Marinoscillum furvescens	Bacteroidetes	300	1	Cap4_small ORF	1	1842
519	WP_147864933.1		Mesonia sp. K4-1	Bacteroidetes	301	1	Cap4_small ORF	1	1841
520	WP_137090813.1		Roseivirga pacifica	Bacteroidetes	302	1	Cap4_small ORF	1	1830
521	MBT1690657.1		Fulvivirgaceae bacterium PWU37	Bacteroidetes	303	1	Cap4_small ORF	1	1835

522	MBS15070 33.1		Bacteroid etes bacterium	Bacteroidetes	304	1	Cap4_small ORF	1	1840
523	WP_15507 6198.1		Flavobact erium sp. MC2016-06	Bacteroidetes	305	1	Cap4_small ORF	1	1839
524	MBS15958 14.1		Bacteroid etes bacterium	Bacteroidetes	306	1	Cap4_small ORF	1	1837
525	WP_16633 6991.1		Sphingob acterium sp. IMCC256 78	Bacteroidetes	307	1	Cap4_small ORF	1	1839
526	WP_07287 7812.1		Muricaud a antarctica	Bacteroidetes	308	1	Cap4_small ORF	1	1837
527	WP_02675 2125.1		Sediminib acterium sp. C3	Bacteroidetes	309	1	Cap4_small ORF	1	1837
528	WP_12712 7804.1		Pseudofla vitalea rhizospha erae	Bacteroidetes	310	1	Cap4_small ORF	1	1837
529	MBS06473 40.1		Verrucom icrobia bacterium	Verrucomicro bia	311	1	Cap4_small ORF	1	1839
530	RBQ29471. 1		Arcobacte r sp. CECT 9230	Proteobacteria	312	1	Cap4_small ORF	1	1779
531	WP_10591 6069.1		Aliarcoba cter cryaeroph ilus	Proteobacteria	313	3	Cap4_small ORF	1	1782
532	WP_14106 1316.1		Aliarcoba cter cryaeroph ilus	Proteobacteria	313	3	Cap4_small ORF	0	
533	WP_06615 2737.1		Aliarcoba cter cryaeroph ilus	Proteobacteria	313	3	Cap4_small ORF	0	
534	WP_13012 6035.1		Pseudoalt eromonas sp. CO342X	Proteobacteria	314	1	Cap4_small ORF	1	1827
535	WP_04600 4462.1		Pseudoalt eromonas rubra	Proteobacteria	315	1	Cap4_small ORF	1	1826
536	WP_14302 1066.1		Belnapia rosea	Proteobacteria	316	2	Cap4_small ORF	1	1794
537	SDB48338. 1		Belnapia rosea	Proteobacteria	316	2	Cap4_small ORF	0	
538	WP_05575 6330.1		unclassifi ed Sphingom onas	Proteobacteria	317	1	Cap4_small ORF	1	1790
539	MBI56298 96.1		Elusimicr obia bacterium	Elusimicrobia	318	1	Cap4_small ORF	1	1832
540	VFJ42444. 1		Candidatu s Kentron sp. DK	Proteobacteria	319	1	Cap4_small ORF	1	1853
541	WP_00862 0089.1		Magnetos pirillum caucaseu m	Proteobacteria	320	1	Cap4_small ORF	1	1818
542	WP_00272 7977.1		Magnetos pirillum	Proteobacteria	321	1	Cap4_small ORF	1	1810

			molischianum						
543	MBI11983 02.1		Phenylobacterium sp.	Proteobacteria	322	1	Cap4	1	1905
544	CDX20111 .1		Mesorhizobium sp. ORS 3324	Proteobacteria	323	1	Cap4	1	1904
545	WP_16503 0825.1		Mesorhizobium camelthorni	Proteobacteria	324	1	Cap4	1	1915
546	AKQ41547 .2		Aurantiacibacter atlanticus	Proteobacteria	325	2	Cap4	1	1921
547	WP_16148 5441.1		Aurantiacibacter atlanticus	Proteobacteria	325	2	Cap4	0	
548	WP_15706 3835.1		Phaeobacter sp. 11ANDIMAR09	Proteobacteria	326	2	Cap4	1	1883
549	KPD10366. 1		Phaeobacter sp. 11ANDIMAR09	Proteobacteria	326	2	Cap4	0	
550	WP_07561 4768.1		Rhizobium taibaishanense	Proteobacteria	327	1	Cap4	1	1876
551	MBS50509 .1		Sphingobium sp.	Proteobacteria	328	2	Cap4	1	1882
552	MAX16492 .1		Sphingobium sp.	Proteobacteria	328	2	Cap4	0	
553	WP_13015 6344.1		Sphingomonas populi	Proteobacteria	329	1	Cap4	1	1900
554	WP_07601 4487.1		Pseudomonas sp. ATCC PTA-122608	Proteobacteria	330	1	Cap4	1	1892
555	WP_09536 4275.1		Stenotrophomonas rhizophila	Proteobacteria	331	1	Cap4	1	1882
556	WP_13052 3748.1		unclassified Pseudoxanthomonas	Proteobacteria	332	1	Cap4	1	1882
557	TGQ70983. 1		bacterium M00.F.Ca.ET.205.01.1.1	Unclassified	333	1	Cap4	1	1917
558	WP_14568 1115.1		Azospirillum brasilense	Proteobacteria	334	1	Cap4	1	1915
559	WP_16100 5426.1		Duganella sp. FT135W	Proteobacteria	335	1	Cap4	1	1916
560	WP_18375 4809.1		Pseudochelatococcus contaminans	Proteobacteria	336	1	Cap4	1	1915
561	WP_20203 6730.1		Agrobacterium fabrum	Proteobacteria	337	1	Cap4	1	1933
562	WP_15496 2989.1		Agrobacterium	Proteobacteria	338	1	Cap4	1	1915

			tumefaciens						
563	WP_155039922.1		Paracoccus litorisediminis	Proteobacteria	339	1	Cap4	1	1922
564	PZQ45636.1		Rhodovulum sulfidophilum	Proteobacteria	340	1	Cap4	1	1921
565	WP_197708121.1		Hartmannibacter diazotrophicus	Proteobacteria	341	1	Cap4	1	1920
566	WP_210563969.1		Chromobacterium violaceum	Proteobacteria	342	1	Cap4	1	1924
567	WP_175817829.1		Burkholderia diffusa	Proteobacteria	343	2	Cap4	1	1915
568	WP_089429257.1		Burkholderia sp. HI2500	Proteobacteria	343	2	Cap4	0	
569	WP_043177463.1		Burkholderia cepacia	Proteobacteria	344	1	Cap4	1	1915
570	WP_196426245.1		Pseudomonas aeruginosa	Proteobacteria	345	1	Cap4	1	1915
571	WP_044391693.1		Pseudomonas	Proteobacteria	346	3	Cap4	1	1922
572	OPK01787.1		Pseudomonas veronii	Proteobacteria	346	3	Cap4	0	
573	KPX72787.1		Pseudomonas syringae pv. maculicola	Proteobacteria	346	3	Cap4	0	
574	WP_169904308.1		Pseudomonas	Proteobacteria	347	2	Cap4	1	1922
575	WP_123335569.1		Pseudomonas chlororaphis	Proteobacteria	347	2	Cap4	0	
576	WP_196869176.1		Polaromonas sp. CG_9.11	Proteobacteria	348	2	Cap4	1	1902
577	MBG6075351.1		Polaromonas sp. CG_9.11	Proteobacteria	348	2	Cap4	0	
578	MBL8423739.1		Candidatus Accumulibacter phosphatis	Proteobacteria	349	1	Cap4	1	1915
579	WP_179539345.1		Pseudomonas flavescens	Proteobacteria	350	1	Cap4	1	1920
580	WP_148790447.1		Enterobacter cloacae	Proteobacteria	351	3	Cap4	1	1918
581	WP_025757705.1		Enterobacter cloacae	Proteobacteria	351	3	Cap4	0	
582	VAK78416.1		Enterobacter cloacae	Proteobacteria	351	3	Cap4	0	

583	WP_038747012.1		Burkholderia pseudomallei	Proteobacteria	352	2	Cap4	1	1921
584	KGC45797.1		Burkholderia pseudomallei	Proteobacteria	352	2	Cap4	0	
585	WP_186192801.1		Burkholderia gladioli	Proteobacteria	353	1	Cap4	1	1921
586	SEB25125.1		Variovorax sp. YR216	Proteobacteria	354	1	Cap4	1	1892
587	WP_114845998.1		Dyella tabacisoli	Proteobacteria	355	3	Cap4	1	1915
588	WP_212071673.1		Burkholderia ambifaria	Proteobacteria	355	3	Cap4	0	
589	WP_212118947.1		Burkholderia ambifaria	Proteobacteria	355	3	Cap4	0	
590	WP_063570106.1		Achromobacter ruhlandii	Proteobacteria	356	1	Cap4	1	1914
591	WP_060150733.1		Burkholderia multivorans	Proteobacteria	357	1	Cap4	1	1914
592	TBR74199.1		Burkholderiaceae bacterium	Proteobacteria	358	1	Cap4	1	1915
593	WP_006451195.1		Xanthomonas hortorum	Proteobacteria	359	1	Cap4	1	1915
594	WP_012584755.1		Thauera sp. MZ1T	Proteobacteria	360	3	Cap4	1	1915
595	TXH08733.1		Gammaproteobacteria bacterium	Proteobacteria	360	3	Cap4	0	
596	WP_095598160.1		Pseudomonas sp. PIC25	Proteobacteria	360	3	Cap4	0	
597	WP_179588002.1		Pigmentiphaga litoralis	Proteobacteria	361	1	Cap4	1	1915
598	WP_075076766.1		Stenotrophomonas	Proteobacteria	362	2	Cap4	1	1915
599	WP_016452367.1		Delftia	Proteobacteria	362	2	Cap4	0	
600	WP_183894703.1		Pseudomonas mendocina	Proteobacteria	363	1	Cap4	1	1921
601	WP_128677678.1		Pseudomonas aeruginosa	Proteobacteria	364	5	Cap4	1	1921
602	WP_208217650.1		Pseudomonas stutzeri	Proteobacteria	364	5	Cap4	0	
603	WP_059228111.1		Pseudomonas sp. EpS/L25	Proteobacteria	364	5	Cap4	0	
604	WP_134240915.1		Pseudomonas aeruginosa	Proteobacteria	364	5	Cap4	0	
605	WP_121373432.1		Pseudomonas	Proteobacteria	364	5	Cap4	0	

			aeruginosa						
606	WP_095602896.1		Halomonas humidisoli	Proteobacteria	365	1	Cap4	1	1919
607	WP_066452619.1		Castellaniella caeni	Proteobacteria	366	1	Cap4	1	1915
608	WP_056139590.1		Duganella sp. Leaf61	Proteobacteria	367	1	Cap4	1	1915
609	WP_183681877.1		unclassified Janthinobacterium	Proteobacteria	368	1	Cap4	1	1913
610	WP_070258275.1		Janthinobacterium lividum	Proteobacteria	369	3	Cap4	1	1914
611	WP_070302776.1		Janthinobacterium sp. HH104	Proteobacteria	369	3	Cap4	0	
612	WP_070289229.1		Janthinobacterium sp. HH106	Proteobacteria	369	3	Cap4	0	
613	WP_015666588.1		Bradyrhizobium oligotrophicum	Proteobacteria	370	1	Cap4	1	1917
614	WP_213253131.1		Bradyrhizobium sp. sBnM-33	Proteobacteria	371	1	Cap4	1	1917
615	WP_005990224.1		Desulfovibrio fructosivorans	Proteobacteria	372	1	Cap4	1	1924
616	WP_102111204.1		Niveispirillum cyanobacteriorum	Proteobacteria	373	1	Cap4	1	1940
617	WP_174028052.1		Agrobacterium rhizogenes	Proteobacteria	374	1	Cap4	1	1923
618	WP_106795716.1		Rhizobium sp. H4	Proteobacteria	375	1	Cap4	1	1943
619	WP_027488451.1		Rhizobium undicola	Proteobacteria	376	1	Cap4	1	1937
620	MBL4828995.1		Aliivibrio sp.	Proteobacteria	377	1	Cap4	1	1880
621	PUA27436.1		Cellvibrio sp. 79	Proteobacteria	378	1	Cap4	1	1874
622	MBC8028502.1		Pyrinomonadaceae bacterium	Acidobacteria	379	1	Cap4	1	1887
623	MBU4201455.1		Verrucomicrobia bacterium	Verrucomicrobia	380	1	Cap4	1	1885
624	MBI4620341.1		Desulfobacterales bacterium	Proteobacteria	381	1	Cap4	1	1882
625	MBN2592708.1		Sedimentisphaerales bacterium	Planctomycetes	382	1	Cap4	1	1884
626	MBF2758891.1		Ectothiorhodospiraceae bacterium AqS1	Proteobacteria	383	1	Cap4	1	1849

627	WP_036202640.1		Marinobacter sp. MCTG268	Proteobacteria	384	1	Cap4	1	1886
628	WP_049869384.1		Paraburkholderia phytofirmans	Proteobacteria	385	2	Cap4	1	1880
629	ACD20777.1		Paraburkholderia phytofirmans PsJN	Proteobacteria	385	2	Cap4	0	
630	NQZ16874.1		Idiomarina sp.	Proteobacteria	386	1	Cap4	1	1878
631	WP_008248130.1		Zhongshania aliphaticivorans	Proteobacteria	387	1	Cap4	1	1878
632	WP_082085219.1		Magnetospira sp. QH-2	Proteobacteria	388	2	Cap4	1	1883
633	CCQ72458.1		Magnetospira sp. QH-2	Proteobacteria	388	2	Cap4	0	
634	WP_161816019.1		Steroidobacter agaridevorans	Proteobacteria	389	2	Cap4	1	1883
635	GFE84412.1		Steroidobacter agaridevorans	Proteobacteria	389	2	Cap4	0	
636	MBI4480217.1		Acidobacteria bacterium	Acidobacteria	390	1	Cap4	1	1884
637	RJP17957.1		Candidatus Abyssubacteria bacterium SURF_5	Candidatus Abyssubacteria	391	1	Cap4	1	1879
638	WP_111641800.1		Marinimicrobium alkaliphilum	Proteobacteria	392	1	Cap4	1	1885
639	MBU1420054.1		Proteobacteria bacterium	Proteobacteria	393	1	Cap4	1	1885
640	MBN1271128.1		Candidatus Aminicenantes bacterium	Unclassified	394	1	Cap4	1	1885
641	WP_013970782.1		Pseudomonas putida	Proteobacteria	395	3	Cap4	1	1894
642	WP_054926058.1		Pseudomonas sp. NBRC 111142	Proteobacteria	395	3	Cap4	0	
643	WP_054888530.1		Pseudomonas sp. NBRC 111118	Proteobacteria	395	3	Cap4	0	
644	WP_168084026.1		Pseudomonas	Proteobacteria	396	1	Cap4	1	1886
645	WP_036997970.1		Pseudomonas chlororaphis	Proteobacteria	397	1	Cap4	1	1879
646	WP_153386316.1		Pseudomonas helleri	Proteobacteria	398	1	Cap4	1	1884

647	WP_100217142.1		Pseudomonas tolaasii	Proteobacteria	399	1	Cap4	1	1882
648	WP_120528930.1		Coralloccus exercitus	Proteobacteria	400	1	Cap4	1	1874
649	WP_206797438.1		Coralloccus sp. NCSPR001	Proteobacteria	401	1	Cap4	1	1874
650	RKI17867.1		Coralloccus sp. AB030	Proteobacteria	402	2	Cap4	1	1879
651	WP_199745904.1		Coralloccus sp. AB030	Proteobacteria	402	2	Cap4	0	
652	WP_141643613.1		Myxococcus llanfairpwllgwyngyllgogerychwyrndrobwllllantysiliogogogochensis	Proteobacteria	403	1	Cap4	1	1879
653	WP_167336242.1		Xanthomonas sacchari	Proteobacteria	404	1	Cap4	1	1880
654	WP_108746679.1		Stenotrophomonas maltophilia group	Proteobacteria	405	1	Cap4	1	1880
655	VUQ16054.1		Stenotrophomonas maltophilia	Proteobacteria	406	4	Cap4	1	1879
656	WP_132810108.1		Stenotrophomonas sp. ASS1	Proteobacteria	406	4	Cap4	0	
657	WP_180844250.1		Stenotrophomonas maltophilia	Proteobacteria	406	4	Cap4	0	
658	MBH1701744.1		Stenotrophomonas maltophilia	Proteobacteria	406	4	Cap4	0	
659	MBE7423791.1		Zoogloeaceae bacterium	Proteobacteria	407	2	Cap4	1	1882
660	MBI4988351.1		Rhodocyclales bacterium	Proteobacteria	407	2	Cap4	0	
661	WP_158219724.1		Ideonella sp. A 288	Proteobacteria	408	1	Cap4	1	1878
662	WP_114130199.1		Cupriavidus necator	Proteobacteria	409	1	Cap4	1	1879
663	WP_059834010.1		Burkholderia sp. MSMB1835	Proteobacteria	410	3	Cap4	1	1881
664	WP_175824223.1		Burkholderia sp. BCC0419	Proteobacteria	410	3	Cap4	0	
665	WP_166955976.1		Burkholderia sp. D-99	Proteobacteria	410	3	Cap4	0	
666	WP_156969313.1		Lysobacter	Proteobacteria	411	1	Cap4	1	1882
667	WP_096376343.1		Lysobacter	Proteobacteria	412	1	Cap4	1	1883

			enzymogenes						
668	WP_192027778.1		Pseudoxanthomonas sp. CAU 1598	Proteobacteria	413	1	Cap4	1	1883
669	PLC02856.1		Variovorax sp. RO1	Proteobacteria	414	2	Cap4	1	1884
670	WP_180839037.1		Variovorax sp. RO1	Proteobacteria	414	2	Cap4	0	
671	WP_183026421.1		Variovorax sp. UMC13	Proteobacteria	415	2	Cap4	1	1883
672	MBB1602015.1		Variovorax sp. UMC13	Proteobacteria	415	2	Cap4	0	
673	WP_161499728.1		Limnohabitans sp. TS-CS-82	Proteobacteria	416	2	Cap4	1	1883
674	PQA81913.1		Limnohabitans sp. TS-CS-82	Proteobacteria	416	2	Cap4	0	
675	MBC7395552.1		Variovorax sp.	Proteobacteria	417	1	Cap4	1	1883
676	WP_157613585.1		Variovorax paradoxus	Proteobacteria	418	1	Cap4	1	1867
677	MBD8655634.1		Oxalobacteraceae sp. CFBP 13730	Proteobacteria	419	1	Cap4_short	1	1599
678	WP_060498883.1		Pseudomonas	Proteobacteria	420	1	Cap4_short	1	1572
679	WP_183701822.1		Pseudomonas fluorescens	Proteobacteria	421	2	Cap4_short	1	1574
680	WP_177020193.1		unclassified Pseudomonas	Proteobacteria	421	2	Cap4_short	0	
681	WP_183141710.1		Pseudomonas savastanoi	Proteobacteria	422	2	Cap4_short	1	1573
682	RMQ97099.1		Pseudomonas savastanoi pv. glycinea	Proteobacteria	422	2	Cap4_short	0	
683	WP_146746239.1		Rhodovulum viride	Proteobacteria	423	1	Cap4	1	1872
684	WP_165828113.1		Rhodovulum sp. BSW8	Proteobacteria	424	2	Cap4	1	1871
685	RBO53162.1		Rhodovulum sp. BSW8	Proteobacteria	424	2	Cap4	0	
686	WP_047039675.1		Klebsiella aerogenes	Proteobacteria	425	1	Cap4	1	1870
687	WP_034834205.1		Endozoicomonas numazuensis	Proteobacteria	426	1	Cap4	1	1872
688	MBO6581984.1		Hyphomonas sp.	Proteobacteria	427	1	Cap4	1	1877
689	ESQ14270.1		Thiohalocapsa sp. PB-PSB1	Proteobacteria	428	2	Cap4	1	1874
690	MBL4540318.1		Rhodobacteraceae bacterium	Proteobacteria	428	2	Cap4	0	

691	WP_206682582.1		Gemmobacter straminiformis	Proteobacteria	429	2	Cap4	1	1874
692	MBC2835161.1		Gemmobacter straminiformis	Proteobacteria	429	2	Cap4	0	
693	MBB93460.1		Rhodobacteraceae bacterium	Proteobacteria	430	1	Cap4	1	1873
694	WP_202673318.1		Serratia fonticola	Proteobacteria	431	1	Cap4	1	1867
695	WP_139964370.1	KvAvs3	Klebsiella variicola	Proteobacteria	432	20	Cap4	1	1866
696	WP_200009871.1		Citrobacter freundii	Proteobacteria	432	20	Cap4	0	
697	WP_044711644.1		Citrobacter freundii	Proteobacteria	432	20	Cap4	0	
698	WP_181518166.1		Citrobacter freundii	Proteobacteria	432	20	Cap4	0	
699	WP_174361563.1		Citrobacter gillenii	Proteobacteria	432	20	Cap4	0	
700	WP_176984315.1		Serratia liquefaciens	Proteobacteria	432	20	Cap4	0	
701	WP_045441742.1		Citrobacter sp. S-77	Proteobacteria	432	20	Cap4	0	
702	WP_100194215.1		Citrobacter freundii	Proteobacteria	432	20	Cap4	0	
703	WP_175576746.1		Yersinia enterocolitica	Proteobacteria	432	20	Cap4	0	
704	WP_157985453.1		Pectobacterium parmentieri	Proteobacteria	432	20	Cap4	0	
705	WP_202785623.1		Enterobacter asburiae	Proteobacteria	432	20	Cap4	0	
706	WP_075808385.1		Raoultella ornithinolytica	Proteobacteria	432	20	Cap4	0	
707	AZS56095.1		Pectobacterium parmentieri	Proteobacteria	432	20	Cap4	0	
708	WP_205545364.1		Pectobacterium brasiliense	Proteobacteria	432	20	Cap4	0	
709	WP_050088136.1		Yersinia intermedia	Proteobacteria	432	20	Cap4	0	
710	WP_140475773.1		Ewingella americana	Proteobacteria	432	20	Cap4	0	
711	WP_044203233.1		Pectobacterium	Proteobacteria	432	20	Cap4	0	
712	WP_085117486.1		Serratia proteamaculans	Proteobacteria	432	20	Cap4	0	
713	WP_205538168.1		Pectobacterium versatile	Proteobacteria	432	20	Cap4	0	
714	MBN3059436.1		Pectobacterium versatile	Proteobacteria	432	20	Cap4	0	
715	WP_106843532.1		Aeromonas veronii	Proteobacteria	433	1	Cap4	1	1871
716	WP_193860571.1		Aeromonas veronii	Proteobacteria	434	1	Cap4	1	1871

717	WP_136808727.1		Desulfopila sp. IMCC35005	Proteobacteria	435	1	Cap4	1	1868
718	WP_092350897.1		Desulfuromusa kysingii	Proteobacteria	436	1	Cap4	1	1872
719	WP_022966737.1		Pseudomonas caeni	Proteobacteria	437	1	Cap4	1	1869
720	VCU61052.1		Epibacterium mobile	Proteobacteria	438	2	Cap4	1	1872
721	WP_126138136.1		Epibacterium mobile	Proteobacteria	438	2	Cap4	0	
722	HDZ07499.1		Pseudohongiella sp.	Proteobacteria	439	1	Cap4	1	1873
723	WP_046107775.1		Devosia geojensis	Proteobacteria	440	1	Cap4	1	1867
724	WP_127519184.1		Sinorhizobium meliloti	Proteobacteria	441	1	Cap4	1	1864
725	WP_168278901.1		Rhizobium leguminosarum	Proteobacteria	442	3	Cap4	1	1865
726	WP_163874514.1		Rhizobium laguerreae	Proteobacteria	442	3	Cap4	0	
727	WP_180696942.1		Rhizobium changzhiense	Proteobacteria	442	3	Cap4	0	
728	WP_127326634.1		Mesorhizobium sp. M7A.F.Ca.MR.245.00.0.0	Proteobacteria	443	1	Cap4	1	1875
729	TIO79319.1		Mesorhizobium sp.	Proteobacteria	444	2	Cap4	1	1872
730	TJW02177.1		Mesorhizobium sp.	Proteobacteria	444	2	Cap4	0	
731	WP_020808916.1		Agrobacterium	Proteobacteria	445	3	Cap4	1	1872
732	WP_152849892.1		Brucella intermedia	Proteobacteria	445	3	Cap4	0	
733	WP_174012839.1		Agrobacterium tumefaciens	Proteobacteria	445	3	Cap4	0	
734	KRD46723.1		Acidovorax sp. Root275	Proteobacteria	446	2	Cap4	1	1873
735	WP_198928138.1		Acidovorax sp. Root275	Proteobacteria	446	2	Cap4	0	
736	WP_183303886.1		Comamonas terrigena	Proteobacteria	447	1	Cap4	1	1872
737	WP_035373824.1		Acetobacter persici	Proteobacteria	448	1	Cap4	1	1865
738	PDT81821.1		Sinorhizobium sp. BJ1	Proteobacteria	449	2	Cap4	1	1865
739	WP_210202630.1		Sinorhizobium sp. BJ1	Proteobacteria	449	2	Cap4	0	
740	WP_097155695.1		Cohaesibacter	Proteobacteria	450	1	Cap4	1	1865

			gelatinilyticus						
741	WP_212491133.1		Bradyrhizobium liaoningense	Proteobacteria	451	1	Cap4	1	1870
742	WP_210205639.1		Pseudaminobacter salicylatoxidans	Proteobacteria	452	2	Cap4	1	1871
743	PWJ83728.1		Pseudaminobacter salicylatoxidans	Proteobacteria	452	2	Cap4	0	
744	WP_026107083.1		Dyella ginsengisoli	Proteobacteria	453	1	Cap4	1	1874
745	WP_190971884.1		Bordetella bronchiseptica	Proteobacteria	454	6	Cap4	1	1872
746	WP_015040930.1		Bordetella bronchiseptica	Proteobacteria	454	6	Cap4	0	
747	WP_134839053.1		Kerstersia gyiorum	Proteobacteria	454	6	Cap4	0	
748	QET69295.1		Bordetella bronchiseptica	Proteobacteria	454	6	Cap4	0	
749	KCV55666.1		Bordetella bronchiseptica 7E71	Proteobacteria	454	6	Cap4	0	
750	WP_191579041.1		Achromobacter insolitus	Proteobacteria	454	6	Cap4	0	
751	VUP95512.1		Stenotrophomonas maltophilia	Proteobacteria	455	2	Cap4	1	1872
752	MBH1529354.1		Stenotrophomonas maltophilia	Proteobacteria	455	2	Cap4	0	
753	MBD8643127.1		Stenotrophomonas sp. CFBP 13724	Proteobacteria	456	2	Cap4	1	1872
754	WP_197974309.1		Stenotrophomonas sp. CFBP 13724	Proteobacteria	456	2	Cap4	0	
755	RJG01630.1		Noviherbaspirillum sp. K1S02-23	Proteobacteria	457	2	Cap4	1	1871
756	WP_199688262.1		Noviherbaspirillum sp. K1S02-23	Proteobacteria	457	2	Cap4	0	
757	NNG25530.1		Massilia aromaticivorans	Proteobacteria	458	2	Cap4	1	1870
758	WP_212761056.1		Massilia aromaticivorans	Proteobacteria	458	2	Cap4	0	
759	KHA77842.1		Janthinobacterium lividum	Proteobacteria	459	2	Cap4	1	1874
760	WP_205412554.1		Janthinobacterium lividum	Proteobacteria	459	2	Cap4	0	

761	RSZ55082.1		Massilia atriviolacea	Proteobacteria	460	2	Cap4	1	1870
762	WP_198420959.1		Massilia atriviolacea	Proteobacteria	460	2	Cap4	0	
763	WP_210562949.1		Chromobacterium violaceum	Proteobacteria	461	1	Cap4	1	1871
764	WP_058132227.1		Pseudomonas aeruginosa	Proteobacteria	462	14	Cap4	1	1868
765	WP_171950735.1		Pseudomonas aeruginosa	Proteobacteria	462	14	Cap4	0	
766	WP_160296220.1		Pseudomonas aeruginosa	Proteobacteria	462	14	Cap4	0	
767	KSP58726.1		Pseudomonas aeruginosa	Proteobacteria	462	14	Cap4	0	
768	WP_061196300.1		Pseudomonas aeruginosa	Proteobacteria	462	14	Cap4	0	
769	WP_061195977.1		Pseudomonas aeruginosa	Proteobacteria	462	14	Cap4	0	
770	WP_061195254.1		Pseudomonas aeruginosa	Proteobacteria	462	14	Cap4	0	
771	WP_049950042.1		Pseudomonas aeruginosa	Proteobacteria	462	14	Cap4	0	
772	WP_134305908.1		Pseudomonas aeruginosa	Proteobacteria	462	14	Cap4	0	
773	MBP0226420.1		Pseudomonas aeruginosa	Proteobacteria	462	14	Cap4	0	
774	WP_070654069.1		Pseudomonas aeruginosa	Proteobacteria	462	14	Cap4	0	
775	SVK43557.1		Acinetobacter baumannii	Proteobacteria	462	14	Cap4	0	
776	WP_109413191.1		Pseudomonas aeruginosa	Proteobacteria	462	14	Cap4	0	
777	WP_137137892.1		Pseudomonas asiatica	Proteobacteria	462	14	Cap4	0	
778	AGE82295.1		Pseudomonas syringae pv. actinidiae	Proteobacteria	463	7	Cap4	1	1870
779	WP_017682848.1		Pseudomonas syringae	Proteobacteria	463	7	Cap4	0	

780	WP_210672407.1		Pseudomonas protegens	Proteobacteria	463	7	Cap4	0	
781	WP_085741148.1		Pseudomonas syringae	Proteobacteria	463	7	Cap4	0	
782	OSN39168.1		Pseudomonas syringae pv. actinidiae	Proteobacteria	463	7	Cap4	0	
783	WP_080644453.1		Pseudomonas protegens	Proteobacteria	463	7	Cap4	0	
784	AGL84532.1		Pseudomonas protegens CHA0	Proteobacteria	463	7	Cap4	0	
785	WP_168758817.1		Pseudomonas umsongensis	Proteobacteria	464	1	Cap4	1	1870
786	WP_095050681.1		Pseudomonas sp. Irchel s3h9	Proteobacteria	465	1	Cap4	1	1870
787	WP_150734435.1		Pseudomonas fluorescens	Proteobacteria	466	1	Cap4	1	1868
788	WP_102062182.1		Pseudomonas sp. QC2	Proteobacteria	467	3	Cap4	1	1868
789	SAM33843.1		Pseudomonas sp. 1 R 17	Proteobacteria	467	3	Cap4	0	
790	WP_083221221.1		Pseudomonas sp. 1 R 17	Proteobacteria	467	3	Cap4	0	
791	WP_097276823.1		Burkholderia sp. D7	Proteobacteria	468	1	Cap4	1	1871
792	WP_079503607.1		Paraburkholderia hospita	Proteobacteria	469	1	Cap4	1	1871
793	QHE90393.1		Pandoraea fibrosis	Proteobacteria	470	2	Cap4	1	1867
794	WP_201278438.1		Pandoraea fibrosis	Proteobacteria	470	2	Cap4	0	
795	AZQ54374.1		Burkholderia cenocepacia	Proteobacteria	471	2	Cap4	1	1871
796	WP_206514407.1		Burkholderia cenocepacia	Proteobacteria	471	2	Cap4	0	
797	WP_175226136.1		Paraburkholderia humisilvae	Proteobacteria	472	2	Cap4	1	1870
798	CAB3752592.1		Paraburkholderia humisilvae	Proteobacteria	472	2	Cap4	0	
799	WP_212180667.1		Burkholderia vietnamiensis	Proteobacteria	473	1	Cap4	1	1871
800	WP_208450126.1		Burkholderia ambifaria	Proteobacteria	474	1	Cap4	1	1870

801	WP_216521656.1		Tissierella sp. MSJ-40	Firmicutes	475	1	Cap4	1	1824
802	WP_061309128.1		Clostridium botulinum	Firmicutes	476	1	Cap4	1	1805
803	WP_010964745.1		Clostridium	Firmicutes	477	2	Cap4	1	1804
804	WP_017750479.1		Clostridium tyrobutyricum	Firmicutes	477	2	Cap4	0	
805	WP_069973853.1		Geosporobacter ferrireducens	Firmicutes	478	1	Cap4	1	1830
806	HHY82268.1		Clostridiales bacterium	Firmicutes	479	1	Cap4	1	1832
807	WP_036724482.1		Paenibacillus sp. FSL R7-277	Firmicutes	480	1	Cap4	1	1829
808	WP_145410825.1		Paenibacillus xylanexedens	Firmicutes	481	1	Cap4	1	1826
809	PHV72393.1		Lachnospiraceae bacterium	Firmicutes	482	1	Cap4	1	1820
810	WP_178139231.1		Hathewaya proteolytica	Firmicutes	483	1	Cap4	1	1824
811	WP_129028212.1		Clostridium tetani	Firmicutes	484	1	Cap4	1	1830
812	HHV13072.1		Clostridiales bacterium	Firmicutes	485	1	Cap4	1	1829
813	WP_097015330.1		Anaerocolumna aminovalerica	Firmicutes	486	3	Cap4	1	1829
814	MBU5310475.1		Tissierella carlieri	Firmicutes	486	3	Cap4	0	
815	WP_216554187.1		Tissierella carlieri	Firmicutes	486	3	Cap4	0	
816	WP_146602543.1		Novipirellula aureliae	Planctomycetes	487	1	Cap4_small ORF	1	1885
817	KAA3600031.1		Candidatus Scalindua sp.	Planctomycetes	488	1	Cap4_small ORF	1	1843
818	WP_094602871.1		Sporomusa silvacetica	Firmicutes	489	1	Cap4_small ORF	1	1843
819	WP_068186743.1		Macellibacteroides sp. HH-ZS	Bacteroidetes	490	1	Cap4_small ORF	1	1833
820	WP_130855558.1		Olivibacter jilunii	Bacteroidetes	491	1	Cap4_small ORF	1	1849
821	WP_132397066.1		Pedobacter sp. CF074	Bacteroidetes	492	1	Cap4_small ORF	1	1837
822	WP_131554830.1		Pedobacter sp. RP-1-13	Bacteroidetes	493	1	Cap4_small ORF	1	1838

823	WP_086947453.1		Sphingobacterium sp. JB170	Bacteroidetes	494	1	Cap4_small ORF	1	1836
824	WP_055144806.1		Jiulongibacter sediminis	Bacteroidetes	495	1	Cap4_small ORF	1	1841
825	MBC7525461.1		Flavobacterium sp.	Bacteroidetes	496	1	Cap4_small ORF	1	1838
826	WP_039107859.1		Flavobacterium sp. AED	Bacteroidetes	497	1	Cap4_small ORF	1	1836
827	SFB91135.1		Algibacter lectus	Bacteroidetes	498	1	Cap4_small ORF	1	1858
828	WP_034889475.1		Gillisia sp. Hel_I_29	Bacteroidetes	499	1	Cap4_small ORF	1	1854
829	WP_209404747.1		Pseudozobellia sp. WGM2	Bacteroidetes	500	1	Cap4_small ORF	1	1862
830	WP_117021111.1		Arenibacter sp. P308M17	Bacteroidetes	501	1	Cap4_small ORF	1	1862
831	WP_111410434.1		Flavobacterium aquariorum	Bacteroidetes	502	1	Cap4_small ORF	1	1862
832	WP_123879860.1		Chryseobacterium carnipullorum	Bacteroidetes	503	1	Cap4_small ORF	1	1860
833	WP_047096976.1		Chryseobacterium indologenes	Bacteroidetes	504	6	Cap4_small ORF	1	1856
834	WP_002978688.1		Chryseobacterium gleum	Bacteroidetes	504	6	Cap4_small ORF	0	
835	WP_138402056.1		Chryseobacterium indologenes	Bacteroidetes	504	6	Cap4_small ORF	0	
836	WP_120231228.1		Chryseobacterium	Bacteroidetes	504	6	Cap4_small ORF	0	
837	WP_088583894.1		Chryseobacterium indologenes	Bacteroidetes	504	6	Cap4_small ORF	0	
838	WP_034848631.1		Elizabethkingia anophelis	Bacteroidetes	504	6	Cap4_small ORF	0	
839	WP_073220979.1		Aequorivita viscosa	Bacteroidetes	505	1	Cap4_small ORF	1	1859
840	WP_067150180.1		Tamlana agarivorans	Bacteroidetes	506	1	Cap4_small ORF	1	1862
841	WP_041516098.1		Flavobacterium hibernum	Bacteroidetes	507	1	Cap4_small ORF	1	1866
842	WP_076506214.1		Chryseobacterium shigense	Bacteroidetes	508	1	Cap4_small ORF	1	1857
843	WP_040278309.1		Psychroserpens damuponensis	Bacteroidetes	509	1	Cap4_small ORF	1	1845
844	WP_019388854.1		Arenitalea lutea	Bacteroidetes	510	1	Cap4_small ORF	1	1851
845	WP_187561719.1		Kordia aestuariivivens	Bacteroidetes	511	1	Cap4_small ORF	1	1852

846	WP_006990322.1		Gillisia limnaea	Bacteroidetes	512	1	Cap4_small ORF	1	1849
847	WP_006987912.1		Gillisia limnaea	Bacteroidetes	513	1	Cap4_small ORF	1	1852
848	WP_055436658.1		Lacinutrix algicola	Bacteroidetes	514	1	Cap4_small ORF	1	1854
849	WP_111682188.1		Winogradskyella tangerina	Bacteroidetes	515	1	Cap4_small ORF	1	1854
850	MBF0137168.1		Magnetococcales bacterium	Proteobacteria	516	1	Cap4	1	1866
851	WP_066178558.1		[Arcobacter] porcinus	Proteobacteria	517	1	TIR	1	1607
852	WP_148571801.1	AcAvs3	Aliarcobacter cryaerophilus	Proteobacteria	518	1	TIR	1	1809
853	WP_012083877.1		Sulfurovum sp. NBC37-1	Proteobacteria	519	1	TIR	1	1776
854	HAS88330.1		Desulfovibrio sp.	Proteobacteria	520	1	Cap4	1	1800
855	WP_146683611.1		Limihaloglobus sulfuriphilus	Planctomycetes	521	1	Cap4_small ORF	1	1838
856	PYQ27322.1		Acidobacteria bacterium	Acidobacteria	522	1	Cap4_small ORF	1	1799
857	WP_143195829.1		Archangium sp. Cb G35	Proteobacteria	523	2	Cap4_small ORF	1	1892
858	OJT20011.1		Archangium sp. Cb G35	Proteobacteria	523	2	Cap4_small ORF	0	
859	BCS35486.1		Vicinamibacteria bacterium TBR-22	Acidobacteria	524	1	Cap4_small ORF	1	1874
860	WP_162907940.1		Allorhizocola rhizosphaerae	Actinobacteria	525	1	Cap4_small ORF	1	1879
861	WP_197844218.1		Pseudomonas parafulva	Proteobacteria	526	1	Cap4_small ORF	1	1848
862	WP_155769229.1		Pseudomonas putida	Proteobacteria	527	1	Cap4_small ORF	1	1847
863	WP_136385680.1		Azoarcus rhizosphaerae	Proteobacteria	528	1	Cap4_small ORF	1	1852
864	WP_177105190.1		Pseudomonas gingeri	Proteobacteria	529	1	Cap4_small ORF	1	1848
865	WP_150933344.1		Microbispora cellulosiformans	Actinobacteria	530	1	Cap4_small ORF	1	1817
866	WP_157910582.1		Frankia sp. BMG5.36	Actinobacteria	531	2	Cap4_small ORF	1	1797
867	OHV47811.1		Frankia sp. BMG5.36	Actinobacteria	531	2	Cap4_small ORF	0	
868	WP_189595992.1		Streptomyces flavoviridis	Actinobacteria	532	1	Cap4_small ORF	1	1814

869	WP_157901132.1		Streptomyces davaonensis	Actinobacteria	533	2	Cap4_small ORF	1	1814
870	CCK31985.1		Streptomyces davaonensis JCM 4913	Actinobacteria	533	2	Cap4_small ORF	0	
871	WP_188289751.1		Streptomyces sp. TRM68416	Actinobacteria	534	1	Cap4_small ORF	1	1806
872	PIG11499.1		Streptomyces sp. 1121.2	Actinobacteria	535	2	Cap4_small ORF	1	1804
873	WP_143688148.1		Streptomyces sp. 1121.2	Actinobacteria	535	2	Cap4_small ORF	0	
874	WP_015612833.1		Streptomyces fulvissimus	Actinobacteria	536	1	Cap4_small ORF	1	1801
875	WP_143573585.1		Streptomyces acidiscabies	Actinobacteria	537	1	Cap4_small ORF	1	1801
876	AWW36357.1		Streptomyces cadmiisoli	Actinobacteria	538	2	Cap4_small ORF	1	1803
877	WP_162624868.1		Streptomyces cadmiisoli	Actinobacteria	538	2	Cap4_small ORF	0	
878	WP_165440384.1		Streptomyces sp. CNZ288	Actinobacteria	539	1	Cap4_small ORF	1	1802
879	WP_159059662.1		Streptomyces	Actinobacteria	540	3	Cap4_small ORF	1	1802
880	WP_061334573.1		Streptomyces sp. LBUM 1481	Actinobacteria	540	3	Cap4_small ORF	0	
881	MBP5874305.1		Streptomyces sp. LBUM 1477	Actinobacteria	540	3	Cap4_small ORF	0	
882	MBN1461109.1		Armatimonadetes bacterium	Armatimonadetes	541	1	Cap4_small ORF	1	1836
883	VTR01910.1		Clostridioides difficile	Firmicutes	542	3	Cap4_small ORF	1	1828
884	VTS53095.1		Gordonia terrae	Actinobacteria	542	3	Cap4_small ORF	0	
885	WP_004019792.1		Gordonia terrae	Actinobacteria	542	3	Cap4_small ORF	0	
886	WP_128970146.1		Rhodococcus opacus	Actinobacteria	543	1	Cap4_small ORF	1	1835
887	MBC14845.1		Rhodothermaceae bacterium	Bacteroidetes	544	1	Cap4_small ORF	1	1878
888	SMC07650.1		Rubrobacter radiotolerans DSM 5868	Actinobacteria	545	2	Cap4_small ORF	1	1882
889	WP_084362652.1		Rubrobacter radiotolerans	Actinobacteria	545	2	Cap4_small ORF	0	

890	SDT93178.1		Verrucomicrobium sp. GAS474	Verrucomicrobia	546	2	Cap4	1	1837
891	WP_157895107.1		Verrucomicrobium sp. GAS474	Verrucomicrobia	546	2	Cap4	0	
892	MPR31822.1		Cytophagaceae bacterium SJW1-29	Bacteroidetes	547	1	Cap4_small ORF	1	1881
893	WP_162386972.1		Spirosoma endbachense	Bacteroidetes	548	2	Cap4_small ORF	1	1885
894	QHV96566.1		Spirosoma endbachense	Bacteroidetes	548	2	Cap4_small ORF	0	
895	WP_015333049.1		Fibrella aestuarina	Bacteroidetes	549	1	Cap4_small ORF	1	1881
896	AHM61617.1		Flammeovirgaceae bacterium 311	Bacteroidetes	550	1	Cap4_small ORF	1	1875
897	TDN37031.1		Hymenobacter sp. UV11	Bacteroidetes	551	2	Cap4_small ORF	1	1876
898	WP_133658947.1		Hymenobacter sp. UV11	Bacteroidetes	551	2	Cap4_small ORF	0	
899	TVT36967.1		Hymenobacter sp. Fur1	Actinobacteria	552	1	Cap4_small ORF	1	1876
900	MBU3994120.1		Alphaproteobacteria bacterium	Proteobacteria	553	1	Cap4	1	1855
901	WP_058807484.1		unclassified Sphingopyxis	Proteobacteria	554	1	Cap4	1	1893
902	WP_062770719.1		Tistrella mobilis	Proteobacteria	555	1	Cap4	1	1801
903	WP_014747231.1		Tistrella mobilis	Proteobacteria	556	1	Cap4	1	1801
904	GEO82775.1		Pararhodospirillum oryzae	Proteobacteria	557	2	Cap4	1	1883
905	WP_147164802.1		Pararhodospirillum oryzae	Proteobacteria	557	2	Cap4	0	
906	WP_114836301.1		Agrobacterium tumefaciens	Proteobacteria	558	1	Cap4	1	1876
907	WP_141683738.1		Rhizobium sp. AC27/96	Proteobacteria	559	2	Cap4	1	1884
908	OCJ05377.1		Rhizobium sp. AC27/96	Proteobacteria	559	2	Cap4	0	
909	WP_126514219.1		Pseudomonas taetrolens	Proteobacteria	560	3	Cap4	1	1877
910	SED18013.1		Pseudomonas taetrolens	Proteobacteria	560	3	Cap4	0	
911	KMM82728.1		Pseudomonas taetrolens	Proteobacteria	560	3	Cap4	0	

912	WP_162254864.1		Pseudomonas sp. Root401	Proteobacteria	561	3	Cap4	1	1871
913	KQW40919.1		Pseudomonas sp. Root401	Proteobacteria	561	3	Cap4	0	
914	WP_112196689.1		Pseudomonas sp. LG1E9	Proteobacteria	561	3	Cap4	0	
915	PIW60826.1		Shewanella sp. CG12_big_fil_rev_8_21_14_0_65_47_15	Proteobacteria	562	1	Cap4	1	1875
916	WP_087465078.1		Cellvibrio sp. PSBB006	Proteobacteria	563	1	Cap4	1	1877
917	WP_133680890.1		Cellvibrio sp. YR554	Proteobacteria	564	1	Cap4	1	1878
918	WP_141117927.1		Leisingera sp. JC1	Proteobacteria	565	2	Cap4_small ORF	1	1858
919	OBY26275.1		Leisingera sp. JC1	Proteobacteria	565	2	Cap4_small ORF	0	
920	WP_159809666.1		Litoreibacter roseus	Proteobacteria	566	1	Cap4_small ORF	1	1861
921	WP_149534186.1		Tardiphaga sp. P9-11	Proteobacteria	567	1	Cap4_small ORF	1	1875
922	RYD50763.1		Sphingobacteriales bacterium	Bacteroidetes	568	1	Cap4_small ORF	1	1861
923	WP_065376949.1		Ensifer adhaerens	Proteobacteria	569	2	Cap4_small ORF	1	1879
924	WP_104668461.1		Ensifer adhaerens	Proteobacteria	569	2	Cap4_small ORF	0	
925	RWP08244.1		Mesorhizobium sp.	Proteobacteria	570	1	Cap4_small ORF	1	1884
926	TAK50650.1		Betaproteobacteria bacterium	Proteobacteria	571	1	Cap4_small ORF	1	1877
927	WP_140521376.1		unclassified Mesorhizobium	Proteobacteria	572	1	Cap4_small ORF	1	1878
928	RUU06849.1		Mesorhizobium sp. USDA-HM6	Proteobacteria	573	1	Cap4_small ORF	1	1873
929	MBU0651650.1		Proteobacteria bacterium	Proteobacteria	574	1	Cap4_small ORF	1	1890
930	MBI5436204.1		Nitrosomonadales bacterium	Proteobacteria	575	1	Cap4_small ORF	1	1886
931	ANA40318.1		Geobacter anodireducens	Proteobacteria	576	1	Cap4_small ORF	1	1893
932	WP_136815170.1		Desulfopila sp. IMCC35006	Proteobacteria	577	1	Cap4_small ORF	1	1879
933	MBI2353660.1		Deltaproteobacteria bacterium	Proteobacteria	578	1	Cap4_small ORF	1	1879
934	WP_158819896.1		Granulicella sp. S156	Acidobacteria	579	1	Cap4_small ORF	1	1858

935	WP_165590611.1		Ralstonia solanacearum	Proteobacteria	580	1	Cap4_small ORF	1	1863
936	WP_188678147.1		Amylibacter cionae	Proteobacteria	581	2	Cap4_small ORF	1	1867
937	GGA29925.1		Amylibacter cionae	Proteobacteria	581	2	Cap4_small ORF	0	
938	NEY91989.1		Rhodobacteraceae bacterium KMS-5	Proteobacteria	582	1	Cap4_small ORF	1	1863
939	WP_152009079.1		Hoeflea sp. EC-HK425	Proteobacteria	583	1	Cap4_small ORF	1	1864
940	WP_158702908.1		Melaminivora sp. SC2-9	Proteobacteria	584	2	Cap4_small ORF	1	1869
941	AVO50642.1		Melaminivora sp. SC2-9	Proteobacteria	584	2	Cap4_small ORF	0	
942	OSQ39091.1		Thalassospira sp. MCCC 1A01428	Proteobacteria	585	2	Cap4_small ORF	1	1855
943	WP_139835700.1		Thalassospira sp. MCCC 1A01428	Proteobacteria	585	2	Cap4_small ORF	0	
944	WP_196608579.1		Cupidesulfovibrio oxamicus	Proteobacteria	586	1	Cap4_small ORF	1	1859
945	WP_047409283.1		Caulobacter sp. OV484	Proteobacteria	587	1	Cap4_small ORF	1	1866
946	MBI5446726.1		Deltaproteobacteria bacterium	Proteobacteria	588	1	Cap4_small ORF	1	1863
947	MBU0525851.1		Gammaproteobacteria bacterium	Proteobacteria	589	2	Cap4_small ORF	1	1862
948	MBU0837113.1		Gammaproteobacteria bacterium	Proteobacteria	589	2	Cap4_small ORF	0	
949	WP_213671535.1		Roseococcus sp. XZZS9	Proteobacteria	590	1	Cap4_small ORF	1	1851
950	WP_209568777.1		Sphingomonas sp. PvP018	Proteobacteria	591	1	Cap4_small ORF	1	1855
951	WP_009108300.1		Desulfovibrio sp. U5L	Proteobacteria	592	1	Cap4_small ORF	1	1856
952	WP_048847964.1		Acetobacter indonesiensis	Proteobacteria	593	1	Cap4_small ORF	1	1853
953	WP_184002727.1		Stakelama sediminis	Proteobacteria	594	1	Cap4_small ORF	1	1858
954	WP_175359773.1		Pseudomonas brassicacearum	Proteobacteria	595	2	Cap4_small ORF	1	1852
955	WP_093530318.1		Pseudomonas sp. NFPP24	Proteobacteria	595	2	Cap4_small ORF	0	
956	WP_123344118.1		Pseudomonas brassicacearum	Proteobacteria	596	6	Cap4_small ORF	1	1852

957	WP_104502918.1		Pseudomonas orientalis	Proteobacteria	596	6	Cap4_small ORF	0	
958	WP_134786373.1		Pseudomonas sp. S150	Proteobacteria	596	6	Cap4_small ORF	0	
959	WP_154905925.1		Pseudomonas fluorescens	Proteobacteria	596	6	Cap4_small ORF	0	
960	WP_096138975.1		Pseudomonas congelans	Proteobacteria	596	6	Cap4_small ORF	0	
961	WP_045185202.1		Pseudomonas sp. 5	Proteobacteria	596	6	Cap4_small ORF	0	
962	WP_147407002.1		Pseudomonas sp. 91RF	Proteobacteria	597	2	Cap4_small ORF	1	1852
963	RIJ09707.1		Pseudomonas sp. 91RF	Proteobacteria	597	2	Cap4_small ORF	0	
964	MBT9505730.1		Rhodoferax sp.	Proteobacteria	598	1	Cap4_small ORF	1	1872
965	WP_034351439.1		Comamonas thiooxydans	Proteobacteria	599	2	Cap4_small ORF	1	1896
966	KGG96075.1		Comamonas thiooxydans	Proteobacteria	599	2	Cap4_small ORF	0	
967	WP_069047813.1		Hydrogenophaga sp. RAC07	Proteobacteria	600	3	Cap4_small ORF	1	1898
968	MBA4211654.1		Polaromonas sp.	Proteobacteria	600	3	Cap4_small ORF	0	
969	AOF85118.1		Hydrogenophaga sp. RAC07	Proteobacteria	600	3	Cap4_small ORF	0	
970	MAD35383.1		Tistrella sp.	Proteobacteria	601	1	Cap4_small ORF	1	1885
971	WP_160770219.1		Alteraurantiacibacter buctensis	Proteobacteria	602	1	Cap4_small ORF	1	1885
972	PCJ42297.1		Alphaproteobacteria bacterium	Proteobacteria	603	1	Cap4_small ORF	1	1867
973	WP_075256153.1		Herbaspirillum sp. WT00C	Proteobacteria	604	2	Cap4_small ORF	1	1860
974	WP_180669400.1		unclassified Herbaspirillum	Proteobacteria	604	2	Cap4_small ORF	0	
975	WP_156521438.1		Burkholderia sp. MSMB1589WGS	Proteobacteria	605	1	Cap4_small ORF	1	1864
976	WP_175891404.1		Burkholderia cepacia	Proteobacteria	606	1	Cap4_small ORF	1	1863
977	HCA56555.1		Blastocatellia bacterium	Acidobacteria	607	1	Cap4_small ORF	1	1891
978	WP_108034678.1		unclassified Methylobacterium	Proteobacteria	608	2	Cap4_small ORF	1	1871

979	PTW93903.1		Methylobacterium sp. GV104	Proteobacteria	608	2	Cap4_small ORF	0	
980	WP_152031703.1		Longilinea arvoryzae	Chloroflexi	609	1	Cap4_small ORF	1	1880
981	WP_110520295.1		Herpetosiphon llansteffanensis	Chloroflexi	610	1	Cap4_small ORF	1	1877
982	WP_130663683.1		Rhizobium leguminosarum	Proteobacteria	611	1	Cap4_small ORF	1	1871
983	WP_174157561.1		Rhizobium lusitanum	Proteobacteria	612	1	Cap4_small ORF	1	1875
984	WP_108402672.1		Limnohabitans curvus	Proteobacteria	613	1	Cap4_small ORF	1	1874
985	WP_000801635.1		Vibrio cholerae	Proteobacteria	614	5	Cap4_small ORF	1	1881
986	WP_032478953.1		Vibrio cholerae	Proteobacteria	614	5	Cap4_small ORF	0	
987	WP_119781002.1		Vibrio cholerae	Proteobacteria	614	5	Cap4_small ORF	0	
988	WP_142626524.1		Vibrio cholerae	Proteobacteria	614	5	Cap4_small ORF	0	
989	HBC3855484.1		Vibrio parahaemolyticus	Proteobacteria	614	5	Cap4_small ORF	0	
990	WP_160970693.1		Pseudomaricurvus sp. HS19	Proteobacteria	615	1	Cap4_small ORF	1	1881
991	WP_168451132.1		Spongiibacter sp. KMU-166	Proteobacteria	616	1	Cap4_small ORF	1	1880
992	WP_010322421.1		Marinobacterium stanieri	Proteobacteria	617	1	Cap4_small ORF	1	1880
993	WP_199470127.1		Marinomonas sp. C1424	Proteobacteria	618	1	Cap4_small ORF	1	1881
994	WP_111589538.1		Aeromonas salmonicida	Proteobacteria	619	3	Cap4_small ORF	1	1881
995	WP_203762646.1		Aeromonas caviae	Proteobacteria	619	3	Cap4_small ORF	0	
996	WP_197930878.1		Aeromonas allosaccharophila	Proteobacteria	619	3	Cap4_small ORF	0	
997	WP_171734027.1		Vibrio sp. 99-70-13A1	Proteobacteria	620	2	Cap4_small ORF	1	1881
998	WP_210470699.1		Vibrio crassostreae	Proteobacteria	620	2	Cap4_small ORF	0	
999	WP_076652841.1		unclassified Vibrio	Proteobacteria	621	1	Cap4_small ORF	1	1880
1000	WP_171388284.1		Vibrio crassostreae	Proteobacteria	622	1	Cap4_small ORF	1	1880
1001	WP_068439999.1		Providencia heimbachae	Proteobacteria	623	5	Cap4_small ORF	1	1880
1002	WP_198641393.1		Providencia rettgeri	Proteobacteria	623	5	Cap4_small ORF	0	

1003	WP_163767657.1		Providencia stuartii	Proteobacteria	623	5	Cap4_small ORF	0	
1004	WP_210849090.1		Proteus mirabilis	Proteobacteria	623	5	Cap4_small ORF	0	
1005	WP_068908482.1		Providencia heimbachae	Proteobacteria	623	5	Cap4_small ORF	0	
1006	WP_047101804.1		Vibrio alginolyticus	Proteobacteria	624	3	Cap4_small ORF	1	1881
1007	EGQ7973182.1		Vibrio parahaemolyticus	Proteobacteria	624	3	Cap4_small ORF	0	
1008	MBM5091637.1		Vibrio parahaemolyticus	Proteobacteria	624	3	Cap4_small ORF	0	
1009	WP_011263536.1		Aliivibrio fischeri	Proteobacteria	625	4	Cap4_small ORF	1	1881
1010	WP_063646023.1		Aliivibrio fischeri	Proteobacteria	625	4	Cap4_small ORF	0	
1011	WP_140271718.1		Vibrio parahaemolyticus	Proteobacteria	625	4	Cap4_small ORF	0	
1012	HAS6673266.1		Vibrio parahaemolyticus	Proteobacteria	625	4	Cap4_small ORF	0	
1013	MBF4433411.1		Vibrio anguillarum	Proteobacteria	626	1	Cap4_small ORF	1	1878
1014	HAS8388902.1		Vibrio vulnificus	Proteobacteria	627	1	Cap4_small ORF	1	1881
1015	WP_145238727.1		Pseudoalteromonas neustonica	Proteobacteria	628	1	Cap4_small ORF	1	1887
1016	MAE01560.1		Pseudoalteromonas sp.	Proteobacteria	629	1	Cap4_small ORF	1	1888
1017	WP_149613980.1		Pseudoalteromonas fuliginea	Proteobacteria	630	2	Cap4_small ORF	1	1887
1018	WP_182755703.1		Pseudoalteromonas sp. SR44-2	Proteobacteria	630	2	Cap4_small ORF	0	
See also Data S6 of Gao et al., Science 377, eabm4096 (2022), which is incorporated herein by reference as if expressed in its entirety herein.									

Table 6. Avs4 homologs and domain architectures.

#	Accession	Name	Organism	Phylum	Tree Order	*A*	*B*	*C*	*D*	*E*	*F*
1	WP_188400997.1		Fictibacillus barbaricus	Firmicutes	1	1	Mrr	1			1370
2	WP_098149241.1		Bacillus pseudomycoides	Firmicutes	2	1	Mrr	1			1367
3	WP_098321775.1		Bacillus cereus group	Firmicutes	3	1	Mrr	1			1367
4	PFT25286.1		Bacillus thuringiensis	Firmicutes	4	1	Mrr	1			1368
5	WP_046664035.1		Bacillaceae	Firmicutes	5	2	Mrr	1			1369
6	WP_213379628.1		Bacillus subtilis	Firmicutes	5	2	Mrr	0			
7	WP_014115823.1		Bacillus spizizenii	Firmicutes	6	1	Mrr	1			1369
8	WP_052169443.1		unclassified Exiguobacterium	Firmicutes	7	2	Mrr	1			1371
9	AOT01664.1		Exiguobacterium sp. U13-1	Firmicutes	7	2	Mrr	0			
10	WP_047795013.1		Exiguobacterium sp. JLM-2	Firmicutes	8	1	Mrr	1			1368
11	WP_113970935.1		Bacillus aquimaris	Firmicutes	9	1	Mrr	1			1372
12	WP_066391044.1		Neobacillus mesonae	Firmicutes	10	1	Mrr	1			1371
13	WP_078817948.1		Sporosarcina newyorkensis	Firmicutes	11	1	Mrr	1			1372
14	WP_204190304.1		Mammaliicoccus sciuri	Firmicutes	12	1	Mrr	1			1377
15	WP_200406860.1		Lactococcus taiwanensis	Firmicutes	13	1	Mrr	1			1371
16	EOD94087.1		Enterococcus faecalis EnGen0059	Firmicutes	14	4	Mrr	1			1372
17	WP_129236555.1		Enterococcus faecalis	Firmicutes	14	4	Mrr	0			
18	WP_108355536.1		Enterococcus faecalis	Firmicutes	14	4	Mrr	0			
19	WP_025191754.1		Enterococcus faecalis	Firmicutes	14	4	Mrr	0			
20	WP_098637867.1		Bacillus toyonensis	Firmicutes	15	1	Mrr	1			1372
21	WP_143415941.1		Geobacillus sp. E263	Firmicutes	16	2	Mrr	1			1374
22	WP_151706212.1		Bacillus sp. JC-7	Firmicutes	16	2	Mrr	0			
23	WP_078501671.1		Paenibacillus selenitireducens	Firmicutes	17	2	Mrr	1			1383
24	OPA74758.1		Paenibacillus selenitireducens	Firmicutes	17	2	Mrr	0			
25	SCY73416.1		Paenibacillus polysaccharolyticus	Firmicutes	18	2	Mrr	1			1378
26	WP_090920527.1		Paenibacillus polysaccharolyticus	Firmicutes	18	2	Mrr	0			
27	MBI5459740.1		Methanobacterium sp.	Euryarchaeota	19	1	Mrr	1			1396
28	KJS46398.1		Peptococcaceae bacterium BRH_c23	Firmicutes	20	2	Mrr	1			1374
29	HBW35373.1		Desulfosporosinus sp.	Firmicutes	20	2	Mrr	0			
30	WP_200901045.1		Thermincola ferriacetica	Firmicutes	21	1	Mrr	1			1377

31	WP_015325489.1		Methanomethylovorans hollandica	Euryarchaeota	22	1	Mrr	1			1376
32	MBF0405297.1		Nitrospirae bacterium	Nitrospirae	23	1	Mrr	1			1369
33	HAR34929.1		Desulfobacter sp.	Proteobacteria	24	1	Mrr	1			1376
34	WP_028301103.1		Oceanospirillum beijerinckii	Proteobacteria	25	2	Mrr	1			1376
35	WP_015487192.1		Thalassolituus oleivorans	Proteobacteria	25	2	Mrr	0			
36	WP_001548013.1		Enterobacterales	Proteobacteria	26	1	Mrr	1			1375
37	EGI4487100.1		Escherichia coli	Proteobacteria	27	17	Mrr	1			1376
38	WP_180552803.1		Cronobacter sakazakii	Proteobacteria	27	17	Mrr	0			
39	EAR1298370.1		Salmonella enterica	Proteobacteria	27	17	Mrr	0			
40	EHJ9221241.1		Salmonella enterica	Proteobacteria	27	17	Mrr	0			
41	EAN4380156.1		Salmonella enterica	Proteobacteria	27	17	Mrr	0			
42	WP_200014774.1		Citrobacter freundii	Proteobacteria	27	17	Mrr	0			
43	WP_107183325.1		Escherichia coli	Proteobacteria	27	17	Mrr	0			
44	WP_063660718.1		Enterobacteriaceae	Proteobacteria	27	17	Mrr	0			
45	HBC0557824.1		Citrobacter freundii	Proteobacteria	27	17	Mrr	0			
46	WP_110826314.1		Escherichia coli	Proteobacteria	27	17	Mrr	0			
47	WP_181629359.1		Citrobacter sp. RHBSTW-01013	Proteobacteria	27	17	Mrr	0			
48	EFL4459666.1		Escherichia coli	Proteobacteria	27	17	Mrr	0			
49	EFN4843203.1		Escherichia coli	Proteobacteria	27	17	Mrr	0			
50	WP_074561656.1		Proteus mirabilis	Proteobacteria	27	17	Mrr	0			
51	WP_108717236.1		Proteus mirabilis	Proteobacteria	27	17	Mrr	0			
52	WP_048219194.1		Enterobacter sp. BIDMC93	Proteobacteria	27	17	Mrr	0			
53	ECO9822345.1		Salmonella enterica	Proteobacteria	27	17	Mrr	0			
54	WP_044068927.1	EcAvs4	Escherichia coli	Proteobacteria	28	114	Mrr	1			1376
55	WP_199946422.1		Citrobacter sp. FDAARGOS_156	Proteobacteria	28	114	Mrr	0			
56	EFH5285965.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
57	EFO0859672.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
58	WP_033734906.1		Pantoea vagans	Proteobacteria	28	114	Mrr	0			
59	EHN8834177.1		Enterobacter hormaechei	Proteobacteria	28	114	Mrr	0			
60	WP_172695350.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
61	WP_021571021.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
62	WP_117035078.1		Klebsiella pneumoniae	Proteobacteria	28	114	Mrr	0			
63	WP_176452971.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			

64	WP_033544383.1		Enterobacteriaceae	Proteobacteria	28	114	Mrr	0			
65	WP_094029139.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
66	WP_049044714.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
67	WP_054622895.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
68	WP_094319745.1		Enterobacteriaceae	Proteobacteria	28	114	Mrr	0			
69	WP_001552206.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
70	WP_112032778.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
71	WP_208483203.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
72	EGO4029856.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
73	WP_104774795.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
74	EFH8083720.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
75	WP_163436353.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
76	WP_200984368.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
77	EEW1994578.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
78	HBC7633164.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
79	HAW6939794.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
80	WP_187797006.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
81	EFP4144916.1		Shigella flexneri	Proteobacteria	28	114	Mrr	0			
82	HAI5003889.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
83	EHP7220878.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
84	HBA4112287.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
85	HBE4809472.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
86	WP_136691559.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
87	HAZ7427615.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
88	WP_042029849.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
89	WP_059328861.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
90	EFK6762968.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
91	WP_104719827.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
92	HAW1228248.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
93	WP_061362265.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
94	WP_124064767.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
95	WP_105279299.1		Escherichia	Proteobacteria	28	114	Mrr	0			
96	HAL3721365.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
97	WP_202793005.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
98	WP_097334928.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			

99	NAF97958.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
100	WP_001572194.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
101	WP_130723354.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
102	HAW3733422.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
103	HAW1903184.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
104	EFH5419041.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
105	WP_000240574.1		Enterobacteriaceae	Proteobacteria	28	114	Mrr	0			
106	HAL7437996.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
107	WP_201519025.1		Escherichia ruysiae	Proteobacteria	28	114	Mrr	0			
108	EFO4396035.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
109	WP_054540921.1		Enterobacterales	Proteobacteria	28	114	Mrr	0			
110	HAY1549976.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
111	WP_099147840.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
112	HAO2598317.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
113	HAY2509485.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
114	WP_096911637.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
115	WP_087534746.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
116	HAV9903515.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
117	HAJ3500309.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
118	HAN4503974.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
119	WP_064232198.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
120	EFD9830649.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
121	HAN5138863.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
122	WP_129964656.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
123	WP_001697684.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
124	NAH15491.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
125	WP_113848840.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
126	WP_106466489.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
127	EGM0635455.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
128	WP_137517133.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
129	WP_075278014.1		Pectobacterium brasiliense	Proteobacteria	28	114	Mrr	0			
130	WP_194967140.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
131	HBE4389818.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
132	EFE8969780.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
133	WP_153677274.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			

134	WP_171995315.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
135	HAU979137.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
136	WP_114429508.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
137	EFJ8098256.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
138	WP_178282190.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
139	MBS8710648.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
140	WP_000240573.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
141	WP_137565047.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
142	WP_181211447.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
143	WP_096924963.1		Klebsiella pneumoniae	Proteobacteria	28	114	Mrr	0			
144	EGS5097797.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
145	WP_119719661.1		Klebsiella pneumoniae	Proteobacteria	28	114	Mrr	0			
146	WP_208064910.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
147	WP_062914738.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
148	HAM9465060.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
149	HAI2686192.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
150	EBJ3544487.1		Salmonella enterica	Proteobacteria	28	114	Mrr	0			
151	HAJ7922833.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
152	EFN4021456.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
153	WP_105453333.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
154	EEY3885972.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
155	WP_154123502.1		Enterobacter hormaechei	Proteobacteria	28	114	Mrr	0			
156	MBH3047561.1		Serratia marcescens	Proteobacteria	28	114	Mrr	0			
157	HAL1199981.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
158	WP_109954997.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
159	WP_028127231.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
160	WP_115201953.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
161	HBB1771698.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
162	WP_139589827.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
163	EAC1583121.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
164	EHA3733173.1		Citrobacter freundii	Proteobacteria	28	114	Mrr	0			
165	AVU37279.1		Serratia marcescens	Proteobacteria	28	114	Mrr	0			
166	HAL2560926.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			
167	WP_195806943.1		Escherichia coli	Proteobacteria	28	114	Mrr	0			

168	NRH21331.1		Candidatus Gracilibacteria bacterium	Unclassified	29	1	Mrr	1			1390
169	OFX26993.1		Bacteroidetes bacterium GWA2_31_9	Bacteroidetes	30	1	Mrr	1			1386
170	RPJ79756.1		Alphaproteobacteria bacterium	Proteobacteria	31	2	Mrr	1			1390
171	RPJ73014.1		Alphaproteobacteria bacterium	Proteobacteria	31	2	Mrr	0			
172	MBI5408231.1		Nitrospirae bacterium	Nitrospirae	32	1	Mrr	1			1387
173	OGV34651.1		Lentisphaerae bacterium GWF2_45_14	Lentisphaerae	33	1	Mrr	1			1385
174	KKT80763.1		Parcubacteria group bacterium GW2011_GWF2_44_8b	Unclassified	34	2	Mrr	1			1411
175	OHA60551.1		Candidatus Vogelbacteria bacterium RIFOXYD1_FULL_51_18	Unclassified	34	2	Mrr	0			
176	NOX64602.1		Chlorobi bacterium	Chlorobi	35	1	Mrr	1			1363
177	NPA51760.1		Aquificae bacterium	Aquificae	36	1	Mrr	1			1372
178	WP_114838292.1		Halarcobacter bivalviorum	Proteobacteria	37	1	Mrr	1			1377
179	KAF3982703.1		Methylococcales symbiont of Hymedesmia sp. n. MRB-2018	Proteobacteria	38	1	Mrr	1			1388
180	WP_207688654.1		Desulfonema limicola	Proteobacteria	39	1	Mrr	1			1379
181	WP_100788505.1		Leptospira adleri	Spirochaetes	40	1	Mrr	1			1381
182	WP_111570772.1		Halanaerobium saccharolyticum	Firmicutes	41	2	Mrr	1			1389
183	WP_149796740.1		Halanaerobium congolense	Firmicutes	41	2	Mrr	0			
184	WP_100542212.1		Lysinibacillus xylanilyticus	Firmicutes	42	1	Mrr	1			1389
185	MBN1054909.1		Clostridium botulinum	Firmicutes	43	1	Mrr	1			1387
186	WP_077830518.1		Clostridium beijerinckii	Firmicutes	44	2	Mrr	1			1391
187	WP_171786355.1		Clostridium beijerinckii	Firmicutes	44	2	Mrr	0			
188	WP_099716613.1		Clostridium sp. LS	Firmicutes	45	2	Mrr	1			1391
189	EKQ56698.1		Clostridium sp. Maddingley MBC34-26	Firmicutes	45	2	Mrr	0			
190	MBT3176925.1		Desulfobacula sp.	Proteobacteria	46	1	Mrr	1			1372
191	HAH62478.1		Treponema sp.	Spirochaetes	47	1	Mrr	1			1381
192	MBR4791219.1		Treponema sp.	Spirochaetes	48	1	Mrr	1			1365
193	MBR5401687.1		Treponema sp.	Spirochaetes	49	1	Mrr	1			1365
194	CUV65504.1		Sulfurovum sp. enrichment culture clone C5	Proteobacteria	50	1	Mrr	1			1375
195	WP_200415481.1		Arcobacter sp. FWKO B	Proteobacteria	51	1	Mrr	1			1369
196	WP_083191842.1		[Arcobacter] porcinus	Proteobacteria	52	3	Mrr	1			1371

197	WP_083202012.1		Aliarcobacter thereius	Proteobacteria	52	3	Mrr	0			
198	WP_083190897.1		Aliarcobacter thereius	Proteobacteria	52	3	Mrr	0			
199	MBP7749014.1		Aliarcobacter sp.	Proteobacteria	53	1	Transmembrane	1			1366
200	MBE0492665.1		Sulfurospirillum sp.	Proteobacteria	54	1	Transmembrane	1			1368
201	WP_059369241.1		Treponema endosymbiont of Eucomonympha sp.	Spirochaetes	55	1	Mrr	1			1389
202	WP_133822409.1		Rahnella sp. BIGb0236	Proteobacteria	56	1	Mrr	1			1381
203	WP_109937829.1		Enterobacter sp. HPCN14	Proteobacteria	57	1	Mrr	1			1380
204	EDW27933183.1		Salmonella enterica	Proteobacteria	58	1	Mrr	1			1380
205	MBQ0780900.1		Thalassolituus oleivorans	Proteobacteria	59	1	Mrr	1			1378
206	TXK94859.1		Methylococcaceae bacterium CS5	Proteobacteria	60	2	Mrr	1			1343
207	TXK94841.1		Methylococcaceae bacterium CS4	Proteobacteria	60	2	Mrr	0			
208	WP_156801525.1		Psychromonas sp. CNPT3	Proteobacteria	61	2	Mrr	1			1374
209	AGH819301		Psychromonas sp. CNPT3	Proteobacteria	61	2	Mrr	0			
210	WP_046019608.1		Marinomonas sp. S3726	Proteobacteria	62	1	Mrr	1			1375
211	WP_144224124.1		Shewanella algae	Proteobacteria	63	2	Mrr	1			1375
212	WP_144135094.1		Shewanella algae	Proteobacteria	63	2	Mrr	0			
213	WP_107274799.1		Photobacterium iliopiscarium	Proteobacteria	64	1	Mrr	1			1375
214	WP_069438825.1		Pseudoalteromonas tetraodonis	Proteobacteria	65	1	Mrr	1			1375
215	PCJ24636.1		SAR86 cluster bacterium	Proteobacteria	66	1	Mrr	1			1378
216	WP_204665432.1		Fusibacter tunisiensis	Firmicutes	67	1	Mrr	1			1357
217	WP_147732835.1		Listeria innocua	Firmicutes	68	1	Mrr	1			1372
218	WP_132704185.1		Reinekea marinisedimentorum	Proteobacteria	69	1	Mrr	1			1380
219	MBP6367029.1		Nitrosomonas sp.	Proteobacteria	70	1	Mrr	1			1375
220	WP_201981043.1		Aeromonas veronii	Proteobacteria	71	1	Mrr	1			1380
221	WP_208638775.1		Vibrio rhodolitus	Proteobacteria	72	1	Mrr	1			1379
222	EGR1137689.1		Vibrio cholerae	Proteobacteria	73	6	Mrr	1			1379
223	WP_186437140.1		Vibrio atlanticus	Proteobacteria	73	6	Mrr	0			
224	TVU58084.1		Vibrio atlanticus	Proteobacteria	73	6	Mrr	0			
225	EGQ8443239.1		Vibrio cholerae	Proteobacteria	73	6	Mrr	0			
226	WP_185833761.1		Vibrio cholerae	Proteobacteria	73	6	Mrr	0			
227	WP_130087294.1		Aliivibrio finisterrensis	Proteobacteria	73	6	Mrr	0			
228	WP_192928639.1		Photobacterium alginatilyticum	Proteobacteria	74	5	Mrr	1			1380

229	WP_208892996.1		Vibrio anguillarum	Proteobacteria	74	5	Mrr	0			
230	WP_196298308.1		Vibrio campbellii	Proteobacteria	74	5	Mrr	0			
231	AQP37667.1		Vibrio anguillarum	Proteobacteria	74	5	Mrr	0			
232	NBI54572.1		Photobacterium alginatilyticum	Proteobacteria	74	5	Mrr	0			
233	WP_181897503.1		Grimontia hollisae	Proteobacteria	75	1	Mrr	1			1379
234	WP_176776477.1		Nitrosomonas oligotropha	Proteobacteria	76	1	Mrr	1			1380
235	WP_004657592.1		Acinetobacter proteolyticus	Proteobacteria	77	1	Mrr	1			1380
236	WP_171293664.1		Acinetobacter haemolyticus	Proteobacteria	78	1	Mrr	1			1380
237	WP_004581102.1		Marinobacter nanhaiticus	Proteobacteria	79	1	Mrr	1			1380
238	WP_011370258.1		Hydrogenovibrio crunogenus	Proteobacteria	80	1	Mrr	1			1380
239	NHI95667.1		Proteus mirabilis	Proteobacteria	81	5	Mrr	1			1380
240	WP_201090200.1		Entomomonas sp. F2A	Proteobacteria	81	5	Mrr	0			
241	WP_207779893.1		Proteus mirabilis	Proteobacteria	81	5	Mrr	0			
242	WP_185171857.1		Proteus mirabilis	Proteobacteria	81	5	Mrr	0			
243	HBC6051646.1		Proteus mirabilis	Proteobacteria	81	5	Mrr	0			
244	ECN6815935.1		Salmonella enterica subsp. enterica serovar Typhimurium	Proteobacteria	82	34	Mrr	1			1380
245	WP_187265346.1		Escherichia coli	Proteobacteria	82	34	Mrr	0			
246	TXQ11293.1		Escherichia coli	Proteobacteria	82	34	Mrr	0			
247	WP_203067935.1		Salmonella enterica	Proteobacteria	82	34	Mrr	0			
248	EFD4033533.1		Escherichia coli	Proteobacteria	82	34	Mrr	0			
249	HAD4838318.1		Salmonella enterica subsp. enterica serovar Typhi str. CT18	Proteobacteria	82	34	Mrr	0			
250	EFM4226080.1		Escherichia coli	Proteobacteria	82	34	Mrr	0			
251	WP_172837556.1		Escherichia coli	Proteobacteria	82	34	Mrr	0			
252	WP_016245977.1		Escherichia coli	Proteobacteria	82	34	Mrr	0			
253	WP_210115209.1		Escherichia coli	Proteobacteria	82	34	Mrr	0			
254	WP_172903590.1		Escherichia coli	Proteobacteria	82	34	Mrr	0			
255	WP_210434251.1		Escherichia coli	Proteobacteria	82	34	Mrr	0			
256	WP_187439486.1		Escherichia coli	Proteobacteria	82	34	Mrr	0			
257	MBF2854060.1		Escherichia coli	Proteobacteria	82	34	Mrr	0			
258	HAW8037787.1		Escherichia coli	Proteobacteria	82	34	Mrr	0			
259	EFI6347447.1		Escherichia coli	Proteobacteria	82	34	Mrr	0			
260	EFR7065199.1		Escherichia coli	Proteobacteria	82	34	Mrr	0			
261	EEW5978513.1	Ec2Avs4	Escherichia coli	Proteobacteria	82	34	Mrr	0			

262	EFI567636 3.1		Escherichia coli	Proteoba cteria	82	34	Mrr	0			
263	EGF53056 91.1		Escherichia coli	Proteoba cteria	82	34	Mrr	0			
264	MZI03168. 1		Escherichia coli	Proteoba cteria	82	34	Mrr	0			
265	EFW6870 634.1		Shigella sonnei	Proteoba cteria	82	34	Mrr	0			
266	EEZ76543 38.1		Escherichia coli	Proteoba cteria	82	34	Mrr	0			
267	WP_01623 9529.1		Escherichia coli	Proteoba cteria	82	34	Mrr	0			
268	TGW4102 4.1		Escherichia coli	Proteoba cteria	82	34	Mrr	0			
269	EFE08878 13.1		Escherichia coli	Proteoba cteria	82	34	Mrr	0			
270	EFK34822 12.1		Escherichia coli	Proteoba cteria	82	34	Mrr	0			
271	EFI297851 8.1		Escherichia coli	Proteoba cteria	82	34	Mrr	0			
272	WP_20408 2033.1		Escherichia coli	Proteoba cteria	82	34	Mrr	0			
273	EFG57095 19.1		Escherichia coli	Proteoba cteria	82	34	Mrr	0			
274	HAO5558 442.1		Escherichia coli	Proteoba cteria	82	34	Mrr	0			
275	EGJ51791 08.1		Escherichia coli	Proteoba cteria	82	34	Mrr	0			
276	EEY56908 65.1		Escherichia coli	Proteoba cteria	82	34	Mrr	0			
277	WP_17421 7567.1		Escherichia coli	Proteoba cteria	82	34	Mrr	0			
278	WP_06107 7503.1		Citrobacter amalonaticus	Proteoba cteria	83	1	Mrr	1			1381
279	WP_06392 6238.1		Enterobacter hormaechei	Proteoba cteria	84	10	Mrr	1			1382
280	VAE07880 .1		Enterobacter hormaechei	Proteoba cteria	84	10	Mrr	0			
281	WP_20281 5543.1		Enterobacter hormaechei	Proteoba cteria	84	10	Mrr	0			
282	CZW0580 9.1		Enterobacter hormaechei	Proteoba cteria	84	10	Mrr	0			
283	WP_20281 7076.1		Enterobacter hormaechei	Proteoba cteria	84	10	Mrr	0			
284	SAG45166 .1		Enterobacter hormaechei	Proteoba cteria	84	10	Mrr	0			
285	SAC93766 .1		Enterobacter hormaechei	Proteoba cteria	84	10	Mrr	0			
286	WP_20281 5497.1		Enterobacter hormaechei	Proteoba cteria	84	10	Mrr	0			
287	WP_15032 0268.1		Enterobacter hormaechei	Proteoba cteria	84	10	Mrr	0			
288	WP_06321 5922.1		Enterobacter sp. DSM 30060	Proteoba cteria	84	10	Mrr	0			
289	MBK0712 707.1		Klebsiella aerogenes	Proteoba cteria	85	3	Mrr	1			1383
290	WP_20089 1392.1		Klebsiella aerogenes	Proteoba cteria	85	3	Mrr	0			
291	KJP12292. 1		[Enterobacter] aerogenes	Proteoba cteria	85	3	Mrr	0			
292	WP_14200 2429.1		Serratia marcescens	Proteoba cteria	86	3	Mrr	1			1383
293	MBH3205 598.1		Serratia marcescens	Proteoba cteria	86	3	Mrr	0			
294	WP_19782 3328.1		Serratia ureilytica	Proteoba cteria	86	3	Mrr	0			
295	WP_19470 1566.1		Fusibacter sp. Q10-2	Firmicute s	87	1	Mrr	1			1368
296	WP_15186 1669.1		Alkaliphilus pronyensis	Firmicute s	88	1	Mrr	1			1365

297	WP_101539862.1		Anaerococcus octavius	Firmicutes	89	1	Mrr	1			1374
298	WP_072911239.1		Desulfotomaculum aeronauticum	Firmicutes	90	1	Mrr	1			1380
299	MBA1336236.1		Firmicutes bacterium	Firmicutes	91	1	Mrr	1			1394
300	TDJ09138.1		Deltaproteobacteria bacterium	Proteobacteria	92	1	Mrr	1			1390
301	TRZ48792.1		bacterium	Unclassified	93	1	Mrr	1			1384
302	MBI4706514.1		Candidatus Omnitrophica bacterium	Candidatus Omnitrophica	94	1	Mrr	1			1389
303	PKP60153.1		Candidatus Atribacteria bacterium HGW-Atribacteria-1	Atribacterota	95	1	Mrr	1			1385
304	HBY57712.1		Candidatus Atribacteria bacterium	Atribacterota	96	1	Mrr	1			1387
305	OGN97055.1		Chloroflexi bacterium RBG_13_51_36	Chloroflexi	97	1	Mrr	1			1385
306	TRZ51984.1		Dehalococcoidia bacterium	Chloroflexi	98	1	Mrr	1			1393
307	MBL0240403.1		Chloracidobacterium sp.	Acidobacteria	99	1	Mrr	1			1386
308	KKS93412.1		Candidatus Collierbacteria bacterium GW2011_GWC2_43_12	Unclassified	100	1	Mrr	1			1394
309	OGY51351.1		Candidatus Buchananbacteria bacterium RIFCSPLOWO2_01_FULL_39_33	Unclassified	101	1	Mrr	1			1414
310	HAT1982218.1		Legionella pneumophila	Proteobacteria	102	1	Mrr	1			1401
311	WP_058468133.1		Fluoribacter gormanii	Proteobacteria	103	1	Mrr	1			1399
312	WP_070322492.1		Legionella pneumophila	Proteobacteria	104	3	Mrr	1			1399
313	WP_061398562.1		Legionella pneumophila	Proteobacteria	104	3	Mrr	0			
314	HAT1881200.1		Legionella pneumophila	Proteobacteria	104	3	Mrr	0			
315	KOY12443.1		Bradyrhizobium diazoefficiens	Proteobacteria	105	4	Mrr	1			1415
316	KGJ65563.1		Bradyrhizobium diazoefficiens SEMIA 5080	Proteobacteria	105	4	Mrr	0			
317	WP_182869608.1		Bradyrhizobium diazoefficiens	Proteobacteria	105	4	Mrr	0			
318	WP_131234196.1		Bradyrhizobium diazoefficiens	Proteobacteria	105	4	Mrr	0			
319	WP_087877677.1		Bradyrhizobium sp. ORS 285	Proteobacteria	106	2	Mrr	1			1414
320	CCD90273.1		Bradyrhizobium sp. ORS 285	Proteobacteria	106	2	Mrr	0			
321	WP_053389900.1		Citrobacter freundii complex	Proteobacteria	107	1	Mrr	1			1417
322	WP_145962016.1		Sulfuriferula thiophila	Proteobacteria	108	1	Mrr	1			1415
323	RTL55636.1		Rhodocyclaceae bacterium	Proteobacteria	109	1	Mrr	1			1417

324	OGS66405.1		Gallionellales bacterium GWA2_54_124	Proteobacteria	110	1	Mrr	1			1416
325	WP_111376293.1		Flavobacterium nitrogenifigens	Bacteroidetes	111	1	Mrr	1			1375
326	WP_210645956.1		Flavobacterium sp. P4023	Bacteroidetes	112	1	Mrr	1			1384
327	WP_093024089.1		Pustulibacterium marinum	Bacteroidetes	113	1	Mrr	1			1376
328	WP_163385394.1		Cyclobacterium jeungdonense	Bacteroidetes	114	1	Mrr	1			1381
329	WP_203458146.1		Lacinutrix sp. WUR7	Bacteroidetes	115	1	Mrr	1			1390
330	WP_154917081.1		Maribacter flavus	Bacteroidetes	116	1	Mrr	1			1392
331	KAF5425055.1		ANME-2 cluster archaeon	Euryarchaeota	117	1	Mrr	1			1386
332	WP_131607490.1		Pedobacter sp. RP-3-8	Bacteroidetes	118	1	Mrr	1			1388
333	WP_112374127.1		Sphingobacterium multivorum	Bacteroidetes	119	1	Mrr	1			1388
334	MBC7888490.1		Ferruginibacter sp.	Bacteroidetes	120	1	Mrr	1			1381
335	OYZ01010.1		Sphingobacteriia bacterium 28-36-52	Bacteroidetes	121	1	Mrr	1			1381
336	MBP6589559.1		Chitinophagaceae bacterium	Bacteroidetes	122	1	Mrr	1			1378
337	MBN8720078.1		Sediminibacterium magnilacihabitans	Bacteroidetes	123	1	Mrr	1			1378
338	BBL62648.1		Methanobrevibacter arboriphilus	Euryarchaeota	124	1	Mrr	1			1418
339	MBR6993696.1		Methanobrevibacter sp.	Euryarchaeota	125	1	Mrr	1			1365
340	WP_042693958.1		Methanobrevibacter oralis	Euryarchaeota	126	1	Mrr	1			1364
341	MBR2236329.1		Prevotella sp.	Bacteroidetes	127	1	Mrr	1			1290
342	WP_182021347.1		Bacteroides intestinalis	Bacteroidetes	128	2	Mrr	1			1330
343	RGV57265.1		Bacteroides intestinalis	Bacteroidetes	128	2	Mrr	0			
344	MBD5183176.1		Bacteroidales bacterium	Bacteroidetes	129	2	Mrr	1			1359
345	MBD5372204.1		Bacteroides sp.	Bacteroidetes	129	2	Mrr	0			
346	WP_036490665.1		Neisseria mucosa	Proteobacteria	130	1	Mrr	1			1369
347	WP_045444928.1		Psychrobacter sp. JCM 18902	Proteobacteria	131	2	Mrr	1			1372
348	WP_201497530.1		Psychrobacter arenosus	Proteobacteria	131	2	Mrr	0			
349	WP_201595075.1		Psychrobacter fulvigenes	Proteobacteria	132	1	Mrr	1			1376
350	WP_073552104.1		Elizabethkingia meningoseptica	Bacteroidetes	133	1	Mrr	1			1372
351	WP_062676173.1		Chryseobacterium cucumeris	Bacteroidetes	134	2	Mrr	1			1371
352	WP_065080903.1		Elizabethkingia miricola	Bacteroidetes	134	2	Mrr	0			
353	WP_163399676.1		Flavobacterium fluviatile	Bacteroidetes	135	1	Mrr	1			1374
354	WP_019541419.1		Proteiniphilum acetatigenes	Bacteroidetes	136	1	Mrr	1			1376
355	WP_103475132.1		Arenibacter catalasegens	Bacteroidetes	137	1	Mrr	1			1381
356	WP_085501037.1		Arenibacter troitsensis	Bacteroidetes	138	1	Mrr	1			1377

357	WP_204347091.1		Flavihalobacter algicola	Bacteroidetes	139	1	Mrr	1			1371
358	WP_179335829.1		Winogradskyella costae	Bacteroidetes	140	1	Mrr	1			1371
359	MBP1631271.1		Bacteroidetes bacterium	Bacteroidetes	141	1	Mrr	1			1381
360	WP_006263341.1		Myroides odoratimimus	Bacteroidetes	142	1	Mrr	1			1374
361	WP_013666869.1		Sphingobacterium sp. 21	Bacteroidetes	143	1	Mrr	1			1373
362	NQU67036.1		Candidatus Marinimicrobia bacterium	Candidatus Marinimicrobia	144	1	Mrr	1			1374
363	WP_047842108.1		Paenibacillus sp. VT-400	Firmicutes	145	1	Mrr	1			1374
364	WP_109220090.1		Ignatzschineria cameli	Proteobacteria	146	1	Mrr	1			1380
365	WP_204434487.1		Bacteroides mediterraneensis	Bacteroidetes	147	1	Mrr	1			1369
366	MBO5382468.1		Bacteroides sp.	Bacteroidetes	148	1	Mrr	1			1369
367	MBO5853287.1		Bacteroidales bacterium	Bacteroidetes	149	1	Mrr	1			1369
368	MBP8945795.1		Paludibacteraceae bacterium	Bacteroidetes	150	1	Mrr	1			1372
369	WP_055301401.1		Bacteroides thetaiotaomicron	Bacteroidetes	151	1	Mrr	1			1373
370	NLN49918.1		Clostridiales bacterium	Firmicutes	152	1	Mrr	1			1372
371	MBP8960272.1		Bacteroidales bacterium	Bacteroidetes	153	1	Mrr	1			1373
372	MBR9776730.1		Cytophagales bacterium	Bacteroidetes	154	1	Mrr	1			1372
373	WP_024480421.1		Cellulophaga baltica	Bacteroidetes	155	1	Mrr	1			1372
374	WP_208889374.1		Polaribacter sejongensis	Bacteroidetes	156	1	Mrr	1			1372
375	WP_007096343.1		Kordia algicida	Bacteroidetes	157	1	Mrr	1			1372
376	WP_025838174.1		Prevotella scopos	Bacteroidetes	158	4	Mrr	1			1372
377	MBF1479172.1		Prevotella pallens	Bacteroidetes	158	4	Mrr	0			
378	WP_008564061.1		Prevotella maculosa	Bacteroidetes	158	4	Mrr	0			
379	WP_088401358.1		Prevotella nigrescens	Bacteroidetes	158	4	Mrr	0			
380	KAB2858924.1		Flavobacteriales bacterium	Bacteroidetes	159	1	Mrr	1			1372
381	WP_072885472.1		Chryseobacterium takakiae	Bacteroidetes	160	1	Mrr	1			1364
382	MBK9413844.1		Bacteroidetes bacterium	Bacteroidetes	161	1	Mrr	1			1372
383	WP_094876672.1		Paenibacillus sp. VTT E-133291	Firmicutes	162	1	Mrr	1			1373
384	WP_047684511.1		Paenibacillus chondroitinus	Firmicutes	163	1	Mrr	1			1372
385	WP_144705222.1		Cohnella sp. G13	Firmicutes	164	1	Mrr	1			1372
386	WP_211142838.1		Clostridium aciditolerans	Firmicutes	165	1	MBL + serine protease	1			1363
387	EHI0762614.1		Clostridium botulinum	Firmicutes	166	1	MBL + serine protease	1			1362
388	WP_056999717.1		Carnobacterium maltaromaticum	Firmicutes	167	4	Mrr	1			1381
389	WP_141299568.1		Carnobacterium maltaromaticum	Firmicutes	167	4	Mrr	0			

390	WP_157455011.1		Carnobacterium maltaromaticum	Firmicutes	167	4	Mrr	0			
391	WP_069698087.1		Enterococcus rivorum	Firmicutes	167	4	Mrr	0			
392	WP_132416040.1		Paenibacillus sp. 18JY21-1	Firmicutes	168	1	Mrr	1			1386
393	WP_054402668.1		Paenibacillus solani	Firmicutes	169	1	Mrr	1			1384
394	WP_063480133.1		Paenibacillus glucanolyticus	Firmicutes	170	1	Mrr	1			1384
395	WP_098606232.1		Bacillus thuringiensis	Firmicutes	171	1	Mrr	1			1366
396	WP_014469645.1		Bacillus subtilis group	Firmicutes	172	5	Mrr	1			1367
397	WP_174201878.1		Bacillus amyloliquefaciens	Firmicutes	172	5	Mrr	0			
398	WP_013350856.1		Bacillus amyloliquefaciens	Firmicutes	172	5	Mrr	0			
399	WP_061573685.1		Bacillus amyloliquefaciens	Firmicutes	172	5	Mrr	0			
400	WP_065118603.1		Bacillus amyloliquefaciens	Firmicutes	172	5	Mrr	0			
401	WP_097959852.1		Bacillus toyonensis	Firmicutes	173	1	Mrr	1			1377
402	PED57384.1		Bacillus anthracis	Firmicutes	174	1	Mrr	1			1377
403	NBH13559.1		Lachnospiraceae bacterium	Firmicutes	175	1	Mrr	1			1395
404	WP_118725504.1		Clostridium sp. AM42-4	Firmicutes	176	2	Mrr	1			1402
405	HCW29438.1		Roseburia sp.	Firmicutes	176	2	Mrr	0			
406	WP_147354930.1		Clostridium sp. AF20-7	Firmicutes	177	15	Mrr	1			1387
407	RHT76589.1		Clostridium sp. AM28-20LB	Firmicutes	177	15	Mrr	0			
408	WP_182448395.1		Clostridium sp. OM05-5BH	Firmicutes	177	15	Mrr	0			
409	WP_182436630.1		Clostridium sp. OF13-4	Firmicutes	177	15	Mrr	0			
410	RHP40436.1		Clostridium sp. AF32-7AC	Firmicutes	177	15	Mrr	0			
411	RHR03838.1		Clostridium sp. AF20-7	Firmicutes	177	15	Mrr	0			
412	WP_182430894.1		Clostridium sp. AF50-3	Firmicutes	177	15	Mrr	0			
413	RHO87798.1		Clostridium sp. AF37-7	Firmicutes	177	15	Mrr	0			
414	MBD9191217.1		Roseburia inulinivorans	Firmicutes	177	15	Mrr	0			
415	WP_182431250.1		unclassified Clostridium	Firmicutes	177	15	Mrr	0			
416	WP_182457030.1		Clostridium sp. TF11-13AC	Firmicutes	177	15	Mrr	0			
417	WP_147353586.1		unclassified Clostridium	Firmicutes	177	15	Mrr	0			
418	WP_147355497.1		unclassified Clostridium	Firmicutes	177	15	Mrr	0			
419	WP_182434832.1		Clostridium sp. AM48-13	Firmicutes	177	15	Mrr	0			
420	WP_147353109.1		Clostridium sp. AF32-7AC	Firmicutes	177	15	Mrr	0			
421	WP_117858929.1		Ruminococcus sp. AF34-12	Firmicutes	178	2	Mrr	1			1369
422	MBS6408364.1		Ruminococcus bicirculans	Firmicutes	178	2	Mrr	0			

423	MBR5683229.1		Ruminococcus sp.	Firmicutes	179	1	Mrr	1			1369
424	WP_087325348.1		Faecalibacterium sp. An58	Firmicutes	180	2	Mrr	1			1372
425	WP_087382848.1		Faecalibacterium sp. An122	Firmicutes	180	2	Mrr	0			
426	WP_136890703.1		Dysosmobacter welbionis	Firmicutes	181	2	Mrr	1			1373
427	WP_119311100.1		Oscillibacter sp. PEA192	Firmicutes	181	2	Mrr	0			
428	WP_195567575.1		Ruminococcus sp. 1001262B_160229_C12	Firmicutes	182	2	Mrr	1			1379
429	SCH64811.1		uncultured Blautia sp.	Firmicutes	182	2	Mrr	0			
430	WP_173775414.1		Blautia massiliensis	Firmicutes	183	1	Mrr	1			1387
431	WP_118514389.1		Ruminococcus sp. OF02-6	Firmicutes	184	1	Mrr	1			1384
432	HBB47045.1		Blautia sp.	Firmicutes	185	1	Mrr	1			1387
433	MBL6516838.1		Prevotella stercorea	Bacteroidetes	186	2	Mrr	1			1323
434	MBP5422524.1		Paludibacteraceae bacterium	Bacteroidetes	186	2	Mrr	0			
435	MBO4818313.1		Bacteroidales bacterium	Bacteroidetes	187	1	Mrr	1			1337
436	MBR6438090.1		Bacteroidales bacterium	Bacteroidetes	188	1	Mrr	1			1329
437	MBR6863352.1		Bacteroidales bacterium	Bacteroidetes	189	1	Mrr	1			1328
438	MBQ0088612.1		Prevotellaceae bacterium	Bacteroidetes	190	1	Mrr	1			1328
439	MBO7573824.1		Bacilli bacterium	Firmicutes	191	1	Putative REase	1			1310
440	PWL90250.1		Lachnospiraceae bacterium	Firmicutes	192	1	Putative REase	1			1370
441	ETP71505.1		Lachnospiraceae bacterium JC7	Firmicutes	193	1	Putative REase	1			1370
442	MBR4781994.1		Lachnospiraceae bacterium	Firmicutes	194	1	Putative REase	1			1371
443	SDJ24977.1		Lachnospiraceae bacterium G41	Firmicutes	195	1	Putative REase	1			1371
444	MBQ2675050.1		Prevotella sp.	Bacteroidetes	196	1	Putative REase	1			1373
445	MBO6184678.1		Chryseobacterium sp.	Bacteroidetes	197	1	Putative REase	1			1389
446	WP_090174697.1		Eubacterium oxidoreducens	Firmicutes	198	1	Putative REase	1			1374
447	MBP3817002.1		Butyrivibrio sp.	Firmicutes	199	1	Putative REase	1			1374
448	WP_105086413.1		Veillonella sp. T14073-2	Firmicutes	200	1	Putative REase	1			1378
449	MBN2919304.1		Lactobacillus sp.	Firmicutes	201	1	Putative REase	1			1377
450	WP_191520016.1		Lachnoclostridium sp. An138	Firmicutes	202	1	Putative REase	1			1379
451	WP_076778148.1		Lachnoclostridium phocaeense	Firmicutes	203	1	Putative REase	1			1377
452	MBS5430617.1		Lachnospiraceae bacterium	Firmicutes	204	1	Putative REase	1			1377
453	WP_032835136.1		Fusobacterium sp. OBRC1	Fusobacteria	205	1	Putative REase	1			1378
454	WP_029494352.1		Fusobacterium hwasookii	Fusobacteria	206	1	Putative REase	1			1376
455	WP_023039844.1		Fusobacterium	Fusobacteria	207	1	Putative REase	1			1378
456	MBQ9946711.1		Clostridia bacterium	Firmicutes	208	1	Putative REase	1			1370

457	MBO5938728.1		Clostridia bacterium	Firmicutes	209	1	Putative REase	1			1380
458	MBQ6784117.1		Acholeplasmatales bacterium	Tenericutes	210	1	Putative REase	1			1356
459	WP_092753397.1		Hydrogenoanaerobacterium saccharovorans	Firmicutes	211	1	Putative REase	1			1379
460	PKM73707.1		Firmicutes bacterium HGW-Firmicutes-16	Firmicutes	212	1	Putative REase	1			1355
461	NMB97794.1		Clostridiaceae bacterium	Firmicutes	213	1	Putative REase	1			1379
462	WP_027630689.1		Ruminiclostridium cellobioparum	Firmicutes	214	1	Putative REase	1			1383
463	PKM60671.1		Firmicutes bacterium HGW-Firmicutes-4	Firmicutes	215	1	Putative REase	1			1385
464	MBI4857758.1		Acetobacterium woodii	Firmicutes	216	1	Putative REase	1			1385
465	WP_160687102.1		Clostridium sp. C2-6-12	Firmicutes	217	1	Putative REase	1			1386
466	WP_171787670.1		Clostridium beijerinckii	Firmicutes	218	1	Putative REase	1			1385
467	WP_198609934.1		Clostridium perfringens	Firmicutes	219	2	Putative REase	1			1385
468	HAT4289380.1		Clostridium perfringens	Firmicutes	219	2	Putative REase	0			
469	NFL95079.1		Clostridium botulinum	Firmicutes	220	1	Putative REase	1			1385
470	WP_072832181.1		Clostridium collagenovorans	Firmicutes	221	1	Putative REase	1			1385
471	MBS5854345.1		Clostridium sp.	Firmicutes	222	1	Putative REase	1			1365
472	CCY28872.1		Acholeplasma sp. CAG:878	Tenericutes	223	1	Putative REase	1			1374
473	MBR3815720.1		Clostridia bacterium	Firmicutes	224	1	Putative REase	1			1375
474	MBD5516813.1		Lachnospiraceae bacterium	Firmicutes	225	1	Putative REase	1			1379
475	PWM76107.1		Clostridium sp.	Firmicutes	226	1	Putative REase	1			1384
476	WP_204966775.1		Megamonas hypermegale	Firmicutes	227	1	Putative REase	1			1384
477	SCJ25939.1		uncultured Ruminococcus sp.	Firmicutes	228	13	Putative REase	1			1384
478	WP_144363247.1		Lacrimispora amygdalina	Firmicutes	228	13	Putative REase	0			
479	RGF58817.1		Coprobacillus sp. AF36-10BH	Firmicutes	228	13	Putative REase	0			
480	WP_144367465.1		[Ruminococcus] torques	Firmicutes	228	13	Putative REase	0			
481	WP_118173411.1		Roseburia sp. AF20-18LB	Firmicutes	228	13	Putative REase	0			
482	SCI25423.1		uncultured Clostridium sp.	Firmicutes	228	13	Putative REase	0			
483	WP_173904527.1		Blautia obeum	Firmicutes	228	13	Putative REase	0			
484	WP_173738478.1		Blautia schinkii	Firmicutes	228	13	Putative REase	0			
485	MBD9215629.1		Solobacterium sp.	Firmicutes	228	13	Putative REase	0			
486	WP_117639783.1		Blautia obeum	Firmicutes	228	13	Putative REase	0			
487	WP_015528557.1		[Ruminococcus] torques	Firmicutes	228	13	Putative REase	0			
488	WP_118529548.1		Clostridium sp. OF03-18AA	Firmicutes	228	13	Putative REase	0			

489	MBS7001644.1		Clostridiaceae bacterium	Firmicutes	228	13	Putative REase	0			
490	WP_195531255.1		Fusicatenibacter saccharivorans	Firmicutes	229	3	Unknown-4	1			1387
491	SCJ55194.1		uncultured Blautia sp.	Firmicutes	229	3	Unknown-4	0			
492	WP_117823697.1		Coprococcus comes	Firmicutes	229	3	Unknown-4	0			
493	WP_127065448.1		Cellulosilyticum sp. WCF-2	Firmicutes	230	1	Unknown-4	1			1387
494	MBR6125001.1		Candidatus Saccharibacteria bacterium	Unclassified	231	1	Unknown-4	1			1384
495	WP_088631271.1		Fibrobacter sp. UWB3	Fibrobacteres	232	1	Unknown-4	1			1342
496	WP_118356855.1		Holdemanella biformis	Firmicutes	233	1	Unknown-4	1			1380
497	RHO10759.1		Lachnospiraceae bacterium AM21-21	Firmicutes	234	1	Unknown-4	1			1377
498	WP_170824303.1		Hungatella xylanolytica	Firmicutes	235	1	Unknown-4	1			1378
499	RHS80631.1		Firmicutes bacterium AM43-11BH	Firmicutes	236	1	Unknown-4	1			1378
500	QUO23259.1		Clostridiaceae bacterium Marseille-Q4143	Firmicutes	237	1	Unknown-4	1			1378
501	MBP3326682.1		Coprococcus sp.	Firmicutes	238	1	Unknown-4	1			1378
502	WP_173819357.1		Dorea longicatena	Firmicutes	239	3	Unknown-4	1			1378
503	WP_173792470.1		Dorea longicatena	Firmicutes	239	3	Unknown-4	0			
504	RGT13838.1		[Eubacterium] rectale	Firmicutes	239	3	Unknown-4	0			
505	WP_055058083.1		Blautia obeum	Firmicutes	240	1	Unknown-4	1			1378
506	WP_055196036.1		Roseburia intestinalis	Firmicutes	241	4	Unknown-4	1			1377
507	SCJ14605.1		uncultured Eubacterium sp.	Firmicutes	241	4	Unknown-4	0			
508	MBS5466335.1		Clostridium sp.	Firmicutes	241	4	Unknown-4	0			
509	WP_118605197.1		Clostridium sp. AF20-17LB	Firmicutes	241	4	Unknown-4	0			
510	WP_118513783.1		Clostridium sp. AF37-5AT	Firmicutes	242	2	Unknown-4	1			1378
511	MBS7000681.1		Clostridiaceae bacterium	Firmicutes	242	2	Unknown-4	0			
512	WP_143375102.1		Flavobacterium sp. ZT4R6	Bacteroidetes	243	1	Cap4	1			1329
513	NOQ23732.1		Bacteroidales bacterium	Bacteroidetes	244	1	Cap4	1			1377
514	MBO6585324.1		Gracilimonas sp.	Balneolaeota	245	1	MBL + serine protease	1			1371
515	MBN1876728.1		Anaerolineae bacterium	Chloroflexi	246	1	MBL + serine protease	1			1384
516	WP_042371347.1		Bacteroides neonati	Bacteroidetes	247	1	MBL + serine protease	1			1325
517	MBO7570615.1		Bacteroidales bacterium	Bacteroidetes	248	1	MBL + serine protease	1			1272
518	MBP3245773.1		Bacteroidaceae bacterium	Bacteroidetes	249	1	MBL + serine protease	1			1275

519	WP_147287661.1		Prevotella disiens	Bacteroidetes	250	2	MBL + serine protease	1			1274
520	ERJ80282.1		Prevotella disiens JCM 6334 = ATCC 29426	Bacteroidetes	250	2	MBL + serine protease	0			
521	WP_195453668.1		Alistipes communis	Bacteroidetes	251	1	MBL + serine protease	1			1274
522	MBE6333903.1		Bacteroidales bacterium	Bacteroidetes	252	1	MBL + serine protease	1			1274
523	WP_198138464.1		Phocaeicola dorei	Bacteroidetes	253	3	MBL + serine protease	1			1276
524	WP_087388247.1		Phocaeicola dorei	Bacteroidetes	253	3	MBL + serine protease	0			
525	EEO46972.2		Bacteroides dorei 5_1_36/D4	Bacteroidetes	253	3	MBL + serine protease	0			
526	WP_089903773.1		Chitinophaga arvensicola	Bacteroidetes	254	1	MBL + serine protease	1			1443
527	WP_091149868.1		Mucilaginibacter pineti	Bacteroidetes	255	1	MBL + serine protease	1			1441
528	WP_091133995.1		Flavobacterium anhuiense	Bacteroidetes	256	1	MBL + serine protease	1			1447
529	WP_123205908.1		Pedobacter jejuensis	Bacteroidetes	257	2	MBL + serine protease	1			1436
530	RNL52103.1		Pedobacter jejuensis	Bacteroidetes	257	2	MBL + serine protease	0			
531	MPT33625.1		Flavobacterium sp.	Bacteroidetes	258	1	MBL + serine protease	1			1447
532	WP_008463815.1		Flavobacterium	Bacteroidetes	259	1	MBL + serine protease	1			1447
533	WP_041515743.1		Flavobacterium hibernum	Bacteroidetes	260	2	MBL + serine protease	1			1447
534	WP_129053287.1		Flavobacterium sp. YO64	Bacteroidetes	260	2	MBL + serine protease	0			
535	WP_180857642.1		Flavobacterium sp. PXU-55	Bacteroidetes	261	1	MBL + serine protease	1			1450
536	WP_127340193.1		Flavobacterium cupreum	Bacteroidetes	262	1	MBL + serine protease	1			1450
537	WP_044236412.1		Haliscomenobacter hydrossis	Bacteroidetes	263	2	MBL + serine protease	1			1374
538	AEE54389.1		Haliscomenobacter hydrossis DSM 1100	Bacteroidetes	263	2	MBL + serine protease	0			
539	WP_182961599.1		Pedobacter gandavensis	Bacteroidetes	264	1	MBL + serine protease	1			1375
540	MBN8857080.1		Sphingobacteriales bacterium	Bacteroidetes	265	1	MBL + serine protease	1			1395
541	RZK27179.1		Flavobacterium sp.	Bacteroidetes	266	1	MBL + serine protease	1			1387

542	WP_188094992.1		Chitinophaga varians	Bacteroidetes	267	1	MBL + serine protease	1			1380
543	WP_202703809.1		Flavobacterium sp. UGB4466	Bacteroidetes	268	1	MBL + serine protease	1			1365
544	WP_126577040.1		Flavobacterium sp. ZB4P23	Bacteroidetes	269	1	MBL + serine protease	1			1364
545	WP_074491249.1		Mucilaginibacter polytrichastri	Bacteroidetes	270	1	MBL + serine protease	1			1373
546	RZJ76287.1		Flavobacterium sp.	Bacteroidetes	271	1	MBL + serine protease	1			1367
547	WP_090106762.1		Chitinophaga sp. CF118	Bacteroidetes	272	1	MBL + serine protease	1			1368
548	WP_088584160.1		Chryseobacterium indologenes	Bacteroidetes	273	1	MBL + serine protease	1			1364
549	WP_061084765.1		Chryseobacterium indologenes	Bacteroidetes	274	5	MBL + serine protease	1			1363
550	WP_073060373.1		Chryseobacterium oranimense	Bacteroidetes	274	5	MBL + serine protease	0			
551	WP_048506386.1		Chryseobacterium	Bacteroidetes	274	5	MBL + serine protease	0			
552	NPA10255.1		Chlorobi bacterium	Chlorobi	274	5	MBL + serine protease	0			
553	OJZ01186.1		Sphingobacterium sp. 40-24	Bacteroidetes	274	5	MBL + serine protease	0			
554	WP_120230645.1		Chryseobacterium sp. AG363	Bacteroidetes	275	2	MBL + serine protease	1			1364
555	WP_214588874.1		Chryseobacterium sp. ZHDP1	Bacteroidetes	275	2	MBL + serine protease	0			
556	WP_090001576.1		Chryseobacterium taichungense	Bacteroidetes	276	1	MBL + serine protease	1			1359
557	WP_002977363.1		Chryseobacterium	Bacteroidetes	277	2	MBL + serine protease	1			1362
558	WP_027383733.1		Epilithonimonas caeni	Bacteroidetes	277	2	MBL + serine protease	0			
559	WP_157757781.1		Elizabethkingia miricola	Bacteroidetes	278	2	MBL + serine protease	1			1361
560	ATL42240.1		Elizabethkingia miricola	Bacteroidetes	278	2	MBL + serine protease	0			
561	WP_133440843.1		Chryseobacterium sp. NBC 122	Bacteroidetes	279	1	MBL + serine protease	1			1360
562	WP_149698577.1		Chitinophaga sp. CF418	Bacteroidetes	280	1	Cap4	1			1397
563	WP_109698083.1		Chitinophaga deserti	Bacteroidetes	281	1	Cap4	1			1401
564	WP_129131514.1		Lacibacter luteus	Bacteroidetes	282	1	Cap4	1			1392
565	WP_124562402.1		Pedobacter sp. KBW01	Bacteroidetes	283	1	Cap4	1			1387
566	WP_190995483.1		Sphingobacterium sp. DN00404	Bacteroidetes	284	1	Cap4	1			1398

567	WP_146270153.1		Mucilaginibacter sp. MJ1a	Bacteroidetes	285	1	Cap4	1			1406
568	WP_133968896.1		Algibacter lectus	Bacteroidetes	286	1	Cap4	1			1387
569	WP_157802055.1		Flavobacterium sp. 1	Bacteroidetes	287	2	Unknown-3 (5' split)	1			1374
570	PJJ08719.1		Flavobacterium sp. 1	Bacteroidetes	287	2	Unknown-3 (5' split)	0			
571	WP_163399271.1		Flavobacterium fluviatile	Bacteroidetes	288	1	Unknown-3 (5' split)	1			1384
572	WP_136401327.1		Flavobacterium sp. CC-CTC003	Bacteroidetes	289	1	Unknown-3 (5' split)	1			1386
573	WP_027377678.1		Chryseobacterium palustre	Bacteroidetes	290	1	Unknown-3 (5' split)	1			1382
574	WP_191177693.1		Chryseobacterium sp. C-2	Bacteroidetes	291	1	Unknown-3 (5' split)	1			1381
575	WP_056084182.1		Chryseobacterium sp. Leaf394	Bacteroidetes	292	1	Unknown-3 (5' split)	1			1383
576	WP_076352044.1		Chryseobacterium joostei	Bacteroidetes	293	1	Unknown-3 (5' split)	1			1383
577	WP_125440527.1		Hymenobacter perfusus	Bacteroidetes	294	2	Cap4	1			1384
578	WP_190786132.1		Hymenobacter sp. BT646	Bacteroidetes	294	2	Cap4	0			
579	WP_015333105.1		Fibrella aestuarina	Bacteroidetes	295	1	Cap4	1			1395
580	WP_119670501.1		Fibrisoma montanum	Bacteroidetes	296	1	Cap4	1			1394
581	WP_179021813.1		Winogradskyella forsetii	Bacteroidetes	297	1	Cap4	1			1392
582	WP_093406079.1		Salegentibacter flavus	Bacteroidetes	298	1	Cap4	1			1391
583	WP_150313502.1		Salegentibacter sp. R32	Bacteroidetes	299	1	Cap4	1			1390
584	MBC31815.1		Muricauda sp.	Bacteroidetes	300	1	Cap4	1			1391
585	WP_026716886.1		Flavobacterium gelidilacus	Bacteroidetes	301	1	Cap4	1			1391
586	WP_092132465.1		Bizionia paragorgiae	Bacteroidetes	302	1	Cap4	1			1391
587	MBG43686.1		Aequorivita sp.	Bacteroidetes	303	1	Cap4	1			1389
588	WP_068825324.1		Wenyingzhuangia fucanilytica	Bacteroidetes	304	1	Cap4	1			1389
589	WP_067151511.1		Tamlana agarivorans	Bacteroidetes	305	1	Cap4	1			1390
590	WP_076698114.1		Seonamhaeicola sp. S2-3	Bacteroidetes	306	1	Cap4	1			1389
591	WP_147676324.1		Algibacter pacificus	Bacteroidetes	307	1	Cap4	1			1389
592	WP_179414837.1		Mucilaginibacter sp. E4BP6	Bacteroidetes	308	1	Cap4	1			1382
593	MAM33787.1		Micavibrio sp.	Proteobacteria	309	1	Cap4	1			1390
594	WP_100376785.1		Chryseobacterium geocarposphaerae	Bacteroidetes	310	1	Cap4	1			1396
595	MBL7839210.1		Cyclobacteriaceae bacterium	Bacteroidetes	311	1	Cap4	1			1384
596	WP_035074212.1		Anditalea andensis	Bacteroidetes	312	1	Cap4	1			1387

597	OJU74608.1		Bacteroidetes bacterium 47-18	Bacteroidetes	313	1	Cap4	1			1387
598	MBQ6083086.1		Bacteroidales bacterium	Bacteroidetes	314	1	Cap4	1			1388
599	MBL0316986.1		Flavobacteriales bacterium	Bacteroidetes	315	1	Cap4	1			1389
600	WP_119350757.1		Mariniphaga sediminis	Bacteroidetes	316	1	Cap4	1			1389
601	MBN2236291.1		Bacteroidales bacterium	Bacteroidetes	317	1	Cap4	1			1388
602	WP_013996828.1		Capnocytophaga canimorsus	Bacteroidetes	318	1	Cap4	1			1389
603	WP_009415262.1		Capnocytophaga sp. oral taxon 332	Bacteroidetes	319	2	Cap4	1			1384
604	WP_126677638.1		Prevotella koreensis	Bacteroidetes	319	2	Cap4	0			
605	WP_185249583.1		Chryseobacterium bernardetii	Bacteroidetes	320	1	Cap4	1			1396
606	WP_135656677.1		Leptospira vanthielii	Spirochaetes	321	1	Cap4	1			1397
607	WP_119430758.1		Pontibacter oryzae	Bacteroidetes	322	1	Cap4	1			1396
608	WP_208223689.1		Salegentibacter sp. BDJ18	Bacteroidetes	323	1	Cap4	1			1394
609	WP_050696331.1		Prevotella fusca	Bacteroidetes	324	1	Cap4	1			1396
610	WP_095899148.1		Capnocytophaga sputigena	Bacteroidetes	325	1	Cap4	1			1391
611	WP_183696835.1		Alloprevotella rava	Bacteroidetes	326	7	Cap4	1			1391
612	WP_061868008.1		Prevotella intermedia	Bacteroidetes	326	7	Cap4	0			
613	WP_097626537.1		Porphyromonas gingivalis	Bacteroidetes	326	7	Cap4	0			
614	WP_100189852.1		Prevotella intermedia	Bacteroidetes	326	7	Cap4	0			
615	WP_099895843.1		Porphyromonas gingivalis	Bacteroidetes	326	7	Cap4	0			
616	WP_102189115.1		Prevotella timonensis	Bacteroidetes	326	7	Cap4	0			
617	WP_130830081.1		Prevotella sp. Marseille-P4334	Bacteroidetes	326	7	Cap4	0			
618	WP_019207274.1		Phocaeicola abscessus	Bacteroidetes	327	1	Cap4	1			1392
619	OGO82885.1		Clostridiales bacterium GWD2_32_19	Firmicutes	328	1	Cap4	1			1381
620	MBF8275273.1		Candidatus Brocadiaceae bacterium	Planctomycetes	329	1	Cap4	1			1418
621	WP_048046049.1		Methanosarcina mazei	Euryarchaeota	330	2	Cap4	1			1416
622	AAM32737.1		Methanosarcina mazei Go1	Euryarchaeota	330	2	Cap4	0			
623	WP_048128321.1		Methanosarcina sp. WH1	Euryarchaeota	331	1	Cap4	1			1415
624	HHY38871.1		Clostridia bacterium	Firmicutes	332	1	Cap4	1			1410
625	WP_031463739.1		Paenibacillus polymyxa	Firmicutes	333	1	Cap4	1			1416
626	WP_183587924.1		Paenibacillus rhizosphaerae	Firmicutes	334	1	Cap4	1			1412
627	ACK70730.1		Gloeothece citriformis PCC 7424	Cyanobacteria	335	1	Short	1			1408
628	TRT98870.1		Microcystis flos-aquae Ma_QC_C_20070823_S18D	Cyanobacteria	336	1	Short	1			1408

629	NCQ99115.1		Microcystis aeruginosa L211-11	Cyanobacteria	337	1	Short	1			1402
630	TRV10478.1		Microcystis wesenbergii Mw_MB_S_20031200_S109	Cyanobacteria	338	1	Short	1			1406
631	RWX46269.1		Candidatus Electrothrix aarhusiensis	Proteobacteria	339	1	Cap4	1			1392
632	OQW91929.1		Thiotrichaceae bacterium IS1	Proteobacteria	340	1	Cap4	1			1407
633	WP_015159328.1		Chamaesiphon minutus	Cyanobacteria	341	1	Cap4	1			1413
634	WP_193974117.1		Microcoleus sp. LEGE 07076	Cyanobacteria	342	1	Cap4	1			1405
635	WP_193977905.1		Phormidium sp. LEGE 05292	Cyanobacteria	343	1	Cap4	1			1410
636	WP_190569523.1		Alkalinema sp. FACHB-956	Cyanobacteria	344	1	Cap4	1			1464
637	MBE9102622.1		filamentous cyanobacterium LEGE 07170	Cyanobacteria	345	1	Cap4	1			1458
638	WP_086289671.1		Gilliamella apicola	Proteobacteria	346	1	Cap4	1			1408
639	WP_087775949.1	KpAvs4	Klebsiella pneumoniae	Proteobacteria	347	10	Cap4	1			1403
640	WP_117119679.1		Klebsiella pneumoniae	Proteobacteria	347	10	Cap4	0			
641	EBE0342374.1		Salmonella enterica	Proteobacteria	347	10	Cap4	0			
642	EAB7347849.1		Salmonella enterica subsp. enterica	Proteobacteria	347	10	Cap4	0			
643	EAS2170465.1		Salmonella enterica	Proteobacteria	347	10	Cap4	0			
644	EBO0188881.1		Salmonella enterica	Proteobacteria	347	10	Cap4	0			
645	EAM4411786.1		Salmonella enterica	Proteobacteria	347	10	Cap4	0			
646	EBW4678453.1		Salmonella enterica subsp. salamae serovar Sofia	Proteobacteria	347	10	Cap4	0			
647	WP_135727253.1		Klebsiella pneumoniae	Proteobacteria	347	10	Cap4	0			
648	WP_155109577.1		Escherichia alba	Proteobacteria	347	10	Cap4	0			
649	WP_190286914.1		Citrobacter koseri	Proteobacteria	348	1	Cap4	1			1401
650	WP_054877742.1		Yersinia wautersii	Proteobacteria	349	2	Cap4	1			1401
651	WP_042820288.1		Yersinia wautersii	Proteobacteria	349	2	Cap4	0			
652	WP_053119406.1		Pseudomonas thivervalensis	Proteobacteria	350	1	Cap4	1			1390
653	WP_126570271.1		Pseudomonas sp. C 49-2	Proteobacteria	351	1	Cap4	1			1389
654	WP_007938302.1		Pseudomonas sp. GM18	Proteobacteria	352	1	Cap4	1			1390
655	WP_019649735.1		Pseudomonas sp. 45MFCol3.1	Proteobacteria	353	1	Cap4	1			1409
656	WP_123449902.1		Pseudomonas fluorescens	Proteobacteria	354	1	Cap4	1			1409
657	WP_080636242.1		Pseudomonas sp. ATCC 13867	Proteobacteria	355	1	Cap4	1			1403
658	WP_081711566.1		Pseudomonas alcaligenes	Proteobacteria	356	1	Cap4	1			1402
659	WP_174369528.1		Pseudomonas putida	Proteobacteria	357	1	Cap4	1			1399

660	WP_115487569.1		Pseudomonas fluorescens	Proteobacteria	358	2	Cap4	1			1401
661	WP_054917999.1		Pseudomonas sp. NBRC 111138	Proteobacteria	358	2	Cap4	0			
662	WP_082765976.1		Azotobacter vinelandii	Proteobacteria	359	1	Cap4	1			1397
663	WP_110753422.1		Pseudomonas sp. OV529	Proteobacteria	360	1	Cap4	1			1397
664	WP_145913390.1		Pseudomonas putida	Proteobacteria	361	1	Cap4	1			1397
665	WP_110816979.1		Psychrobacter sp. YP14	Proteobacteria	362	1	Cap4	1			1402
666	WP_125530141.1		Acinetobacter pittii	Proteobacteria	363	6	Cap4	1			1401
667	WP_032007338.1		Acinetobacter calcoaceticus/baumannii complex	Proteobacteria	363	6	Cap4	0			
668	WP_034120359.1		Acinetobacter nosocomialis	Proteobacteria	363	6	Cap4	0			
669	WP_114171532.1		Acinetobacter nosocomialis	Proteobacteria	363	6	Cap4	0			
670	WP_068544641.1		Acinetobacter pittii	Proteobacteria	363	6	Cap4	0			
671	WP_153570502.1		Acinetobacter nosocomialis	Proteobacteria	363	6	Cap4	0			
672	WP_121874808.1		Providencia stuartii	Proteobacteria	364	3	Cap4	1			1403
673	WP_092766249.1		Acinetobacter sp. DSM 11652	Proteobacteria	364	3	Cap4	0			
674	WP_196423911.1		Acinetobacter baretiae	Proteobacteria	364	3	Cap4	0			
675	WP_071243655.1		Acinetobacter baumannii	Proteobacteria	365	1	Cap4	1			1403
676	QNX10371.1		Acinetobacter seifertii	Proteobacteria	366	2	Cap4	1			1403
677	QNX50400.1		Acinetobacter seifertii	Proteobacteria	366	2	Cap4	0			
678	WP_151810427.1		Acinetobacter bereziniae	Proteobacteria	367	1	Cap4	1			1392
679	WP_086206455.1		Acinetobacter sp. ANC 4648	Proteobacteria	368	1	Cap4	1			1392
680	WP_151760573.1		Acinetobacter junii	Proteobacteria	369	2	Cap4	1			1391
681	WP_121523193.1		Acinetobacter chengduensis	Proteobacteria	369	2	Cap4	0			
682	WP_207693958.1		Acinetobacter nosocomialis	Proteobacteria	370	2	Cap4	1			1390
683	WP_209048466.1		Acinetobacter nosocomialis	Proteobacteria	370	2	Cap4	0			
684	WP_069120064.1		Acinetobacter pittii	Proteobacteria	371	2	Cap4	1			1391
685	WP_125535194.1		Acinetobacter baumannii	Proteobacteria	371	2	Cap4	0			
686	WP_049067328.1		Acinetobacter baumannii	Proteobacteria	372	2	Cap4	1			1390
687	WP_125508821.1		Acinetobacter baumannii	Proteobacteria	372	2	Cap4	0			
688	WP_193331718.1		Pseudoalteromonas ulvae	Proteobacteria	373	1	Cap4	1			1407
689	WP_044832204.1		Thalassomonas actiniarum	Proteobacteria	374	1	Cap4	1			1407
690	WP_033185147.1		Pseudoalteromonas sp. PLSV	Proteobacteria	375	1	Cap4	1			1408
691	WP_140959106.1		Aestuariibacter sp. GS-14	Proteobacteria	376	1	Cap4	1			1406
692	WP_025382324.1		Yersinia similis	Proteobacteria	377	1	Cap4	1			1406

693	WP_044617938.1		Gynuella sunshinyii	Proteobacteria	378	1	Cap4	1			1406
694	EGR4216333.1		Vibrio cholerae	Proteobacteria	379	2	Cap4	1			1406
695	EGR4195357.1		Vibrio cholerae	Proteobacteria	379	2	Cap4	0			
696	MBG3153183.1		Proteus mirabilis	Proteobacteria	380	2	Cap4	1			1406
697	EHD4470043.1		Providencia rettgeri	Proteobacteria	380	2	Cap4	0			
698	WP_049000993.1		Klebsiella pneumoniae	Proteobacteria	381	35	Cap4	1			1406
699	WP_064359895.1		Raoultella ornithinolytica	Proteobacteria	381	35	Cap4	0			
700	WP_039553962.1		Pectobacterium brasiliense	Proteobacteria	381	35	Cap4	0			
701	HAT8003895.1		Citrobacter koseri	Proteobacteria	381	35	Cap4	0			
702	MBD7431679.1		Klebsiella pneumoniae	Proteobacteria	381	35	Cap4	0			
703	EHG7891490.1		Citrobacter braakii	Proteobacteria	381	35	Cap4	0			
704	WP_117022464.1		Klebsiella pneumoniae	Proteobacteria	381	35	Cap4	0			
705	WP_104928149.1		Klebsiella pneumoniae	Proteobacteria	381	35	Cap4	0			
706	WP_115657375.1		Klebsiella pneumoniae	Proteobacteria	381	35	Cap4	0			
707	WP_101859569.1		Klebsiella pneumoniae	Proteobacteria	381	35	Cap4	0			
708	WP_205534821.1		Pectobacterium brasiliense	Proteobacteria	381	35	Cap4	0			
709	WP_110196389.1		Klebsiella variicola	Proteobacteria	381	35	Cap4	0			
710	WP_043951620.1		Enterobacter cloacae	Proteobacteria	381	35	Cap4	0			
711	ECD9610362.1		Salmonella enterica subsp. salamae	Proteobacteria	381	35	Cap4	0			
712	WP_117136144.1		Klebsiella variicola	Proteobacteria	381	35	Cap4	0			
713	EHJ5406845.1		Salmonella enterica subsp. enterica serovar Wedding	Proteobacteria	381	35	Cap4	0			
714	WP_077139241.1		Klebsiella	Proteobacteria	381	35	Cap4	0			
715	WP_123836899.1		Klebsiella sp. P1CD1	Proteobacteria	381	35	Cap4	0			
716	WP_172863070.1		Citrobacter braakii	Proteobacteria	381	35	Cap4	0			
717	WP_181231379.1		Citrobacter freundii	Proteobacteria	381	35	Cap4	0			
718	WP_047025820.1		Serratia sp. TEL	Proteobacteria	381	35	Cap4	0			
719	WP_117133109.1		Klebsiella variicola	Proteobacteria	381	35	Cap4	0			
720	AFM59754.1		Enterobacter cloacae subsp. dissolvens SDM	Proteobacteria	381	35	Cap4	0			
721	WP_142523944.1		Raoultella ornithinolytica	Proteobacteria	381	35	Cap4	0			
722	WP_019106759.1		Pantoea ananatis	Proteobacteria	381	35	Cap4	0			
723	WP_181616848.1		Enterobacter sp. RHBSTW-00593	Proteobacteria	381	35	Cap4	0			
724	WP_204746251.1		Klebsiella pneumoniae	Proteobacteria	381	35	Cap4	0			

725	WP_057533801.1		Yersinia pseudotuberculosis	Proteobacteria	381	35	Cap4	0			
726	WP_104896941.1		Raoultella ornithinolytica	Proteobacteria	381	35	Cap4	0			
727	KEY42476.1		Pantoea agglomerans	Proteobacteria	381	35	Cap4	0			
728	WP_187506252.1		Pantoea agglomerans	Proteobacteria	381	35	Cap4	0			
729	WP_033772794.1		Pantoea agglomerans	Proteobacteria	381	35	Cap4	0			
730	WP_129062838.1		Pantoea agglomerans	Proteobacteria	381	35	Cap4	0			
731	HAT1612954.1		Raoultella planticola	Proteobacteria	381	35	Cap4	0			
732	WP_072927191.1		Nissabacter archeti	Proteobacteria	381	35	Cap4	0			
733	WP_116686659.1		Pelagibaculum spongiae	Proteobacteria	382	1	Cap4	1			1407
734	WP_054204859.1		Pseudoalteromonas	Proteobacteria	383	1	Cap4	1			1407
735	WP_127555826.1		Saccharospirillum alexandrii	Proteobacteria	384	1	Cap4	1			1407
736	WP_167411574.1		Vibrio diazotrophicus	Proteobacteria	385	3	Cap4	1			1407
737	EHH2512332.1		Vibrio parahaemolyticus	Proteobacteria	385	3	Cap4	0			
738	MBE4384539.1		Vibrio parahaemolyticus	Proteobacteria	385	3	Cap4	0			
739	WP_106339830.1		Photobacterium damselae	Proteobacteria	386	1	Cap4	1			1407
740	WP_210449158.1		Vibrio crassostreae	Proteobacteria	387	1	Cap4	1			1407
741	WP_050912796.1		Vibrio harveyi	Proteobacteria	388	1	Cap4	1			1407
742	WP_163391077.1		Enterovibrio norvegicus	Proteobacteria	389	1	Cap4	1			1407
743	WP_155673474.1		Aliivibrio fischeri	Proteobacteria	390	1	Cap4	1			1407
744	WP_102497299.1		Vibrio splendidus	Proteobacteria	391	12	Cap4	1			1407
745	WP_039441412.1		Vibrio navarrensis	Proteobacteria	391	12	Cap4	0			
746	WP_140064774.1		Vibrio parahaemolyticus	Proteobacteria	391	12	Cap4	0			
747	WP_045628254.1		Vibrio vulnificus	Proteobacteria	391	12	Cap4	0			
748	HAS8402825.1		Vibrio vulnificus	Proteobacteria	391	12	Cap4	0			
749	WP_045589806.1		Vibrio vulnificus	Proteobacteria	391	12	Cap4	0			
750	WP_103194416.1		Vibrio vulnificus	Proteobacteria	391	12	Cap4	0			
751	EGQ7933218.1		Vibrio vulnificus	Proteobacteria	391	12	Cap4	0			
752	EGR0670747.1		Vibrio vulnificus	Proteobacteria	391	12	Cap4	0			
753	WP_088878684.1		Vibrio mediterranei	Proteobacteria	391	12	Cap4	0			
754	WP_129532857.1		Vibrio anguillarum	Proteobacteria	391	12	Cap4	0			
755	WP_009846253.1		Vibrio sp. MED222	Proteobacteria	391	12	Cap4	0			
756	WP_067865661.1		Neptuniibacter marinus	Proteobacteria	392	1	Cap4	1			1407
757	WP_155655510.1		Aliivibrio fischeri	Proteobacteria	393	1	Cap4	1			1402

758	WP_136799979.1		Desulfopila sp. IMCC35004	Proteobacteria	394	1	Cap4	1			1407
759	PIP07838.1		Syntrophobacterales bacterium CG23_combo_of_CG06-09_8_20_14_all_48_27	Proteobacteria	395	2	Cap4	1			1396
760	OIP88723.1		Syntrophaceae bacterium CG2_30_49_12	Proteobacteria	395	2	Cap4	0			
761	WP_129112085.1		Burkholderia pseudomallei	Proteobacteria	396	1	Cap4	1			1381
762	WP_090684887.1		Paraburkholderia phenazinium	Proteobacteria	397	1	Cap4	1			1380
763	WP_198111934.1		Inhella sp. 1Y17	Proteobacteria	398	1	Cap4	1			1377
764	MBD8695914.1		Stenotrophomonas sp. CFBP 13718	Proteobacteria	399	2	Cap4	1			1384
765	WP_197974566.1		Stenotrophomonas sp. CFBP 13718	Proteobacteria	399	2	Cap4	0			
766	PJO54016.1		Stenotrophomonas lactitubi	Proteobacteria	400	2	Cap4	1			1381
767	WP_207764212.1		Stenotrophomonas lactitubi	Proteobacteria	400	2	Cap4	0			
768	MBL8261438.1		Xanthomonadaceae bacterium	Proteobacteria	401	1	Cap4	1			1373
769	WP_049817997.1		Xanthomonas arboricola	Proteobacteria	402	2	Cap4	1			1384
770	WP_055846618.1		Xanthomonas sp. Leaf148	Proteobacteria	402	2	Cap4	0			
771	WP_200379897.1		Rubrivivax gelatinosus	Proteobacteria	403	1	Cap4	1			1386
772	WP_002935408.1		Thauera sp. 27	Proteobacteria	404	1	Cap4	1			1407
773	WP_124467896.1		Burkholderia cenocepacia	Proteobacteria	405	4	Cap4	1			1388
774	WP_069253472.1		Burkholderia cepacia	Proteobacteria	405	4	Cap4	0			
775	KWD58344.1		Burkholderia cepacia	Proteobacteria	405	4	Cap4	0			
776	WP_196488926.1		Burkholderia cepacia	Proteobacteria	405	4	Cap4	0			
777	WP_157273703.1		Thiobacillus denitrificans	Proteobacteria	406	1	Cap4	1			1377
778	WP_124614023.1		Burkholderia sp. Bp9143	Proteobacteria	407	1	Cap4	1			1377
779	KVV23129.1		Burkholderia cepacia	Proteobacteria	408	3	Cap4	1			1381
780	WP_080323514.1		Burkholderia cenocepacia	Proteobacteria	408	3	Cap4	0			
781	WP_196482993.1		Burkholderia cepacia	Proteobacteria	408	3	Cap4	0			
782	WP_209319249.1		Ralstonia solanacearum	Proteobacteria	409	2	Cap4	1			1382
783	NJZ81779.1		Ralstonia solanacearum	Proteobacteria	409	2	Cap4	0			
784	WP_197378064.1		Ralstonia pseudosolanacearum	Proteobacteria	410	1	Cap4	1			1377
785	WP_058428175.1		Pseudomonas syringae	Proteobacteria	411	1	Cap4	1			1379
786	WP_053153871.1		Pseudomonas protegens	Proteobacteria	412	1	Cap4	1			1379
787	WP_057430203.1		Pseudomonas syringae group	Proteobacteria	413	4	Cap4	1			1379
788	WP_122336375.1		Pseudomonas coronafaciens	Proteobacteria	413	4	Cap4	0			

789	WP_209996776.1		Pseudomonas sp. PvP028	Proteobacteria	413	4	Cap4	0			
790	WP_046267053.1		Pseudomonas syringae	Proteobacteria	413	4	Cap4	0			
791	WP_074781343.1		Pseudomonas bauzanensis	Proteobacteria	414	3	Cap4	1			1377
792	WP_185267559.1		Pseudomonas xiamenensis	Proteobacteria	414	3	Cap4	0			
793	WP_058343478.1		Pseudomonas aeruginosa	Proteobacteria	414	3	Cap4	0			
794	WP_165204595.1		Roseimicrobium sp. ORNL1	Verrucomicrobia	415	1	Cap4	1			1409
795	WP_126025965.1		Variovorax guangxiensis	Proteobacteria	416	1	Cap4	1			1366
796	MBI2754589.1		Betaproteobacteria bacterium	Proteobacteria	417	1	Cap4	1			1360
797	WP_121696978.1		Xenophilus sp. E41	Proteobacteria	418	1	Cap4	1			1361
798	MBL9219034.1		Opitutaceae bacterium	Verrucomicrobia	419	1	Cap4	1			1395
799	MBN2560744.1		Phycisphaerae bacterium	Planctomycetes	420	1	Cap4	1			1397
800	MBE7491070.1		Planctomycetes bacterium	Planctomycetes	421	1	Cap4	1			1405
801	KAA0255535.1		Acidobacteria bacterium	Acidobacteria	422	1	Cap4	1			1385
802	MBI3818880.1		Planctomycetes bacterium	Planctomycetes	423	1	Cap4	1			1400
803	RME02116.1		Bacteroidetes bacterium	Bacteroidetes	424	1	Cap4	1			1389
804	WP_155998546.1		Thioalkalivibrio sp. ALM2T	Proteobacteria	425	1	Cap4	1			1402
805	MYB37937.1		Gammaproteobacteria bacterium	Proteobacteria	426	2	Cap4	1			1376
806	MXY54003.1		Gammaproteobacteria bacterium	Proteobacteria	426	2	Cap4	0			
807	MYC71570.1		Gemmatimonadetes bacterium	Gemmatimonadetes	427	1	Cap4	1			1373
808	MYE39588.1		Chloroflexi bacterium	Chloroflexi	428	1	Cap4	1			1414
809	WP_146577824.1		Rhodopirellula pilleata	Planctomycetes	429	1	Cap4	1			1394
810	MBN2293378.1		Pirellulales bacterium	Planctomycetes	430	1	Cap4	1			1397
811	MAT70862.1		Planctomycetaceae bacterium	Planctomycetes	431	1	Cap4	1			1404
812	MAX36745.1		Gimesia sp.	Planctomycetes	432	1	Cap4	1			1403
813	KYF88440.1		Sorangium cellulosum	Proteobacteria	433	1	Cap4	1			1406
814	WP_206844304.1		Methylacidiphilum sp. IT5	Verrucomicrobia	434	1	Cap4	1			1342
815	NPV30225.1		Firmicutes bacterium	Firmicutes	435	1	Cap4	1			1409
816	MBC7083462.1		Firmicutes bacterium	Firmicutes	436	1	Cap4	1			1404
817	MBN2005803.1		Anaerolineae bacterium	Chloroflexi	437	1	Cap4	1			1392
818	WP_015225262.1		Halothece sp. PCC 7418	Cyanobacteria	438	1	TIR	1			1388
819	WP_110519169.1		Herpetosiphon llansteffanensis	Chloroflexi	439	1	Cap4	1			1398
820	WP_089729927.1		Candidatus Thiosymbion oneisti	Proteobacteria	440	1	Cap4	1			1398
821	WP_136524126.1		Geomonas ferrireducens	Proteobacteria	441	1	Cap4	1			1388
822	WP_150300898.1		Pseudomonas profundi	Proteobacteria	442	1	Cap4	1			1392

823	WP_043504263.1		Pseudomonas aeruginosa	Proteobacteria	443	8	Cap4	1			1392
824	WP_172765378.1		Pseudomonas aeruginosa	Proteobacteria	443	8	Cap4	0			
825	WP_114242274.1		Pseudomonas aeruginosa	Proteobacteria	443	8	Cap4	0			
826	WP_065285113.1		Pseudomonas aeruginosa	Proteobacteria	443	8	Cap4	0			
827	WP_074234472.1		Pseudomonas aeruginosa	Proteobacteria	443	8	Cap4	0			
828	MBG6430668.1		Pseudomonas aeruginosa	Proteobacteria	443	8	Cap4	0			
829	MBH9308361.1		Pseudomonas aeruginosa	Proteobacteria	443	8	Cap4	0			
830	WP_182835010.1		Pseudomonas sp. SR9	Proteobacteria	443	8	Cap4	0			
831	WP_139669303.1		Pseudomonas sp. F16(2018)	Proteobacteria	444	1	Cap4	1			1392
832	WP_079783616.1		Pseudomonas sp. LPH1	Proteobacteria	445	1	Cap4	1			1392
833	WP_135296069.1		Pseudomonas fluorescens	Proteobacteria	446	1	Cap4	1			1392
834	WP_099601923.1		Pseudomonas sp. 2995-1	Proteobacteria	447	10	Cap4	1			1392
835	RMM23402.1		Pseudomonas syringae pv. pisi	Proteobacteria	447	10	Cap4	0			
836	WP_024649000.1		Pseudomonas syringae	Proteobacteria	447	10	Cap4	0			
837	WP_020293158.1		Pseudomonas sp. CFII64	Proteobacteria	447	10	Cap4	0			
838	WP_129429846.1		Pseudomonas syringae	Proteobacteria	447	10	Cap4	0			
839	ELS43610.1		Pseudomonas syringae pv. syringae B64	Proteobacteria	447	10	Cap4	0			
840	WP_129422748.1		Pseudomonas syringae	Proteobacteria	447	10	Cap4	0			
841	WP_024663039.1		Pseudomonas syringae	Proteobacteria	447	10	Cap4	0			
842	WP_017254109.1		Pseudomonas tolaasii	Proteobacteria	447	10	Cap4	0			
843	WP_198799822.1		Pseudomonas sp. MF6787	Proteobacteria	447	10	Cap4	0			
844	WP_169857520.1		Pseudomonas mandelii	Proteobacteria	448	1	Cap4	1			1392
845	WP_124315008.1		Pseudomonas chlororaphis	Proteobacteria	449	1	Cap4	1			1392
846	PKO88875.1		Betaproteobacteria bacterium HGW-Betaproteobacteria-12	Proteobacteria	450	1	Cap4	1			1392
847	MBU2844273.1		Acidithiobacillus ferriphilus	Proteobacteria	451	1	Cap4	1			1391
848	WP_164450454.1		Thiorhodococcus minor	Proteobacteria	452	1	Cap4	1			1392
849	WP_186409083.1		Candidatus Accumulibacter aalborgensis	Proteobacteria	453	3	Cap4	1			1392
850	MBN8517195.1		Candidatus Accumulibacter sp.	Proteobacteria	453	3	Cap4	0			
851	MBN8439380.1		Candidatus Accumulibacter sp.	Proteobacteria	453	3	Cap4	0			
852	NLE73997.1		Actinobacteria bacterium	Actinobacteria	454	1	Cap4	1			1392
853	WP_070780869.1		Achromobacter xylosoxidans	Proteobacteria	455	1	Cap4	1			1392
854	WP_065058996.1		Paraburkholderia tropica	Proteobacteria	456	1	Cap4	1			1392

855	WP_060160466.1		Burkholderia ubonensis	Proteobacteria	457	10	Cap4	1			1392
856	WP_095400166.1		Burkholderia ubonensis	Proteobacteria	457	10	Cap4	0			
857	WP_059662344.1		Burkholderia ubonensis	Proteobacteria	457	10	Cap4	0			
858	WP_089483261.1		Burkholderia sp. HI2714	Proteobacteria	457	10	Cap4	0			
859	WP_059947001.1		Burkholderia ubonensis	Proteobacteria	457	10	Cap4	0			
860	WP_060318995.1		Burkholderia ubonensis	Proteobacteria	457	10	Cap4	0			
861	WP_060124385.1		Burkholderia ubonensis	Proteobacteria	457	10	Cap4	0			
862	WP_124491725.1		Burkholderia stagnalis	Proteobacteria	457	10	Cap4	0			
863	WP_060114156.1		Burkholderia ubonensis	Proteobacteria	457	10	Cap4	0			
864	WP_059875663.1		Burkholderia multivorans	Proteobacteria	457	10	Cap4	0			
865	WP_023009724.1		Marinobacter sp. EVN1	Proteobacteria	458	2	Cap4	1			1392
866	WP_135804462.1		Marinobacter confluentis	Proteobacteria	458	2	Cap4	0			
867	WP_205743546.1		Halomonas alimentaria	Proteobacteria	459	3	Cap4	1			1392
868	NAW35536.1		Halomonas alimentaria	Proteobacteria	459	3	Cap4	0			
869	EAT06262.1		delta proteobacterium MLMS-1	Proteobacteria	459	3	Cap4	0			
870	WP_018870814.1		Thioalkalivibrio sp. ALgr3	Proteobacteria	460	1	Cap4	1			1392
871	TAM50784.1		Burkholderiaceae bacterium	Proteobacteria	461	1	Cap4	1			1392
872	MBH1404088.1		Stenotrophomonas maltophilia	Proteobacteria	462	1	Cap4	1			1392
873	MBN8768756.1		Stenotrophomonas sp.	Proteobacteria	463	1	Cap4	1			1391
874	WP_147351593.1		unclassified Collinsella	Actinobacteria	464	2	Mrr	1			1350
875	RHJ39296.1		Collinsella sp. AM10-48	Actinobacteria	464	2	Mrr	0			
876	WP_196191563.1		Streptomyces sp. BRB081	Actinobacteria	465	1	Trypco2 + serine protease	1			1283
877	WP_055535423.1		Streptomyces graminilatus	Actinobacteria	466	1	Cap4	1			1295
878	WP_127445402.1		Streptomyces sp. WAC08452	Actinobacteria	467	8	Cap4	1			1293
879	WP_127438523.1		Streptomyces sp. WAC04189	Actinobacteria	467	8	Cap4	0			
880	MYS42353.1		Streptomyces sp. SID5998	Actinobacteria	467	8	Cap4	0			
881	WP_037642105.1		Streptomyces griseorubens	Actinobacteria	467	8	Cap4	0			
882	WP_164367721.1		Streptomyces diastaticus	Actinobacteria	467	8	Cap4	0			
883	WP_175462819.1		Streptomyces sp. KAI 90	Actinobacteria	467	8	Cap4	0			
884	WP_208898211.1		Streptomyces incarnatus	Actinobacteria	467	8	Cap4	0			
885	AKJ10103.1		Streptomyces incarnatus	Actinobacteria	467	8	Cap4	0			
886	OAA99023.1		Streptomyces sp. FXJ1.172	Actinobacteria	468	2	Cap4	1			1367
887	WP_159396099.1		Streptomyces sp. FXJ1.172	Actinobacteria	468	2	Cap4	0			

888	ABS03723.1		Kineococcus radiotolerans SRS30216 = ATCC BAA-149	Actinobacteria	469	2	Cap4	1			1362
889	WP_157873729.1		Kineococcus radiotolerans	Actinobacteria	469	2	Cap4	0			
890	WP_157746074.1		Micromonospora purpureochromogenes	Actinobacteria	470	1	Cap4	1			1353
891	WP_114032414.1		Sphaerisporangium album	Actinobacteria	471	1	Cap4	1			1357
892	OLB80079.1		Actinobacteria bacterium 13_2_20CM_2_71_6	Actinobacteria	472	1	Transmembrane	1			1402
893	WP_181361350.1	SsAvs4	Streptomyces sp. A244	Actinobacteria	473	2	TIR	1			1350
894	PTH87616.1		Streptomyces sp. A244	Actinobacteria	473	2	TIR	0			
895	PKQ29598.1		Actinobacteria bacterium HGW-Actinobacteria-10	Actinobacteria	474	1	Cap4	1			1341
896	WP_033260210.1		Kitasatospora	Actinobacteria	475	2	Nucleosidase (5' split)	1			1381
897	BAJ33315.1		Kitasatospora setae KM-6054	Actinobacteria	475	2	Nucleosidase (5' split)	0			
898	WP_177245521.1		Streptosporangium canum	Actinobacteria	476	1	Nucleosidase	1			1219
899	WP_194892248.1		Catenulispora pinisilvae	Actinobacteria	477	2	Nucleosidase (5' split)	1			1338
900	WP_194923278.1		Catenulispora pinisilvae	Actinobacteria	477	2	Nucleosidase (5' split)	0			
901	WP_146229093.1		Streptomyces sp. NWU339	Actinobacteria	478	2	Trypco2 + serine protease	1			1337
902	PWI05114.1		Streptomyces sp. NWU339	Actinobacteria	478	2	Trypco2 + serine protease	0			
903	WP_181785285.1		Streptomyces sp. PIP175	Actinobacteria	479	1	Nucleosidase (5' split)	1			1333
904	WP_123819330.1		Kitasatospora niigatensis	Actinobacteria	480	2	Nucleosidase (5' split)	1			1313
905	WP_148089481.1		Kitasatospora cineracea	Actinobacteria	480	2	Nucleosidase (5' split)	0			
906	WP_190030427.1		Streptomyces goshikiensis	Actinobacteria	481	2	Nucleosidase (5' split)	1			1327
907	GHD63344.1		Streptomyces goshikiensis	Actinobacteria	481	2	Nucleosidase (5' split)	0			
908	WP_161217601.1		unclassified Streptomyces	Actinobacteria	482	1	Nucleosidase (5' split)	1			1314
909	RST19787.1		Streptomyces sp. WAC04770	Actinobacteria	483	3	Nucleosidase (5' split)	1			1344
910	QCW80034.1		Streptomyces sp. S6	Actinobacteria	483	3	Nucleosidase (5' split)	0			

911	WP_125620528.1		Streptomyces sp. WAC04770	Actinobacteria	483	3	Nucleosidase (5' split)	0			
912	WP_053558482.1		Streptomyces sp. CFMR 7	Actinobacteria	484	1	Nucleosidase (5' split)	1			1346
913	WP_123528123.1		Streptomyces sp. 840.1	Actinobacteria	485	2	Nucleosidase (5' split)	1			1345
914	NED89970.1		Streptomyces sp. SID11233	Actinobacteria	485	2	Nucleosidase (5' split)	0			
915	WP_150517241.1		Streptomyces subrutilus	Actinobacteria	486	2	Nucleosidase (5' split)	1			1345
916	QEU78149.1		Streptomyces subrutilus	Actinobacteria	486	2	Nucleosidase (5' split)	0			
917	WP_158641955.1		Corallococcus sp. AB018	Proteobacteria	487	2	Nucleosidase	1			1373
918	RUO920688.1		Corallococcus sp. AB018	Proteobacteria	487	2	Nucleosidase	0			
919	WP_134520533.1		Cryobacterium sp. Hh14	Actinobacteria	488	1	Cap4	1			1323
920	WP_167489980.1		Nocardia terpenica	Actinobacteria	489	1	Mrr	1			1307
921	WP_189253503.1		Lentzea flava	Actinobacteria	490	1	Cap4	1			1204
922	WP_061989248.1		Amycolatopsis regifaucium	Actinobacteria	491	1	Cap4	1			1351
923	WP_126641065.1		Embleya hyalina	Actinobacteria	492	1	Trypco2 + serine protease	1			1382
924	WP_199819140.1		Streptomyces sp. FxanaA7	Actinobacteria	493	1	Trypco2 + serine protease	1			1379
925	WP_214957334.1		Streptomyces sp. ISL-43	Actinobacteria	494	1	Trypco2 + serine protease	1			1383
926	WP_191299249.1		Lentzea cavernae	Actinobacteria	495	1	Cap4	1			1367
927	WP_143682219.1		Streptomyces kasugaensis	Actinobacteria	496	5	Cap4	1			1359
928	GGL77361.1		Streptomyces humiferus	Actinobacteria	496	5	Cap4	0			
929	WP_189285124.1		Streptomyces humiferus	Actinobacteria	496	5	Cap4	0			
930	GGT47271.1		Streptomyces purpureus	Actinobacteria	496	5	Cap4	0			
931	WP_189203523.1		Streptomyces purpureus	Actinobacteria	496	5	Cap4	0			
932	WP_158691936.1		Streptomyces sp. CNH099	Actinobacteria	497	1	Cap4	1			1359
933	WP_181801256.1		Streptomyces sp. W18L9	Actinobacteria	498	1	Cap4	1			1359
934	WP_190031913.1		Streptomyces goshikiensis	Actinobacteria	499	1	Cap4	1			1358
935	AXE24131.1		Streptomyces globosus	Actinobacteria	500	2	Cap4	1			1354
936	WP_162793181.1		Streptomyces globosus	Actinobacteria	500	2	Cap4	0			
937	WP_073805250.1		unclassified Streptomyces	Actinobacteria	501	1	Cap4	1			1359
938	WP_123962333.1		Streptomyces sp. TLI_185	Actinobacteria	502	1	Cap4	1			1354
939	WP_166631392.1		Streptomyces sp. GC420	Actinobacteria	503	1	Cap4	1			1340
940	WP_168715495.1		Streptomyces sp. LRa12	Actinobacteria	504	2	Cap4	1			1344
941	MYS54722.1		Streptomyces sp. SID6013	Actinobacteria	504	2	Cap4	0			

942	WP_18994 2587.1		Streptomyces aurantiogriseus	Actinoba cteria	505	2	Cap4	1			1344
943	GGR4799 8.1		Streptomyces aurantiogriseus	Actinoba cteria	505	2	Cap4	0			
944	WP_13110 5708.1		unclassified Ornithinimicrobi um	Actinoba cteria	506	1	Cap4	1			1378
945	MBU2111 359.1		Actinobacteria bacterium	Actinoba cteria	507	2	Cap4	1			1375
946	MBU2073 670.1		Actinobacteria bacterium	Actinoba cteria	507	2	Cap4	0			
947	WP_10216 2227.1		Brevibacterium luteolum	Actinoba cteria	508	8	Cap4	1			1376
948	WP_16731 8431.1		Kocuria sp. KD4	Actinoba cteria	508	8	Cap4	0			
949	WP_15619 1421.1		Corynebacteriu m sp. 1959	Actinoba cteria	508	8	Cap4	0			
950	QGU0097 5.1		Corynebacteriu m sp. 1959	Actinoba cteria	508	8	Cap4	0			
951	WP_01919 3438.1		Corynebacteriu m timonense	Actinoba cteria	508	8	Cap4	0			
952	WP_07382 4974.1		Buchananella hordeovulneris	Actinoba cteria	508	8	Cap4	0			
953	WP_12485 1291.1		Buchananella hordeovulneris	Actinoba cteria	508	8	Cap4	0			
954	WP_19364 3618.1		Dermacoccus barathri	Actinoba cteria	508	8	Cap4	0			
955	WP_13510 3253.1		Micrococcus lylae	Actinoba cteria	509	1	Cap4	1			1376
956	WP_20185 0808.1		Myceligenerans indicum	Actinoba cteria	510	1	Cap4	1			1353
957	WP_12952 0805.1		Agromyces albus	Actinoba cteria	511	1	HTH-Cap4	1			1351
958	WP_19465 4187.1		Curtobacterium flaccumfaciens	Actinoba cteria	512	1	Cap4	1			1350
959	WP_16731 8552.1		Kocuria sp. KD4	Actinoba cteria	513	1	HTH-Cap4	1			1351
960	WP_16289 3159.1		Microbacterium halotolerans	Actinoba cteria	514	1	HTH-Cap4	1			1351
961	WP_15908 4957.1		Planctomonas deserti	Actinoba cteria	515	1	HTH-Cap4	1			1350
962	WP_13420 6329.1		Mycobacteroide s salmoniphilum	Actinoba cteria	516	1	HTH-Cap4	1			1354
963	WP_06942 7390.1		Mycolicibacteriu m porcinum	Actinoba cteria	517	1	HTH-Cap4	1			1352
964	WP_16289 9239.1		Mycobacterium shigaense	Actinoba cteria	518	1	HTH-Cap4	1			1355
965	WP_15793 1179.1		Mycobacteroide s abscessus	Actinoba cteria	519	1	HTH-Cap4	1			1352
966	WP_07961 6993.1		Mycobacteroide s abscessus	Actinoba cteria	520	1	HTH-Cap4	1			1351
967	WP_13325 1014.1		Mycobacterium montefiorense	Actinoba cteria	521	2	HTH-Cap4	1			1351
968	GBG3868 4.1		Mycobacterium montefiorense	Actinoba cteria	521	2	HTH-Cap4	0			
969	WP_15737 1059.1		Arthrobacter sp. Leaf141	Actinoba cteria	522	2	HTH-Cap4	1			1352
970	KQQ9619 3.1		Arthrobacter sp. Leaf141	Actinoba cteria	522	2	HTH-Cap4	0			
971	WP_13320 5853.1		Arthrobacter	Actinoba cteria	523	1	HTH-Cap4	1			1353
972	WP_18279 9640.1		Tessaracoccus sp. MC1679	Actinoba cteria	524	1	HTH-Cap4	1			1354
973	WP_14322 9783.1		Actinophytocola xanthii	Actinoba cteria	525	2	HTH-Cap4	1			1376
974	OLF13473 .1		Actinophytocola xanthii	Actinoba cteria	525	2	HTH-Cap4	0			
975	WP_14346 7020.1		Lentzea kentuckyensis	Actinoba cteria	526	1	HTH-Cap4	1			1362

976	WP_146615447.1		Nonomuraea sp. KC333	Actinobacteria	527	2	HTH-Cap4	1			1352
977	PZG23168.1		Nonomuraea sp. KC333	Actinobacteria	527	2	HTH-Cap4	0			
978	WP_148036139.1		Nonomuraea sp. C10	Actinobacteria	528	1	HTH-Cap4	1			1355
979	OUE17875.1		Clavibacter michiganensis	Actinobacteria	529	2	Cap4	1			1395
980	WP_143332084.1		Clavibacter michiganensis	Actinobacteria	529	2	Cap4	0			
981	QQR74046.1		Holophagales bacterium	Acidobacteria	530	1	Cap4	1			1388
982	WP_142266367.1		Streptomyces sp. SLBN-115	Actinobacteria	531	2	Short	1			1397
983	TQJ55995.1		Streptomyces sp. SLBN-115	Actinobacteria	531	2	Short	0			
984	WP_146197644.1		Promicromonospora sp. AC04	Actinobacteria	532	2	Cap4	1			1392
985	PUB20181.1		Promicromonospora sp. AC04	Actinobacteria	532	2	Cap4	0			
986	WP_089338933.1		Blastococcus sp. DSM 44272	Actinobacteria	533	1	Cap4	1			1397
987	HAL27589.1		Chloroflexi bacterium	Chloroflexi	534	1	Cap4	1			1394
988	WP_129675990.1		Candidatus Chloroploca sp. Khr17	Chloroflexi	535	1	Cap4	1			1392
989	WP_172653096.1		Rhodococcus imtechensis	Actinobacteria	536	1	Cap4	1			1402
990	WP_131825990.1		Mycolicibacter heraklionensis	Actinobacteria	537	2	Cap4	1			1401
991	OMC09057.1		Mycolicibacter heraklionensis	Actinobacteria	537	2	Cap4	0			
992	WP_167164323.1		Mycolicibacterium fluoranthenivorans	Actinobacteria	538	1	Cap4	1			1394
993	WP_206482467.1		Mycobacteroides abscessus	Actinobacteria	539	1	Cap4	1			1401
994	MBP6728183.1		Microthrixaceae bacterium	Actinobacteria	540	1	Cap4	1			1403
995	WP_095798086.1		Kocuria sp. WRN011	Actinobacteria	541	1	Cap4	1			1418
996	HAM23782.1		Actinobacteria bacterium	Actinobacteria	542	1	Cap4	1			1400
997	WP_131770525.1		Candidatus Frankia datiscae	Actinobacteria	543	1	Cap4	1			1399
998	WP_157881958.1		Streptomyces rubellomurinus	Actinobacteria	544	2	Cap4	1			1400
999	KJS63807.1		Streptomyces rubellomurinus	Actinobacteria	544	2	Cap4	0			
1000	WP_104815091.1		Kitasatospora sp. MMS16-BH015	Actinobacteria	545	1	Cap4	1			1399
1001	RZB19598.1		Streptomyces sp. F001	Actinobacteria	546	2	Cap4	1			1399
1002	WP_129798325.1		Streptomyces sp. F001	Actinobacteria	546	2	Cap4	0			
1003	WP_093686000.1		unclassified Streptomyces	Actinobacteria	547	2	Cap4	1			1399
1004	WP_190216023.1		Streptomyces griseosporeus	Actinobacteria	547	2	Cap4	0			
1005	TGG91619.1		Aphanocapsa feldmannii 277cV	Cyanobacteria	548	1	Cap4	1			1402
1006	WP_019348628.1		Mycolicibacterium	Actinobacteria	549	3	Cap4	1			1402
1007	WP_131821149.1		Mycobacterium vulneris	Actinobacteria	549	3	Cap4	0			

1008	OCB45912.1		Mycobacterium vulneris	Actinobacteria	549	3	Cap4	0			
1009	WP_026555688.1		Arthrobacter sp. 35W	Actinobacteria	550	1	Cap4	1			1394
1010	WP_153011629.1		Serinicoccus chungangensis	Actinobacteria	551	1	Cap4	1			1392
1011	WP_141217014.1		Rhodococcus sp. 14-2483-1-2	Actinobacteria	552	2	Cap4	1			1394
1012	OZF33769.1		Rhodococcus sp. 14-2483-1-2	Actinobacteria	552	2	Cap4	0			
1013	WP_149429245.1		Rhodococcus cavernicola	Actinobacteria	553	1	Cap4	1			1394
1014	WP_174187445.1		Nocardia barduliensis	Actinobacteria	554	1	Cap4	1			1360
1015	WP_153935006.1		Aeromicrobium sp. S22	Actinobacteria	555	1	Cap4	1			1362
1016	WP_076673055.1		unclassified Microbacterium	Actinobacteria	556	2	Cap4	1			1358
1017	SDO32857.1		Microbacterium sp. ru370.1	Actinobacteria	556	2	Cap4	0			
1018	WP_137241781.1		Plantibacter flavus	Actinobacteria	557	2	Cap4	1			1357
1019	TKJ98311.1		Plantibacter flavus	Actinobacteria	557	2	Cap4	0			
1020	WP_206824802.1		Microbacterium esteraromaticum	Actinobacteria	558	1	Cap4	1			1355
1021	WP_211878578.1		Pseudarthrobacter sp. NJ-Z5	Actinobacteria	559	1	Cap4	1			1359
1022	KZF02228.1		Rhodococcus sp. EPR-157	Actinobacteria	560	2	Cap4	1			1363
1023	WP_156520537.1		Rhodococcus sp. EPR-157	Actinobacteria	560	2	Cap4	0			
1024	MBQ1442287.1		Renibacterium sp.	Actinobacteria	561	1	Cap4	1			1357
1025	WP_104285656.1		Pseudoclavibacter sp. AY1F1	Actinobacteria	562	1	Cap4	1			1358
1026	NYD16139.1		Modestobacter sp. DSM 46791	Actinobacteria	563	2	Cap4	1			1420
1027	WP_179655182.1		Modestobacter sp. DSM 46791	Actinobacteria	563	2	Cap4	0			
1028	WP_025348374.1		Nocardia nova	Actinobacteria	564	2	Cap4	1			1332
1029	AHH16892.1		Nocardia nova SH22a	Actinobacteria	564	2	Cap4	0			
1030	SCL45623.1		Micromonospora eburnea	Actinobacteria	565	4	Cap4	1			1360
1031	WP_020217611.1		Salinispora arenicola	Actinobacteria	565	4	Cap4	0			
1032	WP_091114996.1		Micromonospora eburnea	Actinobacteria	565	4	Cap4	0			
1033	WP_018800481.1		Salinispora arenicola	Actinobacteria	565	4	Cap4	0			
1034	MBB6536700.1		Micromonospora sp. CNZ286	Actinobacteria	566	2	Cap4	1			1362
1035	WP_047894050.1		Micromonospora	Actinobacteria	566	2	Cap4	0			
1036	WP_095889100.1		Rhodococcus sp. ACPA4	Actinobacteria	567	1	Cap4	1			1360
1037	WP_194583201.1		Microbacterium sp. VKM Ac-2870	Actinobacteria	568	1	Cap4	1			1345
1038	WP_203579983.1		Microbacterium hibisci	Actinobacteria	569	1	Cap4	1			1345
1039	PTW74628.1		Nocardioides sp. CF204	Actinobacteria	570	2	Cap4	1			1358
1040	WP_107701792.1		Nocardioides sp. CF204	Actinobacteria	570	2	Cap4	0			
1041	WP_206436430.1		Corynebacterium sp. YIM 101343	Actinobacteria	571	2	Cap4	1			1358

1042	RSZ61228.1		Corynebacterium sp. YIM 101343	Actinobacteria	571	2	Cap4	0			
1043	WP_198947325.1		Cellulosimicrobium sp. KWT-B	Actinobacteria	572	1	Cap4	1			1359
1044	WP_202637497.1		Rhodococcus ruber	Actinobacteria	573	2	Cap4	1			1363
1045	WP_140189087.1		Rhodococcus ruber	Actinobacteria	573	2	Cap4	0			
1046	WP_211090682.1		Nesterenkonia sp. MY13	Actinobacteria	574	2	Cap4	1			1369
1047	NLS11010.1		Nesterenkonia sp. MY13	Actinobacteria	574	2	Cap4	0			
1048	WP_211604298.1		Phycicoccus sp. BSK3Z-2	Actinobacteria	575	2	Cap4	1			1369
1049	MBR7744779.1		Phycicoccus sp. BSK3Z-2	Actinobacteria	575	2	Cap4	0			
1050	WP_211294341.1		Lentzea kentuckyensis	Actinobacteria	576	1	Cap4	1			1352
1051	WP_211337230.1		Lentzea deserti	Actinobacteria	577	1	Cap4	1			1363
1052	WP_029130601.1		Salinispora pacifica	Actinobacteria	578	4	Trypco2 + serine protease	1			1361
1053	WP_028570838.1		Salinispora tropica	Actinobacteria	578	4	Trypco2 + serine protease	0			
1054	WP_027650030.1		Salinispora pacifica	Actinobacteria	578	4	Trypco2 + serine protease	0			
1055	WP_018722403.1		Salinispora fenicalii	Actinobacteria	578	4	Trypco2 + serine protease	0			
1056	WP_132238902.1		Micromonospora sp. CNZ303	Actinobacteria	579	1	Trypco2 + serine protease	1			1363
1057	WP_099845162.1		Verrucosispora sp. CNZ293	Actinobacteria	580	1	Trypco2 + serine protease	1			1358
1058	WP_206518015.1		Rhodococcus sp. X156	Actinobacteria	581	1	Cap4	1			1361
1059	WP_146483362.1		Streptomyces qinzhouensis	Actinobacteria	582	1	Cap4	1			1356
1060	WP_202460105.1		Streptomyces sp. SID1328	Actinobacteria	583	2	Cap4	1			1360
1061	MYV40166.1		Streptomyces sp. SID1328	Actinobacteria	583	2	Cap4	0			
1062	WP_211266448.1		Streptomyces ochraceisclerotic us	Actinobacteria	584	1	Cap4	1			1361
1063	MBK3584092.1		Streptomyces sp. MBT57	Actinobacteria	585	1	Cap4	1			1361
1064	WP_128503201.1		Streptomyces inhibens	Actinobacteria	586	1	Cap4	1			1361
1065	WP_185931339.1		Streptomyces sp. WAC08401	Actinobacteria	587	3	Trypco2 + serine protease	1			1361
1066	WP_203614280.1		Streptomyces rochei	Actinobacteria	587	3	Trypco2 + serine protease	0			
1067	WP_199577286.1		Streptomyces sp. CRB46	Actinobacteria	587	3	Trypco2 + serine protease	0			
1068	WP_189774664.1		Streptomyces tauricus	Actinobacteria	588	4	Trypco2 + serine protease	1			1361
1069	WP_051859642.1		Streptomyces xanthophaeus	Actinobacteria	588	4	Trypco2 + serine protease	0			

1070	WP_069919157.1		Streptomyces subrutilus	Actinobacteria	588	4	Trypco2 + serine protease	0			
1071	WP_209521939.1		Streptomyces sp. KCTC 0041BP	Actinobacteria	588	4	Trypco2 + serine protease	0			
1072	WP_188266647.1		unclassified Streptomyces	Actinobacteria	589	1	Cap4	1			1361
1073	WP_024761775.1		Streptomyces exfoliatus	Actinobacteria	590	1	Cap4	1			1361
1074	WP_086756090.1		Streptomyces scabiei	Actinobacteria	591	1	Cap4	1			1361
1075	WP_099499828.1		Streptomyces sp. HG99	Actinobacteria	592	4	Cap4	1			1361
1076	WP_206337078.1		Streptomyces dengpaensis	Actinobacteria	592	4	Cap4	0			
1077	WP_215448948.1		Streptomyces sp. ATCC 21386	Actinobacteria	592	4	Cap4	0			
1078	WP_199839877.1		Streptomyces scabiei	Actinobacteria	592	4	Cap4	0			
1079	WP_203634998.1		Streptomyces sp. SID10815	Actinobacteria	593	2	Cap4	1			1358
1080	NEA51428.1		Streptomyces sp. SID10815	Actinobacteria	593	2	Cap4	0			
1081	WP_071369131.1		Streptomyces colonosanans	Actinobacteria	594	3	Cap4	1			1361
1082	GHF19163.1		Streptomyces fumanus	Actinobacteria	594	3	Cap4	0			
1083	WP_190206825.1		Streptomyces fumanus	Actinobacteria	594	3	Cap4	0			
1084	WP_214516845.1		Streptomyces hygroscopicus	Actinobacteria	595	1	Cap4	1			1361
1085	WP_199810457.1		Streptomyces	Actinobacteria	596	2	Cap4	1			1361
1086	KOT54544.1		Streptomyces rimosus subsp. rimosus	Actinobacteria	596	2	Cap4	0			
1087	WP_199822425.1		Streptomyces	Actinobacteria	597	4	Cap4	1			1361
1088	KOT38866.1		Streptomyces caelestis	Actinobacteria	597	4	Cap4	0			
1089	WP_210732867.1		Streptomyces sp. RKND-216	Actinobacteria	597	4	Cap4	0			
1090	THA27867.1		Streptomyces sp. RKND-216	Actinobacteria	597	4	Cap4	0			
1091	WP_214298984.1		Streptomyces sp. Tu10	Actinobacteria	598	3	Cap4	1			1361
1092	WP_199821411.1		Streptomyces europaeiscabiei	Actinobacteria	598	3	Cap4	0			
1093	KND29525.1		Streptomyces europaeiscabiei	Actinobacteria	598	3	Cap4	0			
1094	WP_190077700.1		Streptomyces daghestanicus	Actinobacteria	599	3	Cap4	1			1361
1095	WP_190008730.1		Streptomyces niveoruber	Actinobacteria	599	3	Cap4	0			
1096	GGS84918.1		Streptomyces griseoviridis	Actinobacteria	599	3	Cap4	0			
1097	WP_086371075.1		Leptolyngbya	Cyanobacteria	600	1	Cap4	1			1516
1098	WP_155523800.1		Nodosilinea nodulosa	Cyanobacteria	601	1	Cap4	1			1522
1099	WP_008318602.1		Leptolyngbya sp. PCC 6406	Cyanobacteria	602	1	Cap4	1			1531
1100	WP_096585908.1		Anabaenopsis circularis	Cyanobacteria	603	1	Cap4	1			1545
1101	WP_009786348.1		Lyngbya sp. PCC 8106	Cyanobacteria	604	1	Cap4	1			1506
1102	WP_072722571.1		Planktothrix tepida	Cyanobacteria	605	1	Cap4	1			1441

1103	KXK16364.1		Armatimonadetes bacterium OLB18	Armatimonadetes	606	1	Cap4	1			1430
1104	HBJ85778.1		Verrucomicrobiales bacterium	Verrucomicrobia	607	1	Cap4	1			1477
1105	WP_215434541.1		Akkermansia muciniphila	Verrucomicrobia	608	2	Cap4	1			1454
1106	QWO90549.1		Akkermansia muciniphila	Verrucomicrobia	608	2	Cap4	0			
1107	HAF14961.1		Blastocatellia bacterium	Acidobacteria	609	1	Cap4	1			1448
1108	QDT13449.1		Planctomycetes bacterium K23_9	Planctomycetes	610	1	Cap4	1			1431
1109	WP_146516459.1		Rubripirellula amarantea	Planctomycetes	611	2	Cap4	1			1431
1110	WP_146600548.1		Novipirellula aureliae	Planctomycetes	611	2	Cap4	0			
1111	WP_007340586.1		Rhodopirellula baltica	Planctomycetes	612	3	Cap4	1			1431
1112	EGF25624.1		Rhodopirellula baltica WH47	Planctomycetes	612	3	Cap4	0			
1113	WP_007332049.1		Rhodopirellula baltica	Planctomycetes	612	3	Cap4	0			
1114	MBE7518481.1		Thermoflexaceae bacterium	Chloroflexi	613	1	Cap4	1			1366
1115	MXY37111.1		Dehalococcoidia bacterium	Chloroflexi	614	1	Cap4	1			1371
1116	WP_109817749.1		Micromonospora sp. 5R2A7	Actinobacteria	615	1	Cap4	1			1402
1117	WP_013287186.1		Micromonospora aurantiaca	Actinobacteria	616	1	Cap4	1			1404
1118	WP_146766716.1		Micromonospora noduli	Actinobacteria	617	2	Cap4	1			1406
1119	RAO19584.1		Micromonospora noduli	Actinobacteria	617	2	Cap4	0			
1120	WP_125440676.1		Hymenobacter perfusus	Bacteroidetes	618	1	Cap4	1			1442
1121	WP_157942291.1		Salinibacter altiplanensis	Bacteroidetes	619	1	Cap4	1			1465
1122	MAQ94469.1		Rhodothermaceae bacterium	Bacteroidetes	620	1	Cap4	1			1475
1123	WP_095512653.1		Rubrivirga marina	Rhodothermaeota	621	1	Cap4	1			1475
1124	WP_190361816.1		Cyanobacterium aponinum	Cyanobacteria	622	1	Short	1			1537
1125	WP_066122143.1		Geminocystis sp. NIES-3709	Cyanobacteria	623	1	Short	1			1523
1126	SLM18447.1		uncultured spirochete	Spirochaetes	624	1	Cap4	1			1472
1127	WP_012826160.1		Haliangium ochraceum	Proteobacteria	625	1	Cap4	1			1471
1128	WP_126935491.1		Corallococcus sp. AB018	Proteobacteria	626	2	Cap4	1			1487
1129	RUO89273.1		Corallococcus sp. AB018	Proteobacteria	626	2	Cap4	0			
1130	PZV02821.1		Cyanobium sp.	Cyanobacteria	627	1	Cap4	1			1464
1131	WP_175687336.1		Burkholderia multivorans	Proteobacteria	628	1	Cap4	1			1528
1132	WP_212162832.1		Burkholderia vietnamiensis	Proteobacteria	629	1	Cap4	1			1528
1133	WP_122174622.1		Caballeronia sordidicola	Proteobacteria	630	1	Cap4	1			1528
1134	WP_208717986.1		Pantoea cypripedii	Proteobacteria	631	1	Cap4	1			1452

1135	PKN26063.1		Deltaproteobacteria bacterium HGW-Deltaproteobacteria-22	Proteobacteria	632	1	Cap4	1			1464
1136	PKN47060.1		Deltaproteobacteria bacterium HGW-Deltaproteobacteria-17	Proteobacteria	633	1	Cap4	1			1464
1137	WP_156153004.1		Aeromonas sp. L_1B5_3	Proteobacteria	634	1	Cap4	1			1455
1138	WP_211056997.1		Pseudoalteromonas sp. MMG005	Proteobacteria	635	1	Cap4	1			1456
1139	WP_142572694.1		Vibrio cholerae	Proteobacteria	636	1	Cap4	1			1455
1140	HAS8249109.1		Vibrio vulnificus	Proteobacteria	637	3	Cap4	1			1455
1141	WP_017096219.1		Vibrio tasmaniensis	Proteobacteria	637	3	Cap4	0			
1142	EHI9242011.1		Vibrio vulnificus	Proteobacteria	637	3	Cap4	0			
1143	WP_178310203.1		Vibrio parahaemolyticus	Proteobacteria	638	1	Cap4	1			1455
1144	WP_145213921.1		Gimesia alba	Planctomycetes	639	1	Cap4	1			1444
1145	RJQ47689.1		Nitrospiraceae bacterium	Nitrospirae	640	1	Cap4	1			1465
1146	NEO33816.1		Symploca sp. SIO3C6	Cyanobacteria	641	1	Short	1			1460
1147	WP_190899006.1		Nostoc	Cyanobacteria	642	1	Short	1			1476
1148	MBE9053451.1		Nostocales cyanobacterium LEGE 11386	Cyanobacteria	643	1	Short	1			1476
1149	WP_145192657.1		Gimesia chilikensis	Planctomycetes	644	1	Cap4	1			1464
1150	WP_179915864.1		Halomonas glaciei	Proteobacteria	645	1	Cap4	1			1458
1151	WP_068229937.1		Rheinheimera sp. EpRS3	Proteobacteria	646	1	Cap4	1			1459
1152	WP_196857854.1		Pedobacter sp. CAN_A7	Bacteroidetes	647	1	Cap4	1			1420
1153	WP_089903779.1		Chitinophaga arvensicola	Bacteroidetes	648	1	Cap4	1			1438
1154	WP_113949267.1		Pedobacter miscanthi	Bacteroidetes	649	1	Cap4	1			1415
1155	HGW86899.1		Bacteroidia bacterium	Bacteroidetes	650	1	Cap4	1			1438
1156	WP_208699989.1		Sphingobacterium sp. UDSM-2020	Bacteroidetes	651	1	Cap4	1			1437
1157	WP_091129982.1		Flavobacterium anhuiense	Bacteroidetes	652	1	Cap4	1			1443
1158	WP_089354601.1		Flavobacterium sp. ov086	Bacteroidetes	653	1	Cap4	1			1445
1159	WP_173969642.1		Flavobacterium bizetiae	Bacteroidetes	654	1	Cap4	1			1443
1160	WP_035649296.1		Flavobacterium sp. ASV13	Bacteroidetes	655	1	Cap4	1			1443
1161	WP_119790741.1		Flavobacterium anhuiense	Bacteroidetes	656	1	Cap4	1			1442
1162	WP_196831239.1		Flavobacterium sp. CG_9.1	Bacteroidetes	657	1	Cap4	1			1445
1163	WP_123872512.1		Chryseobacterium bernardetii	Bacteroidetes	658	3	Cap4	1			1439
1164	WP_139785455.1		Chryseobacterium phocaeense	Bacteroidetes	658	3	Cap4	0			

1165	AZB27413.1		Chryseobacterium bernardetii	Bacteroidetes	658	3	Cap4	0			
1166	WP_072403480.1		Flaviramulus basaltis	Bacteroidetes	659	1	Cap4	1			1452
1167	HCU34377.1		Armatimonadetes bacterium	Armatimonadetes	660	1	Cap4	1			1441
1168	RNC81382.1		Phycisphaera sp.	Planctomycetes	661	1	Cap4	1			1496
1169	WP_158791510.1		Granulicella sp. L60	Acidobacteria	662	1	Cap4	1			1450
1170	WP_186695457.1		Edaphobacter sp. 4G125	Acidobacteria	663	1	Cap4	1			1482
1171	MBI5015326.1		Deltaproteobacteria bacterium	Proteobacteria	664	1	Cap4	1			1489
1172	WP_146511378.1		Thalassoglobus neptunius	Planctomycetes	665	1	Cap4	1			1513
1173	WP_068134823.1		Roseimaritima ulvae	Planctomycetes	666	1	Cap4	1			1510
1174	WP_156162438.1		Burkholderia sp. MSHR3999	Proteobacteria	667	2	Cap4	1			1463
1175	KIP17789.1		Burkholderia sp. MSHR3999	Proteobacteria	667	2	Cap4	0			
1176	TAJ91164.1		Reyranella sp.	Proteobacteria	668	1	Cap4	1			1455
1177	MBP1093709.1		Bradyrhizobium japonicum	Proteobacteria	669	1	Cap4	1			1473
1178	WP_198577076.1		Caulobacter hibisci	Proteobacteria	670	1	Cap4	1			1431
1179	WP_115380149.1		Sphingomonas sp. FARSPH	Proteobacteria	671	1	HNH	1			1440
1180	WP_143182766.1		Parasphingorhabdus marina	Proteobacteria	672	1	HNH	1			1443
1181	WP_157136825.1		Sphingomonas sp. PAMC 26617	Proteobacteria	673	1	HNH	1			1451
1182	WP_183671970.1		Pseudomonas umsongensis	Proteobacteria	674	2	Cap4	1			1455
1183	WP_028939503.1		Pseudomonas	Proteobacteria	674	2	Cap4	0			
1184	KJS14419.1		Hoeflea sp. BRH_c9	Proteobacteria	675	1	Cap4	1			1444
1185	WP_207079682.1		Novosphingobium sp. KA1	Proteobacteria	676	1	Cap4	1			1444
1186	WP_116859342.1		Sphingomonas sp. URHD0007	Proteobacteria	677	1	Cap4	1			1440
1187	TPG15803.1		Sphingomonas koreensis	Proteobacteria	678	1	Cap4	1			1446
1188	MBL8552257.1		Hyphomonadaceae bacterium	Proteobacteria	679	1	Cap4	1			1481
1189	WP_172789548.1		Bradyrhizobium sp. WBAH10	Proteobacteria	680	1	Cap4	1			1468
1190	PKN35391.1		Deltaproteobacteria bacterium HGW-Deltaproteobacteria-19	Proteobacteria	681	1	Cap4	1			1474
1191	MBU1057576.1		Proteobacteria bacterium	Proteobacteria	682	1	Cap4	1			1461
1192	RLW65771.1		gamma proteobacterium symbiont of Stewartia floridana	Proteobacteria	683	1	Cap4	1			1461
1193	WP_211393644.1		Bradyrhizobium diazoefficiens	Proteobacteria	684	1	Cap4	1			1453
1194	WP_080878755.1		Nitrospira sp. ND1	Nitrospirae	685	1	Cap4	1			1468
1195	MBP6264194.1		Nitrospira sp.	Nitrospirae	686	1	Cap4	1			1468
1196	WP_142021142.1		Bosea vestrisii	Proteobacteria	687	2	Cap4	1			1461

1197	TQN34848.1		Bosea vestrisii	Proteobacteria	687	2	Cap4	0			
1198	WP_184701987.1		Aminobacter aganoensis	Proteobacteria	688	1	Cap4	1			1472
1199	WP_182575736.1		Aminobacter ciceronei	Proteobacteria	689	2	Cap4	1			1472
1200	WP_210289174.1		Aminobacter ciceronei	Proteobacteria	689	2	Cap4	0			
1201	WP_156874934.1		Sulfitobacter sp. AM1-D1	Proteobacteria	690	1	Cap4	1			1457
1202	WP_139238625.1		Mesorhizobium sp. YR577	Proteobacteria	691	1	Cap4	1			1464
1203	WP_162738207.1		Rhizobium sp. BK463	Proteobacteria	692	1	Cap4	1			1464
1204	WP_145676954.1		Rhizobium lusitanum	Proteobacteria	693	1	Cap4	1			1464
1205	WP_206240177.1		Novosphingobium sp. GeG2	Proteobacteria	694	1	Cap4	1			1455
1206	WP_127904292.1		Solirhodobacter olei	Proteobacteria	695	1	Cap4	1			1456
1207	GEB36257.1		Gluconacetobacter liquefaciens	Proteobacteria	696	3	Cap4	1			1457
1208	WP_141288702.1		Gluconacetobacter liquefaciens	Proteobacteria	696	3	Cap4	0			
1209	RDI39619.1		Gluconacetobacter liquefaciens	Proteobacteria	696	3	Cap4	0			
1210	WP_056003740.1		Frateuria sp. Soil773	Proteobacteria	697	1	Cap4	1			1430
1211	WP_166207389.1		Lysobacter telluris	Proteobacteria	698	1	Cap4	1			1434
1212	WP_083892326.1	CcAvs4	Corallococcus coralloides	Proteobacteria	699	2	Nucleosidase	1			1454
1213	AFE06036.1		Corallococcus coralloides DSM 2259	Proteobacteria	699	2	Nucleosidase	0			
1214	WP_031358153.1		Caballeronia sordidicola	Proteobacteria	700	1	Cap4	1			1452
1215	WP_208459612.1		Paraburkholderia bryophila	Proteobacteria	701	1	Cap4	1			1452
1216	WP_206683615.1		Rhizobium changzhiense	Proteobacteria	702	1	Cap4	1			1440
1217	ANL40244.1		Rhizobium phaseoli	Proteobacteria	703	2	Cap4	1			1449
1218	WP_208613671.1		Rhizobium	Proteobacteria	703	2	Cap4	0			
1219	WP_197496317.1		Rhizobium sp. WYCCWR10014	Proteobacteria	704	3	Cap4	1			1451
1220	WP_197729054.1		Rhizobium leguminosarum	Proteobacteria	704	3	Cap4	0			
1221	OAV57516.1		Rhizobium sp. WYCCWR10014	Proteobacteria	704	3	Cap4	0			
1222	WP_199725042.1		Herbaspirillum	Proteobacteria	705	1	Cap4	1			1459
1223	MBN8486864.1		Burkholderiales bacterium	Proteobacteria	706	1	Cap4	1			1453
1224	WP_205894320.1		Pseudomonas stutzeri	Proteobacteria	707	1	Cap4	1			1456
1225	WP_205760302.1		Aeromonas caviae	Proteobacteria	708	1	Cap4	1			1457
1226	WP_043756940.1		Aeromonas taiwanensis	Proteobacteria	709	1	Cap4	1			1455
1227	TDV32491.1		Stenotrophomonas sp. CC22-02	Proteobacteria	710	2	Cap4	1			1453
1228	WP_208322794.1		Stenotrophomonas sp. CC22-02	Proteobacteria	710	2	Cap4	0			
1229	NYI19916.1		Xanthomonas campestris	Proteobacteria	711	1	Cap4	1			1453

1230	OBV41035.1		Janthinobacterium psychrotolerans	Proteobacteria	712	1	Cap4	1			1454
1231	WP_046967696.1		Luteibacter rhizovicinus	Proteobacteria	713	1	Cap4	1			1448
1232	WP_053549737.1		Desulfuromonas soudanensis	Proteobacteria	714	1	Cap4	1			1454
1233	TXT40388.1		Comamonadaceae bacterium	Proteobacteria	715	1	Cap4	1			1455
1234	WP_211165259.1		Aromatoleum toluolicum	Proteobacteria	716	1	Cap4	1			1454
1235	KVZ49458.1		Burkholderia ubonensis	Proteobacteria	717	2	Cap4	1			1455
1236	WP_198364146.1		Burkholderia ubonensis	Proteobacteria	717	2	Cap4	0			
1237	WP_204149556.1		Pseudomonas sp. MIL9	Proteobacteria	718	1	Cap4	1			1455
1238	CRM24431.1		Pseudomonas sp. 31 E 5	Proteobacteria	719	2	Cap4	1			1455
1239	WP_175404181.1		unclassified Pseudomonas	Proteobacteria	719	2	Cap4	0			
1240	WP_206433041.1		Pseudomonas chlororaphis	Proteobacteria	720	1	Cap4	1			1457
1241	WP_193560092.1		Pseudomonas arsenicoxydans	Proteobacteria	721	1	Cap4	1			1457
1242	AME28267.1		Burkholderia sp. PAMC 26561	Proteobacteria	722	1	Cap4	1			1454
1243	WP_192699477.1		Burkholderia sp. OAS925	Proteobacteria	723	1	Cap4	1			1454
1244	WP_055137897.1		Burkholderia plantarii	Proteobacteria	724	1	Cap4	1			1454
1245	WP_175800085.1		Burkholderia anthina	Proteobacteria	725	1	Cap4	1			1455
1246	WP_208446762.1		Burkholderia gladioli	Proteobacteria	726	5	Cap4	1			1455
1247	KGC04392.1		Burkholderia cepacia	Proteobacteria	726	5	Cap4	0			
1248	WP_059656283.1		Burkholderia ubonensis	Proteobacteria	726	5	Cap4	0			
1249	WP_207626041.1		Burkholderia gladioli	Proteobacteria	726	5	Cap4	0			
1250	EGD04944.1		Burkholderia sp. TJI49	Proteobacteria	726	5	Cap4	0			
1251	WP_186134490.1		Burkholderia gladioli	Proteobacteria	727	1	Cap4	1			1455
1252	HEA06652.1	Archaeon isolate HyVt-335	archaeon	Unclassified Archaea	728	1	Metallophosphoesterase	1			1447
1253	WP_173606509.1		Brevibacillus sp. HB2.2	Firmicutes	729	1	MBL + serine protease	1			1336
1254	WP_211142839.1		Clostridium aciditolerans	Firmicutes	730	1	MBL + serine protease	1			1378
1255	TET47497.1		candidate division TA06 bacterium	Unclassified	731	1	Cap4	1			1422
1256	NAS88381.1		ANME-1 cluster archaeon AG-394-G21	Euryarchaeota	732	1	Cap4	1			1426
1257	AOR37949.1		Candidatus Melainabacteria bacterium MEL.A1	Candidatus Melainabacteria	733	1	Cap4	1			1332
1258	NTZ63207.1		Agrobacterium tumefaciens	Proteobacteria	734	1	Cap4	1			1417
1259	WP_140747008.1		Mesorhizobium sp. B2-3-5	Proteobacteria	735	1	Cap4	1			1457

1260	WP_050402914.1		Bradyrhizobium embrapense	Proteobacteria	736	1	Cap4	1			1432
1261	TMJ52087.1		Alphaproteobacteria bacterium	Proteobacteria	737	1	Cap4	1			1436
1262	WP_157450311.1		Bradyrhizobium sp. ARR65	Proteobacteria	738	1	Cap4	1			1434
1263	WP_211377448.1		Bradyrhizobium diazoefficiens	Proteobacteria	739	1	Cap4	1			1441
1264	RXG856991		Bradyrhizobium zhanjiangense	Proteobacteria	740	2	Cap4	1			1439
1265	WP_164940508.1		Bradyrhizobium zhanjiangense	Proteobacteria	740	2	Cap4	0			
1266	WP_189486166.1		Asticcacaulis endophyticus	Proteobacteria	741	2	Cap4	1			1428
1267	GGZ32512.1		Asticcacaulis endophyticus	Proteobacteria	741	2	Cap4	0			
1268	MBA30206.1		Hyphomonadaceae bacterium	Proteobacteria	742	3	Cap4	1			1417
1269	WP_146740242.1		Hyphomonas sp. GM-8P	Proteobacteria	742	3	Cap4	0			
1270	RAN386400.1		Hyphomonas sp. GM-8P	Proteobacteria	742	3	Cap4	0			
1271	MBU0583096.1		Alphaproteobacteria bacterium	Proteobacteria	743	1	Cap4	1			1410
1272	PVA11426.1		Pelagivirga sediminicola	Proteobacteria	744	2	Cap4	1			1415
1273	WP_133176142.1		Pelagivirga sediminicola	Proteobacteria	744	2	Cap4	0			
1274	PTW63178.1		Breoghania corrubedonensis	Proteobacteria	745	2	Cap4	1			1415
1275	WP_107988647.1		Breoghania corrubedonensis	Proteobacteria	745	2	Cap4	0			
1276	AOY57264.1		Desulfococcus multivorans	Proteobacteria	746	2	Cap4	1			1419
1277	WP_020875279.1		Desulfococcus multivorans	Proteobacteria	746	2	Cap4	0			
1278	MBU0662958.1		Proteobacteria bacterium	Proteobacteria	747	1	Cap4	1			1419
1279	GBE53783.1		bacterium BMS3Bbin14	Unclassified	748	2	Cap4	1			1418
1280	HDO51477.1		Rhizobiales bacterium	Proteobacteria	748	2	Cap4	0			
1281	WP_169668814.1		Rhizobacter sp. SG490	Proteobacteria	749	1	Cap4	1			1409
1282	CAB3721727.1		Achromobacter piechaudii	Proteobacteria	750	2	Cap4	1			1423
1283	WP_061303916.1		Achromobacter piechaudii	Proteobacteria	750	2	Cap4	0			
1284	MBS0193860.1		Proteobacteria bacterium	Proteobacteria	751	1	Cap4	1			1417
1285	WP_184551860.1		Xanthomonas arboricola	Proteobacteria	752	2	Cap4	1			1421
1286	WP_104542909.1		Xanthomonas arboricola	Proteobacteria	752	2	Cap4	0			
1287	WP_184380433.1		Xanthomonas arboricola	Proteobacteria	753	1	Cap4	1			1420
1288	WP_200958309.1		Massilia sp. Root335	Proteobacteria	754	2	Cap4	1			1420
1289	KQV46196.1		Massilia sp. Root335	Proteobacteria	754	2	Cap4	0			
1290	AHZ73669.1		Pseudomonas mandelii JR-1	Proteobacteria	755	3	Cap4	1			1429
1291	WP_010466526.1		Pseudomonas mandelii	Proteobacteria	755	3	Cap4	0			
1292	WP_137807677.1		Pseudomonas	Proteobacteria	755	3	Cap4	0			
1293	WP_185710125.1		Pseudomonas cremoris	Proteobacteria	756	1	Cap4	1			1431
1294	WP_200660703.1		Paraburkholderia	Proteobacteria	757	1	Cap4	1			1431

1295	SFU20986.1		Paraburkholderia aspalathi	Proteobacteria	758	2	Cap4	1			1431
1296	WP_093638279.1		Paraburkholderia aspalathi	Proteobacteria	758	2	Cap4	0			
1297	WP_056654832.1		Rhizobacter sp. Root1221	Proteobacteria	759	1	Cap4	1			1435
1298	WP_089170234.1		Azotobacter chroococcum	Proteobacteria	760	2	Cap4	1			1434
1299	ASL29127.1		Azotobacter chroococcum	Proteobacteria	760	2	Cap4	0			
1300	WP_134466130.1		Paraburkholderia dipogonis	Proteobacteria	761	1	Cap4	1			1431
1301	AEA61788.1		Burkholderia gladioli BSR3	Proteobacteria	762	2	Cap4	1			1430
1302	WP_202947025.1		Burkholderia gladioli	Proteobacteria	762	2	Cap4	0			
1303	WP_107153664.1		Trinickia symbiotica	Proteobacteria	763	1	Cap4	1			1431
1304	WP_038782001.1		Burkholderia pseudomallei	Proteobacteria	764	2	Cap4	1			1432
1305	KGS96707.1		Burkholderia pseudomallei MSHR7498	Proteobacteria	764	2	Cap4	0			
1306	WP_175773048.1		Paraburkholderia phenazinium	Proteobacteria	765	1	Cap4	1			1430
1307	OUS04246.1		Rhodobacterales bacterium 52_120_T64	Proteobacteria	766	1	Cap4	1			1411
1308	WP_058188604.1		Terracidiphilus gabretensis	Acidobacteria	767	1	Cap4	1			1408
1309	WP_206419694.1		Minwuia thermotolerans	Proteobacteria	768	1	Cap4	1			1418
1310	WP_212479254.1		Bradyrhizobium liaoningense	Proteobacteria	769	1	Cap4	1			1409
1311	MBL6938191.1		Alphaproteobacteria bacterium	Proteobacteria	770	1	Cap4	1			1409
1312	MBO6635152.1		Parvibaculum sp.	Proteobacteria	771	1	Cap4	1			1415
1313	WP_154661148.1		Microvirga lotononidis	Proteobacteria	772	2	Cap4	1			1420
1314	EIM28014.1		Microvirga lotononidis	Proteobacteria	772	2	Cap4	0			
1315	ALA18573.1		Chelatococcus sp. CO-6	Proteobacteria	773	2	Cap4	1			1422
1316	WP_156332770.1		unclassified Chelatococcus	Proteobacteria	773	2	Cap4	0			
1317	WP_183687113.1		Rhizobium cellulosilyticum	Proteobacteria	774	1	Cap4	1			1409
1318	WP_143102886.1		Rhizobium sp. NFR07	Proteobacteria	775	3	Cap4	1			1425
1319	SFB60212.1		Rhizobium sp. NFR07	Proteobacteria	775	3	Cap4	0			
1320	WP_210058856.1		Rhizobium petrolearium	Proteobacteria	775	3	Cap4	0			
1321	OGA67699.1		Betaproteobacteria bacterium RIFCSPLOWO2_12_FULL_65_14	Proteobacteria	776	1	Cap4	1			1415
1322	WP_165032485.1		Mesorhizobium camelthorni	Proteobacteria	777	1	Cap4	1			1417
1323	WP_174046563.1		Agrobacterium tumefaciens	Proteobacteria	778	1	Cap4	1			1425
1324	MBC7951686.1		Rhodospirillaceae bacterium	Proteobacteria	779	1	Cap4	1			1425
1325	WP_155121374.1		Bryobacter aggregatus	Acidobacteria	780	1	Cap4	1			1415
1326	PYQ91581.1		Acidobacteria bacterium	Acidobacteria	781	1	Cap4	1			1412

1327	MBU1195725.1		Proteobacteria bacterium	Proteobacteria	782	1	Cap4	1			1419
1328	MBI5605267.1		Deltaproteobacteria bacterium	Proteobacteria	783	1	Cap4	1			1431
1329	NLH41883.1		Planctomycetes bacterium	Planctomycetes	784	1	Cap4	1			1415
1330	NGZ96714.1		Nitrospira sp. WS110	Nitrospirae	785	1	Cap4	1			1432
1331	TKB78351.1		Nitrospira sp.	Nitrospirae	786	1	Cap4	1			1428
1332	WP_186956567.1		Undibacterium sp. NL8W	Proteobacteria	787	1	Cap4	1			1418
1333	WP_027196478.1		Paraburkholderia sprentiae	Proteobacteria	788	1	Cap4	1			1428
1334	WP_169913426.1		Pseudomonas chlororaphis	Proteobacteria	789	1	Cap4	1			1425
1335	WP_011811394.1		Verminephrobacter eiseniae	Proteobacteria	790	1	Cap4	1			1395
1336	WP_102593252.1		unclassified Pseudomonas	Proteobacteria	791	1	Cap4	1			1421
1337	WP_095055580.1		Pseudomonas sp. Irchel s3b2	Proteobacteria	792	1	Cap4	1			1424
1338	MBU0601700.1		Gammaproteobacteria bacterium	Proteobacteria	793	1	Cap4	1			1421
1339	WP_012730266.1		Burkholderia pseudomallei	Proteobacteria	794	3	Cap4	1			1425
1340	WP_076915589.1		Burkholderia pseudomallei	Proteobacteria	794	3	Cap4	0			
1341	EQA90401.1		Burkholderia pseudomallei MSHR338	Proteobacteria	794	3	Cap4	0			
1342	WP_186087567.1		Burkholderia gladioli	Proteobacteria	795	4	Cap4	1			1425
1343	WP_186236924.1		Burkholderia gladioli	Proteobacteria	795	4	Cap4	0			
1344	WP_059859627.1		Burkholderia cepacia	Proteobacteria	795	4	Cap4	0			
1345	WP_186032784.1		Burkholderia	Proteobacteria	795	4	Cap4	0			
1346	WP_175115324.1		Paraburkholderia solisilvae	Proteobacteria	796	3	Cap4	1			1425
1347	WP_175776944.1		Burkholderia anthina	Proteobacteria	796	3	Cap4	0			
1348	CAB3772145.1		Paraburkholderia solisilvae	Proteobacteria	796	3	Cap4	0			
1349	WP_087042341.1		Caballeronia ptereochthonis	Proteobacteria	797	1	Cap4	1			1420
1350	HAL38943.1		Polaromonas sp.	Proteobacteria	798	1	Cap4	1			1421
1351	WP_132979292.1		Pigmentiphaga sp. D-2	Proteobacteria	799	1	Cap4	1			1421
1352	MBP6108196.1		Steroidobacteraceae bacterium	Proteobacteria	800	1	Cap4	1			1421
1353	WP_054622834.1	Ec3Avs4	Escherichia coli	Proteobacteria	801	9	Cap4	1			1422
1354	WP_097495575.1		Escherichia coli	Proteobacteria	801	9	Cap4	0			
1355	HAV9903934.1		Escherichia coli	Proteobacteria	801	9	Cap4	0			
1356	EFE7440959.1		Escherichia coli	Proteobacteria	801	9	Cap4	0			
1357	EGC5562949.1		Escherichia coli	Proteobacteria	801	9	Cap4	0			
1358	WP_115186717.1		Escherichia coli	Proteobacteria	801	9	Cap4	0			
1359	WP_097723085.1		Escherichia coli	Proteobacteria	801	9	Cap4	0			
1360	HAI4397292.1		Escherichia coli	Proteobacteria	801	9	Cap4	0			

1361	EHK31337 20.1		Escherichia coli	Proteoba cteria	801	9	Cap4	0			
1362	PTA95723 .1		Kluyvera sp. Nf5	Proteoba cteria	802	1	Cap4	1			1422
1363	WP_13706 9222.1		Escherichia coli	Proteoba cteria	803	46	Cap4	1			1422
1364	WP_15369 0081.1		Citrobacter braakii	Proteoba cteria	803	46	Cap4	0			
1365	HAU6233 262.1		Citrobacter freundii	Proteoba cteria	803	46	Cap4	0			
1366	EGO47630 36.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1367	WP_10547 6398.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1368	MBI99795 82.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1369	EEZ90259 22.1		Escherichia coli O136	Proteoba cteria	803	46	Cap4	0			
1370	WP_13753 9965.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1371	HBC10298 33.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1372	HBE40645 81.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1373	HBB71932 37.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1374	WP_20415 8894.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1375	WP_08965 0228.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1376	WP_12509 4048.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1377	QMP4425 7.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1378	WP_20462 8364.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1379	WP_18832 2774.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1380	WP_02264 5559.1		Enterobacteriace ae	Proteoba cteria	803	46	Cap4	0			
1381	EFB97660 06.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1382	WP_19450 5745.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1383	WP_08812 9937.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1384	HBC16790 76.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1385	HBB97151 51.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1386	WP_18779 7279.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1387	HAH2937 206.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1388	WP_08481 8554.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1389	EHP64815 89.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1390	EFJ374150 9.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1391	WP_10199 1001.1		Klebsiella quasipneumonia e	Proteoba cteria	803	46	Cap4	0			
1392	WP_20152 5306.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1393	HAW1709 870.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1394	EEV58061 58.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			

1395	HAJ07327 51.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1396	HBC07964 35.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1397	WP_21160 3472.1		Klebsiella pneumoniae	Proteoba cteria	803	46	Cap4	0			
1398	WP_19997 1960.1		Citrobacter koseri	Proteoba cteria	803	46	Cap4	0			
1399	EFB76257 70.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1400	HAI95441 39.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1401	WP_21379 6276.1		Klebsiella aerogenes	Proteoba cteria	803	46	Cap4	0			
1402	EEW2009 735.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1403	HAN9558 771.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1404	HAH4981 808.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1405	WP_16649 4933.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1406	WP_05933 7284.1		Escherichia coli	Proteoba cteria	803	46	Cap4	0			
1407	WP_19706 4012.1		Rahnella aquatilis	Proteoba cteria	803	46	Cap4	0			
1408	WP_08888 3868.1		Raoultella ornithinolytica	Proteoba cteria	803	46	Cap4	0			
1409	WP_19550 4627.1		Citrobacter farmeri	Proteoba cteria	804	1	Cap4	1			1422
1410	WP_20893 6049.1		Acidovorax avenae	Proteoba cteria	805	1	Cap4	1			1405
1411	WP_12194 2444.1		Acidovorax sp. 100	Proteoba cteria	806	1	Cap4	1			1409
1412	WP_05641 4440.1		unclassified Acidovorax	Proteoba cteria	807	3	Cap4	1			1405
1413	WP_05616 9086.1		Acidovorax sp. Leaf78	Proteoba cteria	807	3	Cap4	0			
1414	WP_05642 3078.1		Acidovorax sp. Leaf84	Proteoba cteria	807	3	Cap4	0			
1415	WP_16887 6841.1		Leeia aquatica	Proteoba cteria	808	1	Cap4	1			1419
1416	WP_00824 8757.1		Limnobacter sp. MED105	Proteoba cteria	809	1	Cap4	1			1422
1417	KFX71095 .1		Pseudomonas taeanensis MS-3	Proteoba cteria	810	2	Cap4	1			1422
1418	WP_20092 9594.1		Pseudomonas taeanensis	Proteoba cteria	810	2	Cap4	0			
1419	WP_20567 6616.1		Aquitalea aquatilis	Proteoba cteria	811	1	Cap4	1			1422
1420	WP_21056 2777.1		Chromobacteriu m violaceum	Proteoba cteria	812	1	Cap4	1			1422
1421	WP_05121 1369.1		Sphingomonas sp. LH128	Proteoba cteria	813	2	Cap4	1			1403
1422	EJU15111. 1		Sphingomonas sp. LH128	Proteoba cteria	813	2	Cap4	0			
1423	WP_19209 9563.1		Sphingomonas sp. CFBP 13706	Proteoba cteria	814	1	Cap4	1			1403
1424	WP_15635 9684.1		Sphingomonas sp. Leaf28	Proteoba cteria	815	2	Cap4	1			1403
1425	KQN0878 6.1		Sphingomonas sp. Leaf28	Proteoba cteria	815	2	Cap4	0			
1426	WP_18367 7885.1		unclassified Janthinobacteriu m	Proteoba cteria	816	1	Cap4	1			1406
1427	MBF0628 340.1		Magnetococcale s bacterium	Proteoba cteria	817	1	Cap4	1			1408
1428	WP_10760 8044.1		Sinorhizobium meliloti	Proteoba cteria	818	1	Cap4	1			1411

1429	ACS58884.1		Rhizobium leguminosarum bv. trifolii WSM1325	Proteobacteria	819	2	Cap4	1			1409
1430	WP_050786647.1		Rhizobium leguminosarum	Proteobacteria	819	2	Cap4	0			
1431	WP_101806876.1		Sinorhizobium medicae	Proteobacteria	820	3	Cap4	1			1411
1432	WP_127762961.1		Sinorhizobium medicae	Proteobacteria	820	3	Cap4	0			
1433	WP_127583843.1		Sinorhizobium meliloti	Proteobacteria	820	3	Cap4	0			
1434	WP_057258457.1		unclassified Ensifer	Proteobacteria	821	2	Cap4	1			1411
1435	KQW61030.1		Ensifer sp. Root1252	Proteobacteria	821	2	Cap4	0			
1436	WP_192657683.1		Devosia sp. OAS928	Proteobacteria	822	1	Cap4	1			1397
1437	WP_149148405.1		Agrobacterium sp. B1(2019)	Proteobacteria	823	1	Cap4	1			1397
1438	WP_174048615.1		Agrobacterium rubi	Proteobacteria	824	2	Cap4	1			1393
1439	WP_174003544.1		Agrobacterium rubi	Proteobacteria	824	2	Cap4	0			
1440	WP_038438917.1		Sinorhizobium meliloti	Proteobacteria	825	3	Cap4	1			1398
1441	WP_127643396.1		Sinorhizobium meliloti	Proteobacteria	825	3	Cap4	0			
1442	WP_128205153.1		Sinorhizobium meliloti	Proteobacteria	825	3	Cap4	0			
1443	WP_179695055.1		Pseudomonas moraviensis	Proteobacteria	826	1	Cap4	1			1382
1444	WP_123753504.1		Pseudomonas putida	Proteobacteria	827	1	Cap4	1			1389
1445	WP_203323070.1		Pseudoxanthomonas sp. REN9	Proteobacteria	828	1	Cap4	1			1400
1446	WP_158684135.1		Stenotrophomonas sp. MYb57	Proteobacteria	829	2	Cap4	1			1399
1447	AVJ33236.1		Stenotrophomonas sp. MYb57	Proteobacteria	829	2	Cap4	0			
1448	PZQ34134.1		Stenotrophomonas acidaminiphila	Proteobacteria	830	1	Cap4	1			1400
1449	WP_183856098.1		Rhizobium sp. BK456	Proteobacteria	831	1	Cap4	1			1406
1450	WP_097593929.1		Rhizobium sp. J15	Proteobacteria	832	1	Cap4	1			1411
1451	WP_018496366.1		Rhizobium leguminosarum	Proteobacteria	833	2	Cap4	1			1392
1452	WP_157635855.1		Rhizobium leguminosarum	Proteobacteria	833	2	Cap4	0			
1453	WP_111564300.1		Rhizobium etli	Proteobacteria	834	2	Cap4	1			1419
1454	RAK08040.1		Rhizobium etli	Proteobacteria	834	2	Cap4	0			
1455	WP_127966535.1		Agrobacterium sp. RS6	Proteobacteria	835	1	Cap4	1			1407
1456	KZC77793.1		Sphingobium yanoikuyae	Proteobacteria	836	2	Cap4	1			1415
1457	WP_153039489.1		Sphingobium yanoikuyae	Proteobacteria	836	2	Cap4	0			
1458	WP_155182398.1		Sphingobium sp. CAP-1	Proteobacteria	837	1	Cap4	1			1415
1459	WP_144907369.1		Novosphingobium taihuense	Proteobacteria	838	1	Cap4	1			1410
1460	MAF58853.1		Ponticaulis sp.	Proteobacteria	839	1	Cap4	1			1417
1461	WP_214202249.1		Novosphingobium profundi	Proteobacteria	840	1	Cap4	1			1417

1462	WP_183490836.1		Martelella radicis	Proteobacteria	841	1	Cap4	1			1412
1463	WP_146227961.1		Rhodobacter viridis	Proteobacteria	842	1	Cap4	1			1411
1464	WP_166404242.1		Rhodobacter sp. M37P	Proteobacteria	843	1	Cap4	1			1406
1465	PHR25419.1		Hoeflea sp.	Proteobacteria	844	1	Cap4	1			1412
1466	WP_129158541.1		Bosea sp. Tri-44	Proteobacteria	845	1	Cap4	1			1394
1467	MAM87938.1		unclassified Hahellaceae	Proteobacteria	846	2	Cap4	1			1404
1468	MBQ0747093.1		Marinobacter sp.	Proteobacteria	846	2	Cap4	0			
1469	WP_046903681.1		Altererythrobacter atlanticus	Proteobacteria	847	1	Cap4	1			1411
1470	WP_135215940.1		Bradyrhizobium sp. MOS001	Proteobacteria	848	1	Cap4	1			1419
1471	WP_213290448.1		Bradyrhizobium sp. sGM-13	Proteobacteria	849	1	Cap4	1			1409
1472	WP_120217810.1		Sphingopyxis sp. EG6	Proteobacteria	850	1	Cap4	1			1410
1473	WP_100092286.1		Sphingorhabdus sp. YGSMI21	Proteobacteria	851	1	Cap4	1			1412
1474	WP_170259531.1		Rhizobium laguerreae	Proteobacteria	852	1	Cap4	1			1412
1475	RWH83971.1		Mesorhizobium sp.	Proteobacteria	853	1	Cap4	1			1420
1476	WP_184314051.1		Rhizobium esperanzae	Proteobacteria	854	1	Cap4	1			1425
1477	WP_132537158.1		Rhizobium sp. PP-F2F-G48	Proteobacteria	855	1	Cap4	1			1415
1478	WP_206990102.1		Rhizobium laguerreae	Proteobacteria	856	1	Cap4	1			1415
1479	WP_107930189.1		Sphingomonas aerolata	Proteobacteria	857	1	Cap4	1			1419
1480	WP_132911219.1		Sphingomonas sp. BK235	Proteobacteria	858	1	Cap4	1			1419
1481	GGE01882.1		Aureimonas endophytica	Proteobacteria	859	2	Cap4	1			1420
1482	WP_188908232.1		Aureimonas endophytica	Proteobacteria	859	2	Cap4	0			
1483	WP_141167797.1		Pararhizobium sp. BGMRC6574	Proteobacteria	860	1	Cap4	1			1419
1484	WP_047030938.1		Hoeflea sp. IMCC20628	Proteobacteria	861	1	Cap4	1			1419
1485	WP_002723236.1		Cereibacter sphaeroides	Proteobacteria	862	1	Cap4	1			1419
1486	WP_097932346.1		Monaibacterium marinum	Proteobacteria	863	1	Cap4	1			1419
1487	TNF11693.1		Rhodobacteraceae bacterium	Proteobacteria	864	1	Cap4	1			1418
1488	WP_183474918.1		Limimaricola variabilis	Proteobacteria	865	1	Cap4	1			1418
1489	RWG87387.1		Mesorhizobium sp.	Proteobacteria	866	2	Cap4	1			1415
1490	RWL59689.1		Mesorhizobium sp.	Proteobacteria	866	2	Cap4	0			
1491	QND64702.1		Mesorhizobium loti	Proteobacteria	867	1	Cap4	1			1412
1492	ESY92900.1		Mesorhizobium sp. LNHC229A00	Proteobacteria	868	2	Cap4	1			1420
1493	WP_156939389.1		Mesorhizobium sp. LNHC229A00	Proteobacteria	868	2	Cap4	0			
1494	WP_051505787.1		Mesorhizobium sp. WSM2561	Proteobacteria	869	1	Cap4	1			1412

1495	RWC25855.1		Mesorhizobium sp.	Proteobacteria	870	1	Cap4	1			1412
1496	WP_042126875.1		Pseudomonas japonica	Proteobacteria	871	1	Cap4	1			1412
1497	WP_125861701.1		Pseudomonas xanthomarina	Proteobacteria	872	2	Cap4	1			1411
1498	RRU75153.1		Pseudomonas xanthomarina	Proteobacteria	872	2	Cap4	0			
1499	WP_201021372.1		Pseudomonas sp. Y39-6	Proteobacteria	873	1	Cap4	1			1411
1500	OLY75735.1		Pseudomonas sp. ATCC PTA-122608	Proteobacteria	874	2	Cap4	1			1411
1501	WP_143503562.1		Pseudomonas sp. ATCC PTA-122608	Proteobacteria	874	2	Cap4	0			
1502	WP_083352895.1		Pseudomonas moraviensis	Proteobacteria	875	3	Cap4	1			1410
1503	SDU40395.1		Pseudomonas moraviensis	Proteobacteria	875	3	Cap4	0			
1504	WP_159245232.1		Pseudomonas koreensis	Proteobacteria	875	3	Cap4	0			
1505	WP_141758760.1		Duganella sp. HH105	Proteobacteria	876	1	Cap4	1			1412
1506	WP_188916071.1		Massilia buxea	Proteobacteria	877	3	Cap4	1			1415
1507	GGC19857.1		Massilia buxea	Proteobacteria	877	3	Cap4	0			
1508	WP_155470725.1		Massilia buxea	Proteobacteria	877	3	Cap4	0			
1509	NCE90050.1		Pseudomonas sp. L13	Proteobacteria	878	2	Cap4	1			1410
1510	WP_161904587.1		Pseudomonas sp. L13	Proteobacteria	878	2	Cap4	0			
1511	WP_210692821.1		Pseudomonas protegens	Proteobacteria	879	1	Cap4	1			1419
1512	WP_198706990.1		Pseudomonas paralactis	Proteobacteria	880	1	Cap4	1			1410
1513	WP_146241872.1		Pseudomonas jessenii	Proteobacteria	881	2	Cap4	1			1415
1514	PYY71121.1		Pseudomonas jessenii	Proteobacteria	881	2	Cap4	0			
1515	WP_192182870.1		Mesorhizobium amorphae	Proteobacteria	882	1	Cap4	1			1417
1516	WP_173956967.1		Acetobacter syzygii	Proteobacteria	883	1	Cap4	1			1415
1517	WP_179506258.1		unclassified Sphingomonas	Proteobacteria	884	1	Cap4	1			1418
1518	WP_142882281.1		Rhodopseudomonas palustris	Proteobacteria	885	1	Cap4	1			1414
1519	WP_049433302.1		Stenotrophomonas maltophilia	Proteobacteria	886	1	Cap4	1			1415
1520	WP_131588133.1		Rhizobium pisi	Proteobacteria	887	5	Cap4	1			1415
1521	WP_130716241.1		Rhizobium leguminosarum	Proteobacteria	887	5	Cap4	0			
1522	WP_029875085.1		Rhizobium leguminosarum	Proteobacteria	887	5	Cap4	0			
1523	WP_130769560.1		Rhizobium leguminosarum	Proteobacteria	887	5	Cap4	0			
1524	WP_130759910.1		Rhizobium leguminosarum	Proteobacteria	887	5	Cap4	0			
1525	WP_206990150.1		Rhizobium laguerreae	Proteobacteria	888	1	Cap4	1			1415
1526	EGO93248.1		Acidiphilium sp. PM	Proteobacteria	889	2	Cap4	1			1412
1527	WP_039889346.1		Acidiphilium sp. PM	Proteobacteria	889	2	Cap4	0			
1528	WP_211500661.1		Gluconobacter cerinus	Proteobacteria	890	1	Cap4	1			1418

1529	WP_093084759.1		Sphingobium sp. AP50	Proteobacteria	891	1	Cap4	1			1412
1530	WP_183964003.1		Sagittula marina	Proteobacteria	892	1	Cap4	1			1421
1531	WP_056685510.1		Sphingobium sp. Leaf26	Proteobacteria	893	1	Mrr	1			1405
1532	WP_156649600.1		Methylobacterium sp. Leaf89	Proteobacteria	894	3	Mrr	1			1405
1533	NCP10962.1		Sphingomonadales bacterium	Proteobacteria	894	3	Mrr	0			
1534	KQO73299.1		Methylobacterium sp. Leaf89	Proteobacteria	894	3	Mrr	0			
1535	MBE9604719.1		Acetobacteraceae bacterium H6797	Proteobacteria	895	1	Mrr	1			1405
1536	WP_144423474.1		Sphingopyxis sp. 113P3	Proteobacteria	896	2	Mrr	1			1423
1537	ALC10467.1		Sphingopyxis sp. 113P3	Proteobacteria	896	2	Mrr	0			
1538	WP_126079062.1		unclassified Mesorhizobium	Proteobacteria	897	1	Mrr	1			1424
1539	WP_156941668.1		Mesorhizobium sp. WSM2561	Proteobacteria	898	1	Mrr	1			1424
1540	WP_132975970.1		Thiobaca trueperi	Proteobacteria	899	1	Cap4	1			1448
1541	WP_149201753.1		Azospirillum oryzae	Proteobacteria	900	1	Cap4	1			1446
1542	MBI9080162.1		Pseudodesulfovibrio sp.	Proteobacteria	901	1	Cap4	1			1431
1543	WP_210514475.1		Thalassospira sp. ER-Se-21-Dark	Proteobacteria	902	1	Cap4	1			1443
1544	MBC8492775.1		Chloroflexi bacterium	Chloroflexi	903	1	Cap4_split	1	MBC8492776.1	913	1411
1545	WP_069072355.1		Nostoc sp. KVJ20	Cyanobacteria	904	2	Cap4_split	1	WP_069072354.1	935	1433
1546	WP_073597665.1		Hydrococcus rivularis	Cyanobacteria	904	2	Cap4_split	0	WP_073597664.1	939	
1547	WP_071105463.1		Moorea producens	Cyanobacteria	905	1	Cap4_split	1	WP_071105464.1	923	1420
1548	NET54778.1		Symploca sp. SIO2E6	Cyanobacteria	906	1	Cap4_split	1	NET54779.1	924	1422
1549	WP_190721665.1		Oscillatoria sp. FACHB-1406	Cyanobacteria	907	1	Cap4_split	1	WP_190721667.1	920	1420
1550	WP_193962484.1		Dolichospermum sp. LEGE 00246	Cyanobacteria	908	1	Caspase + EAD8	1	WP_193962485.1	928	1425
1551	WP_168507913.1		Anabaena sp. UHCC 0187	Cyanobacteria	909	1	Caspase split	1	WP_168507915.1	933	1430
1552	WP_187336415.1		Novosphingopyxis sp. PB63	Proteobacteria	910	1	Cap4	1			1397
1553	TAJ80637.1		Sphingobium sp.	Proteobacteria	911	1	Cap4	1			1399
1554	WP_067678364.1		Erythrobacter sp. HI0063	Proteobacteria	912	2	Cap4	1			1399
1555	WP_056696257.1		unclassified Aureimonas	Proteobacteria	912	2	Cap4	0			
1556	WP_023826286.1		Mesorhizobium sp. L103C131B0	Proteobacteria	913	1	Cap4	1			1399
1557	RWN26108.1		Mesorhizobium sp.	Proteobacteria	914	1	Cap4	1			1396
1558	WP_211879885.1		Roseomonas hellenica	Proteobacteria	915	1	Cap4	1			1398

1559	WP_057848734.1		Bradyrhizobium valentinum	Proteobacteria	916	3	Cap4	1			1441
1560	WP_057901107.1		Bradyrhizobium valentinum	Proteobacteria	916	3	Cap4	0			
1561	KRR13022.1		Bradyrhizobium valentinum	Proteobacteria	916	3	Cap4	0			
1562	WP_159646731.1		Sphingorhabdus sp. 109	Proteobacteria	917	1	Cap4	1			1440
1563	WP_046500745.1		Candidatus Filomicrobium marinum	Proteobacteria	918	2	Cap4	1			1446
1564	CFX24687.1		Candidatus Filomicrobium marinum	Proteobacteria	918	2	Cap4	0			
1565	WP_150131414.1		Caulobacter mirabilis	Proteobacteria	919	2	Cap4	1			1445
1566	ATQ43385.1		Caulobacter mirabilis	Proteobacteria	919	2	Cap4	0			
1567	WP_138891443.1		Mycobacterium sp. KBS0706	Actinobacteria	920	1	Cap4	1			1448
1568	WP_183828650.1		Aquamicrobium lusatiense	Proteobacteria	921	1	Cap4	1			1448
1569	WP_212313086.1		unclassified Bradyrhizobium	Proteobacteria	922	1	Cap4	1			1445
1570	WP_011567417.1		Roseobacter denitrificans	Proteobacteria	923	1	Cap4	1			1443
1571	WP_058313483.1		Cognatishimia activa	Proteobacteria	924	1	Cap4	1			1443
1572	WP_014880289.1		Phaeobacter inhibens	Proteobacteria	925	1	Cap4	1			1443
1573	KCZ85078.1		Hyphomonas adhaerens MHS-3	Proteobacteria	926	2	Cap4	1			1488
1574	WP_156942167.1		Hyphomonas adhaerens	Proteobacteria	926	2	Cap4	0			
1575	MBP9233242.1		Hyphomonadaceae bacterium	Proteobacteria	927	1	Cap4	1			1488
1576	WP_147406504.1		Pseudomonas reidholzensis	Proteobacteria	928	2	Cap4	1			1486
1577	SYX89984.1		Pseudomonas reidholzensis	Proteobacteria	928	2	Cap4	0			
1578	WP_130756134.1		Rhizobium leguminosarum	Proteobacteria	929	2	Cap4	1			1460
1579	WP_168305100.1		Rhizobium leguminosarum	Proteobacteria	929	2	Cap4	0			
1580	WP_130808531.1		Rhizobium leguminosarum	Proteobacteria	930	4	Cap4	1			1485
1581	WP_207579709.1		Rhizobium ruizarguesonis	Proteobacteria	930	4	Cap4	0			
1582	WP_130771937.1		Rhizobium leguminosarum	Proteobacteria	930	4	Cap4	0			
1583	WP_130692961.1		Rhizobium leguminosarum	Proteobacteria	930	4	Cap4	0			
1584	WP_062693290.1		Rhizobium sp. Leaf341	Proteobacteria	931	1	Cap4	1			1484
1585	WP_209317755.1		Ralstonia solanacearum	Proteobacteria	932	1	Cap4	1			1463
1586	PZQ45576.1		Rhodovulum sulfidophilum	Proteobacteria	933	1	Cap4	1			1488
1587	WP_207873133.1		Alcaligenes sp. SORT26	Proteobacteria	934	1	Cap4	1			1483
1588	WP_008111132.1		Herbaspirillum sp. YR522	Proteobacteria	935	1	Cap4	1			1480
1589	WP_205711599.1		Janthinobacterium lividum	Proteobacteria	936	2	Cap4	1			1484
1590	NHQ94292.1		Janthinobacterium lividum	Proteobacteria	936	2	Cap4	0			
1591	WP_183755008.1		Pseudochelatococcus contaminans	Proteobacteria	937	1	Cap4	1			1490

1592	WP_149916555.1		Agrobacterium tumefaciens	Proteobacteria	938	1	Cap4	1			1485
1593	WP_212275209.1		Bradyrhizobium sp. AUGA SZCCT0169	Proteobacteria	939	1	Cap4	1			1485
1594	WP_193557291.1		Microvirga pakistanensis	Proteobacteria	940	1	Cap4	1			1480
1595	HAT51263.1		Alphaproteobacteria bacterium	Proteobacteria	941	2	Cap4	1			1481
1596	MBF0604161.1		Nitrospirae bacterium	Nitrospirae	941	2	Cap4	0			
1597	WP_100688340.1		Marinobacter aromaticivorans	Proteobacteria	942	1	Cap4	1			1485
1598	WP_038484064.1		Halomonas campaniensis	Proteobacteria	943	2	Cap4	1			1483
1599	WP_092572432.1		Aidingimonas halophila	Proteobacteria	943	2	Cap4	0			
1600	MBQ0832937.1		Marinobacter sp.	Proteobacteria	944	1	Cap4	1			1488
1601	WP_198024759.1		Bradyrhizobium sp. Cp5.3	Proteobacteria	945	1	Cap4	1			1478
1602	NMX60330.1		Pseudomonas sp. WS 5079	Proteobacteria	946	2	Cap4	1			1475
1603	WP_172311386.1		Pseudomonas sp. WS 5079	Proteobacteria	946	2	Cap4	0			
1604	WP_017683869.1		Pseudomonas syringae	Proteobacteria	947	1	Cap4	1			1480
1605	NMZ99813.1		Pseudomonas lundensis	Proteobacteria	948	4	Cap4	1			1472
1606	WP_169894252.1		Pseudomonas lundensis	Proteobacteria	948	4	Cap4	0			
1607	SEN97945.1		Pseudomonas sp. ok266	Proteobacteria	948	4	Cap4	0			
1608	WP_175559582.1		Pseudomonas sp. ok266	Proteobacteria	948	4	Cap4	0			
1609	WP_054079168.1		Pseudomonas amygdali	Proteobacteria	949	5	Cap4	1			1472
1610	WP_169951626.1		unclassified Pseudomonas	Proteobacteria	949	5	Cap4	0			
1611	WP_122690622.1		Pseudomonas viridiflava	Proteobacteria	949	5	Cap4	0			
1612	WP_122545112.1		Pseudomonas viridiflava	Proteobacteria	949	5	Cap4	0			
1613	WP_059397352.1		Pseudomonas libanensis	Proteobacteria	949	5	Cap4	0			
1614	WP_150706720.1		Pseudomonas fluorescens	Proteobacteria	950	1	Cap4	1			1472
1615	WP_023102062.1		Pseudomonas aeruginosa	Proteobacteria	951	7	Cap4	1			1472
1616	ENY74538.1		Pseudomonas putida TRO1	Proteobacteria	951	7	Cap4	0			
1617	PVZ41015.1		Pseudomonas sp. CC120222-01a	Proteobacteria	951	7	Cap4	0			
1618	WP_191985756.1		Pseudomonas aeruginosa	Proteobacteria	951	7	Cap4	0			
1619	WP_177011105.1		Pseudomonas	Proteobacteria	951	7	Cap4	0			
1620	WP_178083387.1		unclassified Pseudomonas	Proteobacteria	951	7	Cap4	0			
1621	WP_175302142.1		Pseudomonas lundensis	Proteobacteria	951	7	Cap4	0			
1622	WP_122570082.1		Pseudomonas viridiflava	Proteobacteria	952	4	Cap4	1			1468
1623	WP_135196263.1		Pseudomonas fluorescens	Proteobacteria	952	4	Cap4	0			
1624	WP_153649621.1		Pseudomonas sp. CFSAN084952	Proteobacteria	952	4	Cap4	0			

1625	WP_191948583.1		Pseudomonas fluorescens	Proteobacteria	952	4	Cap4	0			
1626	WP_196113741.1		Pseudomonas fulva	Proteobacteria	953	1	Cap4	1			1469
1627	WP_196164972.1		Pseudomonas guariconensis	Proteobacteria	954	1	Cap4	1			1468
1628	WP_186723269.1		Pseudomonas sp. RW1P2	Proteobacteria	955	1	Cap4	1			1466
1629	WP_139176061.1		unclassified Janthinobacterium	Proteobacteria	956	1	Cap4	1			1472
1630	MBN9597938.1		Afipia sp.	Proteobacteria	957	1	Cap4	1			1479
1631	WP_207907326.1		Paucimonas lemoignei	Proteobacteria	958	2	Cap4	1			1478
1632	TCS33100.1		Paucimonas lemoignei	Proteobacteria	958	2	Cap4	0			
1633	WP_212459976.1		Bradyrhizobium manausense	Proteobacteria	959	1	Cap4	1			1482
1634	WP_110818143.1		Pseudomonas syringae	Proteobacteria	960	3	Cap4	1			1473
1635	WP_122456459.1		Pseudomonas viridiflava	Proteobacteria	960	3	Cap4	0			
1636	WP_169858494.1		Pseudomonas lundensis	Proteobacteria	960	3	Cap4	0			
1637	WP_029243653.1		Pseudomonas viridiflava	Proteobacteria	961	1	Cap4	1			1473
1638	WP_122518570.1		Pseudomonas viridiflava	Proteobacteria	962	1	Cap4	1			1473
1639	MBO6548564.1		Rhizobiales bacterium	Proteobacteria	963	1	Cap4	1			1474
1640	WP_028162335.1		Bradyrhizobium elkanii	Proteobacteria	964	2	Cap4	1			1475
1641	WP_130580924.1		Bradyrhizobium sp. Leo170	Proteobacteria	964	2	Cap4	0			
1642	WP_143579275.1		Tardiphaga sp. vice352	Proteobacteria	965	1	Cap4	1			1475
1643	WP_203549321.1		Gluconobacter cerinus	Proteobacteria	966	1	Cap4	1			1474
1644	WP_088521860.1		Rhodoblastus acidophilus	Proteobacteria	967	1	Cap4	1			1476
1645	WP_207587823.1		Rhizobium sp. ZX09	Proteobacteria	968	1	Cap4	1			1477
1646	WP_160870940.1		Shinella sp. AETb1-6	Proteobacteria	969	1	Cap4	1			1476
1647	WP_057834643.1		Bradyrhizobium jicamae	Proteobacteria	970	1	Cap4	1			1476
1648	WP_186194552.1		Burkholderia gladioli	Proteobacteria	971	1	Cap4	1			1474
1649	QET32082.1		Burkholderia multivorans	Proteobacteria	972	3	Cap4	1			1474
1650	WP_190972314.1		Burkholderia multivorans	Proteobacteria	972	3	Cap4	0			
1651	CAB5280108.1		Burkholderia multivorans	Proteobacteria	972	3	Cap4	0			
1652	TAF99113.1		Betaproteobacteria bacterium	Proteobacteria	973	1	Cap4	1			1488
1653	WP_127638394.1		Sinorhizobium meliloti	Proteobacteria	974	1	Cap4	1			1470
1654	RWM76525.1		Mesorhizobium sp.	Proteobacteria	975	1	Cap4	1			1470
1655	PIF74003.1		Variovorax sp. 54	Proteobacteria	976	2	Cap4	1			1451
1656	WP_099791169.1		Variovorax sp. 54	Proteobacteria	976	2	Cap4	0			
1657	WP_182344401.1		Comamonas koreensis	Proteobacteria	977	1	Cap4	1			1455
1658	WP_177213758.1		Variovorax sp. YR750	Proteobacteria	978	1	Cap4	1			1459

1659	WP_198981121.1		Herbaspirillum sp. ASV7	Proteobacteria	979	1	Cap4	1			1459
1660	WP_108301471.1		Limnohabitans sp. JirII-29	Proteobacteria	980	1	Cap4	1			1459
1661	MAG79949.1		Sutterellaceae bacterium	Proteobacteria	981	1	Cap4	1			1459
1662	WP_169802320.1		Hydrogenophaga flava	Proteobacteria	982	1	Cap4	1			1459
1663	WP_057267415.1		Acidovorax sp. Root219	Proteobacteria	983	2	Cap4	1			1459
1664	KRC27513.1		Acidovorax sp. Root219	Proteobacteria	983	2	Cap4	0			
1665	WP_013390847.1		Achromobacter xylosoxidans	Proteobacteria	984	1	Cap4	1			1458
1666	WP_108617525.1		Acidovorax sp. HMWF029	Proteobacteria	985	1	Cap4	1			1458
1667	WP_148193674.1		Pseudomonas syringae	Proteobacteria	986	3	Cap4	1			1459
1668	RMO68834.1		Pseudomonas syringae pv. primulae	Proteobacteria	986	3	Cap4	0			
1669	WP_147464826.1		Pseudomonas syringae group genomosp. 3	Proteobacteria	986	3	Cap4	0			
1670	WP_186598769.1		Pseudomonas taiwanensis	Proteobacteria	987	1	Cap4	1			1462
1671	WP_122575728.1		Pseudomonas viridiflava	Proteobacteria	988	1	Cap4	1			1458
1672	WP_157115953.1		Pseudomonas sp. B10	Proteobacteria	989	3	Cap4	1			1458
1673	WP_170058451.1		Pseudomonas sp. WS 5021	Proteobacteria	989	3	Cap4	0			
1674	SIR22034.1		Pseudomonas sp. B10	Proteobacteria	989	3	Cap4	0			
1675	WP_192345345.1		Pseudomonas sp. PDM09	Proteobacteria	990	1	Cap4	1			1457
1676	AHX16386.1		Dyella jiangningensis	Proteobacteria	991	1	Cap4	1			1457
1677	WP_179476559.1		Rhodanobacter sp. K2T2	Proteobacteria	992	1	Cap4	1			1456
1678	MBN1210300.1		Myxococcaceae bacterium	Proteobacteria	993	1	Cap4	1			1462
1679	RYZ40545.1		Myxococcaceae bacterium	Proteobacteria	994	1	Cap4	1			1466
1680	TAN03267.1		Rhodanobacteraceae bacterium	Proteobacteria	995	1	Cap4	1			1465
1681	WP_169131598.1		Aromatoleum evansii	Proteobacteria	996	1	Cap4	1			1463
1682	KAA5805344.1		Alkalicaulis satelles	Proteobacteria	997	2	Cap4	1			1491
1683	WP_199200455.1		Alkalicaulis satelles	Proteobacteria	997	2	Cap4	0			
1684	WP_091832423.1		Propionispira arboris	Firmicutes	998	1	Cap4	1			1472
1685	WP_139976508.1		Ochrobactrum sp. CGA5	Proteobacteria	999	1	Cap4	1			1516
1686	WP_166601806.1		Rhizobium sp. CRIBSB	Proteobacteria	1000	1	Cap4	1			1511
1687	WP_035034101.1		Aquabacterium sp. NJ1	Proteobacteria	1001	1	Cap4	1			1501
1688	OGB04847.1		Burkholderiales bacterium RIFCSPHIGHO2_12_FULL_6320	Proteobacteria	1002	1	Cap4	1			1505
1689	WP_168106689.1		Ramlibacter lithotrophicus	Proteobacteria	1003	1	Cap4	1			1510
1690	WP_063498192.1		Paraburkholderia phytofirmans	Proteobacteria	1004	1	Cap4	1			1502

1691	WP_111998130.1		Burkholderia cepacia	Proteobacteria	1005	1	Cap4	1			1503
1692	WP_034200522.1		Burkholderia	Proteobacteria	1006	1	Cap4	1			1504
1693	WP_076929124.1		Burkholderia pseudomallei	Proteobacteria	1007	2	Cap4	1			1503
1694	WP_038776393.1		Burkholderia pseudomallei	Proteobacteria	1007	2	Cap4	0			
1695	WP_175787316.1		Burkholderia anthina	Proteobacteria	1008	1	Cap4	1			1503
1696	MBF4009710.1		Burkholderia pseudomallei	Proteobacteria	1009	1	Cap4	1			1503
1697	WP_184856769.1		Acidovorax soli	Proteobacteria	1010	1	Cap4	1			1498
1698	WP_206515131.1		Xanthomonas sp. BRIP62411	Proteobacteria	1011	1	Cap4	1			1500
1699	MBS0557069.1		Proteobacteria bacterium	Proteobacteria	1012	1	Cap4	1			1500
1700	WP_190281010.1		Thermomonas sp. XSG	Proteobacteria	1013	1	Cap4	1			1502
1701	WP_133608090.1		Aquabacterium commune	Proteobacteria	1014	1	Cap4	1			1502
1702	TAM352222.1		Rhodanobacter sp.	Proteobacteria	1015	1	Cap4	1			1502
1703	WP_011798570.1		Polaromonas naphthalenivorans	Proteobacteria	1016	1	Cap4	1			1517
1704	WP_150675894.1		Pseudomonas fluorescens	Proteobacteria	1017	1	Cap4	1			1500
1705	WP_090623950.1		Azotobacter beijerinckii	Proteobacteria	1018	1	Cap4	1			1502
1706	WP_019677666.1		Rheinheimera perlucida	Proteobacteria	1019	1	Cap4	1			1500
1707	WP_169915542.1		Pseudomonas chlororaphis	Proteobacteria	1020	1	Cap4	1			1499
1708	WP_070383454.1		Pseudomonas	Proteobacteria	1021	2	Cap4	1			1504
1709	WP_126525797.1		Pseudomonas aeruginosa	Proteobacteria	1021	2	Cap4	0			
1710	WP_039175673.1		Halomonas hydrothermalis	Proteobacteria	1022	4	Cap4	1			1500
1711	WP_102856795.1		Pseudomonas stutzeri	Proteobacteria	1022	4	Cap4	0			
1712	MBH3941077.1		Pseudomonas aeruginosa	Proteobacteria	1022	4	Cap4	0			
1713	WP_058162051.1		Pseudomonas aeruginosa	Proteobacteria	1022	4	Cap4	0			
1714	WP_058170252.1		Pseudomonas aeruginosa	Proteobacteria	1023	1	Cap4	1			1499
1715	WP_024663354.1		Pseudomonas syringae	Proteobacteria	1024	2	Cap4	1			1497
1716	WP_057436448.1		Pseudomonas syringae group genomosp. 3	Proteobacteria	1024	2	Cap4	0			
1717	WP_104652475.1		Pseudomonas syringae	Proteobacteria	1025	6	Cap4	1			1497
1718	WP_210712935.1		Pseudomonas sp. MWU349	Proteobacteria	1025	6	Cap4	0			
1719	WP_169889516.1		Pseudomonas proteolytica	Proteobacteria	1025	6	Cap4	0			
1720	WP_039590740.1		Pseudomonas frederiksbergensis	Proteobacteria	1025	6	Cap4	0			
1721	WP_177017942.1		Pseudomonas sp. D2002	Proteobacteria	1025	6	Cap4	0			
1722	WP_201230255.1		Pseudomonas sp. TH04	Proteobacteria	1025	6	Cap4	0			
1723	WP_012279545.1		Caulobacter sp. K31	Proteobacteria	1026	1	Cap4	1			1497

1724	WP_056750616.1		Caulobacter sp. Root1472	Proteobacteria	1027	1	Cap4	1			1504
1725	WP_108179135.1		Phreatobacter oligotrophus	Proteobacteria	1028	1	Cap4	1			1505
1726	WP_123656830.1		Salinisphaera japonica	Proteobacteria	1029	1	Cap4	1			1507
1727	WP_186032229.1		Burkholderia gladioli	Proteobacteria	1030	5	Cap4	1			1505
1728	WP_186083557.1		Burkholderia gladioli	Proteobacteria	1030	5	Cap4	0			
1729	WP_186193707.1		Burkholderia gladioli	Proteobacteria	1030	5	Cap4	0			
1730	WP_186183827.1		Burkholderia gladioli	Proteobacteria	1030	5	Cap4	0			
1731	WP_186063252.1		Burkholderia gladioli	Proteobacteria	1030	5	Cap4	0			
1732	WP_185225140.1		Brevundimonas sp. 090558	Proteobacteria	1031	2	Cap4	1			1487
1733	PZU73797.1		Brevundimonas sp.	Proteobacteria	1031	2	Cap4	0			
1734	WP_066628416.1		Brevundimonas vesicularis	Proteobacteria	1032	1	Cap4	1			1496
1735	WP_126005036.1		Sphingomonas koreensis	Proteobacteria	1033	1	Cap4	1			1519
1736	WP_198424461.1		Geomonas limicola	Proteobacteria	1034	2	Cap4	1			1501
1737	GFO67798.1		Geomonas limicola	Proteobacteria	1034	2	Cap4	0			
1738	MBO6506780.1		Kordiimonadaceae bacterium	Proteobacteria	1035	1	Cap4	1			1503
1739	WP_202671363.1		Serratia fonticola	Proteobacteria	1036	1	Cap4	1			1428
1740	WP_050679917.1		Pantoea vagans	Proteobacteria	1037	1	Cap4	1			1429
1741	WP_068910624.1		Acinetobacter sp. WCHAc010034	Proteobacteria	1038	1	Cap4	1			1471
1742	WP_151741360.1		Acinetobacter seifertii	Proteobacteria	1039	1	Cap4	1			1470
1743	WP_134292550.1		Psychrobacter sp. 230	Proteobacteria	1040	1	Cap4	1			1466
1744	WP_096064541.1		Psychrobacter sp. FDAARGOS_221	Proteobacteria	1041	1	Cap4	1			1490
1745	WP_201551902.1		Psychrobacter fjordensis	Proteobacteria	1042	1	MBL + serine protease	1			1485
1746	WP_201613246.1		Psychrobacter immobilis	Proteobacteria	1043	1	MBL + serine protease	1			1486
1747	WP_087550346.1		Acinetobacter sp. WCHAc010052	Proteobacteria	1044	1	MBL + serine protease	1			1485
1748	KHF77843.1		Acinetobacter sp. neg1	Proteobacteria	1045	2	MBL + serine protease	1			1457
1749	WP_047429121.1		Acinetobacter sp. neg1	Proteobacteria	1045	2	MBL + serine protease	0			
1750	SNX45784.1		Acinetobacter puyangensis	Proteobacteria	1046	1	MBL + serine protease	1			1473
1751	WP_100244472.1		Acinetobacter baumannii	Proteobacteria	1047	1	MBL + serine protease	1			1458
1752	WP_155758092.1		Acinetobacter sp. OIFC021	Proteobacteria	1048	3	MBL + serine protease	1			1459

1753	WP_033855729.1		Acinetobacter calcoaceticus/baumannii complex	Proteobacteria	1048	3	MBL + serine protease	0			
1754	ELW83155.1		Acinetobacter sp. OIFC021	Proteobacteria	1048	3	MBL + serine protease	0			
1755	WP_121929437.1		Acinetobacter sp. 2JN-4	Proteobacteria	1049	1	MBL + serine protease	1			1461
1756	WP_168159176.1		Acinetobacter johnsonii	Proteobacteria	1050	1	MBL + serine protease	1			1486
1757	WP_178319542.1		Acinetobacter	Proteobacteria	1051	1	MBL + serine protease	1			1461
1758	WP_058602075.1		Aureimonas ureilytica	Proteobacteria	1052	1	Cap4	1			1506
1759	WP_205206706.1		Asaia sp. As-1742	Proteobacteria	1053	2	Cap4	1			1500
1760	NIE79518.1		Asaia sp. As-1742	Proteobacteria	1053	2	Cap4	0			
1761	WP_080924336.1		Klebsiella quasivariicola	Proteobacteria	1054	8	Unknown-4	1			1489
1762	WP_087529028.1		Klebsiella pneumoniae	Proteobacteria	1054	8	Unknown-4	0			
1763	WP_131635746.1		Enterobacter wuhouensis	Proteobacteria	1054	8	Unknown-4	0			
1764	WP_049091014.1		Klebsiella oxytoca	Proteobacteria	1054	8	Unknown-4	0			
1765	WP_202768595.1		Enterobacter hormaechei	Proteobacteria	1054	8	Unknown-4	0			
1766	EFE6301287.1		Escherichia coli	Proteobacteria	1054	8	Unknown-4	0			
1767	MBS5210931.1		Klebsiella sp.	Proteobacteria	1054	8	Unknown-4	0			
1768	WP_200094311.1		Citrobacter freundii	Proteobacteria	1054	8	Unknown-4	0			
1769	OXS29038.1		Desulfovibrio sp. MES5	Proteobacteria	1055	1	Cap4	1			1504
1770	NQY16085.1		Alteromonas sp.	Proteobacteria	1056	1	Cap4	1			1493
1771	WP_068902014.1		Enterovibrio pacificus	Proteobacteria	1057	1	Cap4	1			1498
1772	WP_039467082.1		Vibrio vulnificus	Proteobacteria	1058	5	Cap4	1			1506
1773	WP_095429676.1		Vibrio vulnificus	Proteobacteria	1058	5	Cap4	0			
1774	EGQ8000088.1		Vibrio vulnificus	Proteobacteria	1058	5	Cap4	0			
1775	WP_095666370.1		Vibrio	Proteobacteria	1058	5	Cap4	0			
1776	WP_102466533.1		Vibrio splendidus	Proteobacteria	1058	5	Cap4	0			
1777	WP_108167906.1		Vibrio	Proteobacteria	1059	5	Cap4	1			1500
1778	WP_198593706.1		Vibrio splendidus	Proteobacteria	1059	5	Cap4	0			
1779	WP_102451005.1		Vibrio splendidus	Proteobacteria	1059	5	Cap4	0			
1780	WP_171308816.1		Vibrio splendidus	Proteobacteria	1059	5	Cap4	0			
1781	WP_210442783.1		Vibrio crassostreae	Proteobacteria	1059	5	Cap4	0			
1782	OYZ35129.1		Campylobacteraleses bacterium 16-40-21	Proteobacteria	1060	1	Cap4	1			1488
1783	WP_012465104.1		Chlorobium limicola	Chlorobi	1061	1	Cap4	1			1499
1784	WP_207765186.1		Pseudomonas stutzeri	Proteobacteria	1062	1	Unknown-4	1			1497

1785	WP_198044916.1		Pseudomonas frederiksbergensis	Proteobacteria	1063	1	Unknown-4	1			1497
1786	WP_115540356.1		Xanthomonas campestris	Proteobacteria	1064	1	Cap4	1			1493
1787	WP_010340396.1		Xanthomonas sacchari	Proteobacteria	1065	1	Cap4	1			1493
1788	VUJ87838.1		Stenotrophomonas maltophilia	Proteobacteria	1066	1	Cap4	1			1493
1789	WP_029554377.1		Verminephrobacter aporrectodeae	Proteobacteria	1067	1	Cap4	1			1508
1790	MBU2645209.1		bacterium	Unclassified	1068	1	Cap4	1			1492
1791	WP_047243867.1		Chromobacterium subtsugae	Proteobacteria	1069	1	Cap4	1			1498
1792	WP_101324799.1		Aeromonas sobria	Proteobacteria	1070	1	Cap4	1			1496
1793	WP_043555732.1		Aeromonas bestiarum	Proteobacteria	1071	1	Cap4	1			1496
1794	NQY37295.1		Alteromonadaceae bacterium	Proteobacteria	1072	1	Cap4	1			1506
1795	WP_148066033.1		Parahaliea aestuarii	Proteobacteria	1073	2	Cap4	1			1500
1796	WP_153844092.1		Halomonas piezotolerans	Proteobacteria	1073	2	Cap4	0			
1797	PLY14975.1		Sedimenticola sp.	Proteobacteria	1074	1	Cap4	1			1500
1798	RDH87718.1		endosymbiont of Escarpia spicata	Proteobacteria	1075	1	Cap4	1			1500
1799	WP_197019021.1		Thioalkalivibrio sp. HK1	Proteobacteria	1076	1	Cap4	1			1484
1800	WP_165787044.1		Halioglobus lutimaris	Proteobacteria	1077	1	Cap4	1			1500
1801	WP_090287471.1		Nitrosomonas mobilis	Proteobacteria	1078	1	Cap4	1			1500
1802	MBK9130386.1		Gammaproteobacteria bacterium	Proteobacteria	1079	1	Cap4	1			1500
1803	WP_090413306.1		Nitrosomonas halophila	Proteobacteria	1080	1	Cap4	1			1500
1804	MBU1224396.1		Gammaproteobacteria bacterium	Proteobacteria	1081	1	Cap4	1			1500
1805	WP_195883302.1		Pseudomonas mendocina	Proteobacteria	1082	1	Cap4	1			1501
1806	WP_014028321.1		Acidithiobacillus ferrivorans	Proteobacteria	1083	1	Cap4	1			1488
1807	WP_073434282.1		Halomonas cupida	Proteobacteria	1084	2	Cap4	1			1499
1808	WP_179930019.1		Halomonas salicampi	Proteobacteria	1084	2	Cap4	0			
1809	WP_101145383.1		Halomonas sp. MES3-P3E	Proteobacteria	1085	1	Cap4	1			1499
1810	KQR76938.1		Burkholderia sp. Leaf177	Proteobacteria	1086	1	Cap4	1			1500
1811	KVO58157.1		Burkholderia stagnalis	Proteobacteria	1087	1	Cap4	1			1500
1812	SAK89708.1		Caballeronia arationis	Proteobacteria	1088	1	Cap4	1			1500
1813	KVC24504.1		Burkholderia pseudomultivorans	Proteobacteria	1089	1	Cap4	1			1500

See also Data S7 of Gao et al., Science 377, eabm4096 (2022), which is incorporated herein by reference as if expressed in its entirety herein.
A- Cluster Size
B – Architecture
C – Representative
D - Gene 2 Accession
E - Gene 2 Length
F - Length exclude-ng NTD

Other Exemplary Domains

[0101] In some embodiments, the engineered protein composition comprises one or more other domains. In some embodiments the one or more additional domains are in the N-terminal region, C-terminal region, or both of the engineered protein of the present invention. In some embodiments, the one or more additional domains are contained between the N-terminal region and the C-terminal region of the engineered protein. In some embodiments, the engineered protein of the present invention contains one or more structural motifs or interaction domains. Exemplary structural motifs and/or interaction domains include, without limitation, a TPR domain, a dimerization domain, an oligomerization domain, a signaling domain, and/or the like. Exemplary dimerization domains include, without limitation, zinc finger domains and leucine zipper domains. In some embodiments, one or more of the domains of the engineered proteins of the present invention allow interaction with other proteins, including by not limited to engineered proteins of the present invention, but others as well, such as those present on target cells and engage in cell signaling.

Engineered Protein Oligomers

[0102] In some embodiments, the engineered proteins of the present invention form oligomers. In some embodiments, effector activity and/or activation of one or more engineered proteins includes oligomer formation. Without being bound by theory in these embodiments, activation occurs upon oligomer formation. In some embodiments, oligomer formation involves binding of a pattern recognition domain to a target polypeptide. In some embodiments, the oligomer is a tetramer, a trimer, or a dimer. In some embodiments, the oligomer is heterogeneous (i.e., contains at least two different engineered protein monomers). In some embodiments, at least two engineered protein monomers are different. In some embodiments, each engineered protein monomer is different. In some embodiments, the at least two different engineered protein monomers have different effector domains. In some embodiments, the oligomer is homogenous (i.e., contains all the same engineered protein monomers).

POLYNUCLEOTIDES AND VECTORS

[0103] Described herein are polynucleotides encoding one or more components (e.g., polypeptides and/or guide polynucleotides) of the programmable pattern recognition proteins, oligomers, or system (such as a detection composition or system) comprising the programmable pattern recognition composition. Also described herein are vectors and vector

systems containing one or more programmable pattern recognition protein or system encoding polynucleotides. As used herein with reference to the relationship between DNA, cDNA, cRNA, RNA, protein/peptides, and the like "corresponding to" or "encoding" (used interchangeably herein) refers to the underlying biological relationship between these different molecules. As such, one of skill in the art would understand that operatively "corresponding to" can direct them to determine the possible underlying and/or resulting sequences of other molecules given the sequence of any other molecule which has a similar biological relationship with these molecules. For example, from a DNA sequence an RNA sequence can be determined and from an RNA sequence a cDNA sequence can be determined.

Polynucleotides

[0104] As used herein, "nucleic acid," "nucleotide sequence," and "polynucleotide" can be used interchangeably herein and can generally refer to a string of at least two base-sugar-phosphate combinations and refers to, among others, single-and double-stranded DNA, DNA that is a mixture of single-and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, polynucleotide as used herein can refer to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions can be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. "Polynucleotide" and "nucleic acids" also encompasses such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including simple and complex cells, inter alia. For instance, the term polynucleotide as used herein can include DNAs or RNAs as described herein that contain one or more modified bases. Thus, DNAs or RNAs including unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. "Polynucleotide", "nucleotide sequences" and "nucleic acids" also includes PNAs (peptide nucleic acids), phosphorothioates, and other variants of the phosphate backbone of native nucleic acids. Natural nucleic acids have a phosphate backbone, artificial nucleic acids can contain other types of backbones, but contain the same bases. Thus, DNAs or RNAs with backbones modified for stability or for other

reasons are “nucleic acids” or "polynucleotides" as that term is intended herein. As used herein, “nucleic acid sequence” and “oligonucleotide” also encompasses a nucleic acid and polynucleotide as defined elsewhere herein.

Codon Optimization

[0105] In some embodiments, the polynucleotide can be codon optimized. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g., about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the “Codon Usage Database” available at www.kazusa.orjp/codon/ and these tables can be adapted in a number of ways. See Nakamura, Y., et al. “Codon usage tabulated from the international DNA sequence databases: status for the year 2000” Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, PA), are also available. In some embodiments, one or more codons (e.g., 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a DNA/RNA-targeting Cas protein corresponds to the most frequently used codon for a particular amino acid. As to codon usage in yeast, reference is made to the online Yeast Genome database available at http://www.yeastgenome.org/community/codon_usage.shtml, or *Codon selection in yeast*, Bennetzen and Hall, J Biol Chem. 1982 Mar 25;257(6):3026-31. As to codon usage in plants including algae, reference is made to *Codon usage in higher plants, green algae, and cyanobacteria*, Campbell and Gowri, Plant Physiol. 1990 Jan; 92(1): 1–11.; as well as *Codon usage in plant genes*, Murray et al, Nucleic Acids Res. 1989 Jan 25;17(2):477-98; or *Selection*

on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages, Morton BR, J Mol Evol. 1998 Apr;46(4):449-59.

[0106] The polynucleotide can be codon optimized for expression in a specific cell-type, tissue type, organ type, and/or subject type. In some embodiments, a codon optimized sequence is a sequence optimized for expression in a eukaryote, e.g., humans (i.e. being optimized for expression in a human or human cell), or for another eukaryote, such as another animal (e.g. a mammal or avian) as is described elsewhere herein. Such codon optimized sequences are within the ambit of the ordinary skilled artisan in view of the description herein. In some embodiments, the polynucleotide is codon optimized for a specific cell type. Such cell types can include, but are not limited to, epithelial cells (including skin cells, cells lining the gastrointestinal tract, cells lining other hollow organs), nerve cells (nerves, brain cells, spinal column cells, nerve support cells (e.g. astrocytes, glial cells, Schwann cells etc.) , muscle cells (e.g. cardiac muscle, smooth muscle cells, and skeletal muscle cells), connective tissue cells (fat and other soft tissue padding cells, bone cells, tendon cells, cartilage cells), blood cells, stem cells and other progenitor cells, immune system cells, germ cells, and combinations thereof. Such codon optimized sequences are within the ambit of the ordinary skilled artisan in view of the description herein. In some embodiments, the polynucleotide is codon optimized for a specific tissue type. Such tissue types can include, but are not limited to, muscle tissue, connective tissue, connective tissue, nervous tissue, and epithelial tissue. Such codon optimized sequences are within the ambit of the ordinary skilled artisan in view of the description herein. In some embodiments, the polynucleotide is codon optimized for a specific organ. Such organs include, but are not limited to, muscles, skin, intestines, liver, spleen, brain, lungs, stomach, heart, kidneys, gallbladder, pancreas, bladder, thyroid, bone, blood vessels, blood, and combinations thereof. Such codon optimized sequences are within the ambit of the ordinary skilled artisan in view of the description herein.

[0107] In some embodiments, a polynucleotide coding sequence encoding one or more elements of programmable pattern recognition proteins or system described herein is codon optimized for expression in particular cells, such as prokaryotic or eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a plant or a mammal, including but not limited to human, or non-human eukaryote or animal or mammal as herein discussed, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate.

Vectors and Vector Systems

[0108] Also provided herein are vectors and vector system that can contain one or more of the programmable pattern recognition protein or system polynucleotides (such as an encoding polynucleotide) described herein. In certain embodiments, the vector can contain one or more polynucleotides encoding one or more elements of a CRISPR-Cas system described herein. The vectors can be useful in producing bacterial, fungal, yeast, plant cells, animal cells, and transgenic animals that can express one or more components of the programmable pattern recognition protein or system described herein. Within the scope of this disclosure are vectors containing one or more of the polynucleotide sequences described herein. One or more of the polynucleotides that are part of the programmable pattern recognition protein or system described herein can be included in a vector or vector system. The vectors and/or vector systems can be used, for example, to express one or more of the polynucleotides in a cell, such as a producer cell, to produce programmable pattern recognition protein or system containing virus particles described elsewhere herein. Other uses for the vectors and vector systems described herein are also within the scope of this disclosure. In general, and throughout this specification, the term "vector" refers to a tool that allows or facilitates the transfer of an entity from one environment to another. In some contexts which will be appreciated by those of ordinary skill in the art, "vector" can be a term of art to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. A vector can be a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elements.

[0109] Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g., circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g., retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses (AAVs)). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain

vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[0110] Recombinant expression vectors can be composed of a nucleic acid (e.g., a polynucleotide) of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which can be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" and "operatively-linked" are used interchangeably herein and further defined elsewhere herein. In the context of a vector, the term "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g., in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell). Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells. These and other embodiments of the vectors and vector systems are described elsewhere herein.

[0111] In some embodiments, the vector can be a bicistronic vector. In some embodiments, a bicistronic vector can be used for one or more elements of the programmable pattern recognition protein or system described herein. In some embodiments, expression of elements of the programmable pattern recognition protein or system described herein can be driven by the CBh promoter or other ubiquitous promoter. Where the element of the programmable pattern recognition protein or system is an RNA, its expression can be driven by a Pol III promoter, such as a U6 promoter. In some embodiments, the two are combined.

[0112] In some embodiments, a vector capable of delivering an effector protein and optionally at least one guide RNA to a cell can be composed of or contain a minimal promoter operably linked to a polynucleotide sequence encoding the effector protein and a second minimal promoter operably linked to a polynucleotide sequence encoding at least one guide

RNA, wherein the length of the vector sequence comprising the minimal promoters and polynucleotide sequences is less than 4.4Kb. In an embodiment, the vector can be a viral vector. In certain embodiments, the viral vector is an is an adeno-associated virus (AAV) or an adenovirus vector.

[0113] In some embodiments, the vector capable of delivering a lentiviral vector for an effector protein and at least one guide RNA to a cell can be composed of or contain a promoter operably linked to a polynucleotide sequence encoding a STAND NTPase, a target containing a pattern recognized by the STAND NTPase, an effector and a second promoter operably linked to a polynucleotide sequence encoding at least one guide RNA, wherein the polynucleotide sequences are in reverse orientation.

[0114] In one embodiment, the invention provides a vector system comprising one or more vectors. In some embodiments, the system comprises: (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences up- or downstream (whichever applicable) of the direct repeat sequence, wherein when expressed, the one or more guide sequence(s) direct(s) sequence-specific binding of the programmable pattern recognition protein or system complex to the one or more target sequence(s) in a eukaryotic cell, wherein the programmable pattern recognition protein or system complex comprises a STAND NTPase polypeptide and/or effector polypeptide complexed with the one or more guide sequence(s) that is hybridized to the one or more target sequence(s); and (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said STAND NTPase polypeptide and/or effector polypeptide, preferably comprising at least one nuclear localization sequence and/or at least one NES; wherein components (a) and (b) are located on the same or different vectors of the system. Where applicable, a tracr sequence may also be provided. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a programmable pattern recognition protein or system complex to a different target sequence in a eukaryotic cell. In some embodiments, the programmable pattern recognition protein or system complex comprises one or more nuclear localization sequences and/or one or more NES of sufficient strength to drive accumulation of said programmable pattern recognition protein or system complex in a detectable amount in or out of the nucleus of a eukaryotic cell. In some embodiments, the first regulatory element is a polymerase III

promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, each of the guide sequences is at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length.

[0115] These and others are further detailed and described elsewhere herein.

Cell-based Vector Amplification and Expression

[0116] Vectors may be introduced and propagated in a prokaryote or prokaryotic cell. In some embodiments, a prokaryote is used to amplify copies of a vector to be introduced into a eukaryotic cell or as an intermediate vector in the production of a vector to be introduced into a eukaryotic cell (e.g., amplifying a plasmid as part of a viral vector packaging system). The vectors can be viral-based or non-viral based. In some embodiments, a prokaryote is used to amplify copies of a vector and express one or more nucleic acids, such as to provide a source of one or more proteins for delivery to a host cell or host organism.

[0117] Vectors can be designed for expression of one or more elements of the programmable pattern recognition protein or system described herein (e.g., nucleic acid transcripts, proteins, enzymes, and combinations thereof) in a suitable host cell. In some embodiments, the suitable host cell is a prokaryotic cell. Suitable host cells include, but are not limited to, bacterial cells, yeast cells, insect cells, and mammalian cells. In some embodiments, the suitable host cell is a eukaryotic cell.

[0118] In some embodiments, the suitable host cell is a suitable bacterial cell. Suitable bacterial cells include, but are not limited to, bacterial cells from the bacteria of the species Escherichia coli. Many suitable strains of E. coli are known in the art for expression of vectors. These include, but are not limited to Pir1, Stbl2, Stbl3, Stbl4, TOP10, XL1 Blue, and XL10 Gold. In some embodiments, the host cell is a suitable insect cell. Suitable insect cells include those from Spodoptera frugiperda. Suitable strains of S. frugiperda cells include, but are not limited to Sf9 and Sf21. In some embodiments, the host cell is a suitable yeast cell. In some embodiments, the yeast cell can be from Saccharomyces cerevisiae. In some embodiments, the host cell is a suitable mammalian cell. Many types of mammalian cells have been developed to express vectors. Suitable mammalian cells include, but are not limited to, HEK293, Chinese Hamster Ovary Cells (CHOs), mouse myeloma cells, HeLa, U2OS, A549, HT1080, CAD, P19, NIH 3T3, L929, N2a, MCF-7, Y79, SO-Rb50, HepG G2, DIKX-X11, J558L, Baby hamster kidney cells (BHK), and chicken embryo fibroblasts (CEFs). Suitable host cells are discussed

further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990).

[0119] In some embodiments, the vector can be a yeast expression vector. Examples of vectors for expression in yeast Saccharomyces cerevisiae include pYepSec1 (Baldari, et al., 1987. EMBO J. 6: 229-234), pMFa (Kuijan and Herskowitz, 1982. Cell 30: 933-943), pJRY88 (Schultz et al., 1987. Gene 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.). As used herein, a "yeast expression vector" refers to a nucleic acid that contains one or more sequences encoding an RNA and/or polypeptide and may further contain any desired elements that control the expression of the nucleic acid(s), as well as any elements that enable the replication and maintenance of the expression vector inside the yeast cell. Many suitable yeast expression vectors and features thereof are known in the art; for example, various vectors and techniques are illustrated in in Yeast Protocols, 2nd edition, Xiao, W., ed. (Humana Press, New York, 2007) and Buckholz, R.G. and Gleeson, M.A. (1991) Biotechnology (NY) 9(11): 1067-72. Yeast vectors can contain, without limitation, a centromeric (CEN) sequence, an autonomous replication sequence (ARS), a promoter, such as an RNA Polymerase III promoter, operably linked to a sequence or gene of interest, a terminator such as an RNA polymerase III terminator, an origin of replication, and a marker gene (e.g., auxotrophic, antibiotic, or other selectable markers). Examples of expression vectors for use in yeast may include plasmids, yeast artificial chromosomes, 2μ plasmids, yeast integrative plasmids, yeast replicative plasmids, shuttle vectors, and episomal plasmids.

[0120] In some embodiments, the vector is a baculovirus vector or expression vector and can be suitable for expression of polynucleotides and/or proteins in insect cells. In some embodiments, the suitable host cell is an insect cell. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., SF9 cells) include the pAc series (Smith, et al., 1983. Mol. Cell. Biol. 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. Virology 170: 31-39). rAAV (recombinant Adeno-associated viral) vectors are preferably produced in insect cells, e.g., Spodoptera frugiperda Sf9 insect cells, grown in serum-free suspension culture. Serum-free insect cells can be purchased from commercial vendors, e.g., Sigma Aldrich (EX-CELL 405).

[0121] In some embodiments, the vector is a mammalian expression vector. In some embodiments, the mammalian expression vector is capable of expressing one or more

polynucleotides and/or polypeptides in a mammalian cell. Examples of mammalian expression vectors include, but are not limited to, pCDM8 (Seed, 1987. Nature 329: 840) and pMT2PC (Kaufman, et al., 1987. EMBO J. 6: 187-195). The mammalian expression vector can include one or more suitable regulatory elements capable of controlling expression of the one or more polynucleotides and/or proteins in the mammalian cell. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, simian virus 40, and others disclosed herein and known in the art. More detail on suitable regulatory elements are described elsewhere herein.

[0122] For other suitable expression vectors and vector systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook, et al., MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

[0123] In some embodiments, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, et al., 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Baneiji, et al., 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, et al., 1985. *Science* 230: 912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546). With regards to these prokaryotic and eukaryotic vectors, mention is made of U.S. Patent 6,750,059, the contents of which are incorporated by reference herein in their entirety. Other embodiments can utilize viral vectors, with regards to which mention is made of U.S. Patent application 13/092,085, the contents of which are incorporated by reference herein in their entirety. Tissue-specific regulatory elements are known in the art and in this regard, mention is made of U.S. Patent 7,776,321, the contents of

which are incorporated by reference herein in their entirety. In some embodiments, a regulatory element can be operably linked to one or more elements of a CRISPR-Cas system so as to drive expression of the one or more elements of the CRISPR-Cas system described herein.

[0124] In some embodiments, the vector can be a fusion vector or fusion expression vector. In some embodiments, fusion vectors add a number of amino acids to a protein encoded therein, such as to the amino terminus, carboxy terminus, or both of a recombinant protein. Such fusion vectors can serve one or more purposes, such as: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. In some embodiments, expression of polynucleotides (such as non-coding polynucleotides) and proteins in prokaryotes can be carried out in Escherichia coli with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion polynucleotides and/or proteins. In some embodiments, the fusion expression vector can include a proteolytic cleavage site, which can be introduced at the junction of the fusion vector backbone or other fusion moiety and the recombinant polynucleotide or protein to enable separation of the recombinant polynucleotide or protein from the fusion vector backbone or other fusion moiety subsequent to purification of the fusion polynucleotide or protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Example fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion E. coli expression vectors include pTrc (Amrann et al., (1988) Gene 69:301-315) and pET 11d (Studier et al., GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

[0125] In some embodiments, one or more vectors driving expression of one or more elements of a programmable pattern recognition proteins or system described herein are introduced into a host cell such that expression of the elements of the engineered delivery system described herein direct formation a programmable pattern recognition protein or system complex at one or more target sites. For example, a programmable pattern recognition protein or system effector protein describe herein and a nucleic acid component (e.g., a guide polynucleotide) can each be operably linked to separate regulatory elements on separate

vectors. RNA(s) of different elements of programmable pattern recognition protein or system described herein can be delivered to an animal, plant, microorganism or cell thereof to produce an animal (e.g., a mammal, reptile, avian, etc.), plant, microorganism or cell thereof that constitutively, inducibly, or conditionally expresses different elements of the programmable pattern recognition protein or system described herein that incorporates one or more elements of the programmable pattern recognition protein or system described herein or contains one or more cells that incorporates and/or expresses one or more elements of the programmable pattern recognition protein or system described herein.

[0126] In some embodiments, two or more of the elements expressed from the same or different regulatory element(s), can be combined in a single vector, with one or more additional vectors providing any components of the system not included in the first vector. In some embodiments, the specific regulator elements used are chosen to reduce or eliminate regulatory element competition, such as promoter competition. Programmable pattern recognition protein or system polynucleotides that are combined in a single vector may be arranged in any suitable orientation, such as one element located 5' with respect to ("upstream" of) or 3' with respect to ("downstream" of) a second element. The coding sequence of one element may be located on the same or opposite strand of the coding sequence of a second element, and oriented in the same or opposite direction. In some embodiments, a single promoter drives expression of a transcript encoding one or more programmable pattern recognition protein or system proteins, embedded within one or more intron sequences (e.g., each in a different intron, two or more in at least one intron, or all in a single intron). In some embodiments, the programmable pattern recognition protein or system polynucleotides can be operably linked to and expressed from the same promoter.

Cell-Free Vector and Polynucleotide Expression

[0127] In some embodiments, the polynucleotide encoding one or more features of the programmable pattern recognition protein or system can be expressed from a vector or suitable polynucleotide in a cell-free in vitro system. In other words, the polynucleotide can be transcribed and optionally translated in vitro. In vitro transcription/translation systems and appropriate vectors are generally known in the art and commercially available. Generally, in vitro transcription and in vitro translation systems replicate the processes of RNA and protein synthesis, respectively, outside of the cellular environment. Vectors and suitable polynucleotides for in vitro transcription can include T7, SP6, T3, promoter regulatory

sequences that can be recognized and acted upon by an appropriate polymerase to transcribe the polynucleotide or vector.

[0128] In vitro translation can be stand-alone (e.g., translation of a purified polyribonucleotide) or linked/coupled to transcription. In some embodiments, the cell-free (or in vitro) translation system can include extracts from rabbit reticulocytes, wheat germ, and/or E. coli. The extracts can include various macromolecular components that are needed for translation of exogenous RNA (e.g., 70S or 80S ribosomes, tRNAs, aminoacyl-tRNA, synthetases, initiation, elongation factors, termination factors, etc.). Other components can be included or added during the translation reaction, including but not limited to, amino acids, energy sources (ATP, GTP), energy regenerating systems (creatine phosphate and creatine phosphokinase (eukaryotic systems)) (phosphoenol pyruvate and pyruvate kinase for bacterial systems), and other co-factors (Mg2+, K+, etc.). As previously mentioned, in vitro translation can be based on RNA or DNA starting material. Some translation systems can utilize an RNA template as starting material (e.g., reticulocyte lysates and wheat germ extracts). Some translation systems can utilize a DNA template as a starting material (e.g., E coli-based systems). In these systems transcription and translation are coupled and DNA is first transcribed into RNA, which is subsequently translated. Suitable standard and coupled cell-free translation systems are generally known in the art and are commercially available.

Vector Features

[0129] The vectors can include additional features that can confer one or more functionalities to the vector, the polynucleotide to be delivered, a virus particle produced there from, or polypeptide expressed thereof. Such features include, but are not limited to, regulatory elements, selectable markers, molecular identifiers (e.g., molecular barcodes), stabilizing elements, and the like. It will be appreciated by those skilled in the art that the design of the expression vector and additional features included can depend on such factors as the choice of the host cell to be transformed, the level of expression desired, etc.

Regulatory Elements

[0130] In certain embodiments, the polynucleotides and/or vectors thereof described herein (such as the programmable pattern recognition protein or system polynucleotides of the present invention) can include one or more regulatory elements that can be operatively linked to the polynucleotide. The term “regulatory element” is intended to include promoters, enhancers, internal ribosomal entry sites (IRES), other expression control elements (e.g., transcription

termination signals, such as polyadenylation signals and poly-U sequences) and cellular localization signals (e.g., nuclear localization signals). Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter can direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g., liver, pancreas), or particular cell types (e.g., lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific. In some embodiments, a vector comprises one or more pol III promoter (e.g., 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g., 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g., 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer) (see, e.g., Boshart et al, Cell, 41:521-530 (1985)), the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter. Also encompassed by the term "regulatory element" are enhancer elements, such as WPRE; CMV enhancers; the R-U5' segment in LTR of HTLV-I (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β-globin (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981).

[0131] In some embodiments, the regulatory sequence can be a regulatory sequence described in U.S. Pat. No. 7,776,321, U.S. Pat. Pub. No. 2011/0027239, and International Patent Publication No. WO 2011/028929, the contents of which are incorporated by reference herein in their entirety. In some embodiments, the vector can contain a minimal promoter. In some embodiments, the minimal promoter is the Mecp2 promoter, tRNA promoter, or U6. In a further embodiment, the minimal promoter is tissue specific. In some embodiments, the length of the vector polynucleotide the minimal promoters and polynucleotide sequences is less than 4.4Kb.

[0132] To express a polynucleotide, the vector can include one or more transcriptional and/or translational initiation regulatory sequences, *e.g.*, promoters, that direct the transcription of the gene and/or translation of the encoded protein in a cell. In some embodiments a constitutive promoter may be employed. Suitable constitutive promoters for mammalian cells are generally known in the art and include, but are not limited to SV40, CAG, CMV, EF-1α, β-actin, RSV, and PGK. Suitable constitutive promoters for bacterial cells, yeast cells, and fungal cells are generally known in the art, such as a T-7 promoter for bacterial expression and an alcohol dehydrogenase promoter for expression in yeast.

[0133] In some embodiments, the regulatory element can be a regulated promoter. "Regulated promoter" refers to promoters that direct gene expression not constitutively, but in a temporally- and/or spatially-regulated manner, and includes tissue-specific, tissue-preferred and inducible promoters. Regulated promoters include conditional promoters and inducible promoters. In some embodiments, conditional promoters can be employed to direct expression of a polynucleotide in a specific cell type, under certain environmental conditions, and/or during a specific state of development. Suitable tissue specific promoters can include, but are not limited to, liver specific promoters (e.g., APOA2, SERPIN A1 (hAAT), CYP3A4, and MIR122), pancreatic cell promoters (e.g., INS, IRS2, Pdx1, Alx3, Ppy), cardiac specific promoters (e.g., Myh6 (alpha MHC), MYL2 (MLC-2v), TNI3 (cTnl), NPPA (ANF), Slc8a1 (Ncx1)), central nervous system cell promoters (SYN1, GFAP, INA, NES, MOBP, MBP, TH, FOXA2 (HNF3 beta)), skin cell specific promoters (e.g., FLG, K14, TGM3), immune cell specific promoters, (e.g., ITGAM, CD43 promoter, CD14 promoter, CD45 promoter, CD68 promoter), urogenital cell specific promoters (e.g., Pbsn, Upk2, Sbp, Fer1l4), endothelial cell specific promoters (e.g., ENG), pluripotent and embryonic germ layer cell specific promoters (e.g., Oct4, NANOG, Synthetic Oct4, T brachyury, NES, SOX17, FOXA2, MIR122), and muscle cell specific promoter (e.g., Desmin). Other tissue and/or cell specific promoters are generally known in the art and are within the scope of this disclosure.

[0134] Inducible/conditional promoters can be positively inducible/conditional promoters (e.g., a promoter that activates transcription of the polynucleotide upon appropriate interaction with an activated activator, or an inducer (compound, environmental condition, or other stimulus) or a negative/conditional inducible promoter (e.g., a promoter that is repressed (e.g., bound by a repressor) until the repressor condition of the promotor is removed (e.g., inducer binds a repressor bound to the promoter stimulating release of the promoter by the repressor or

removal of a chemical repressor from the promoter environment). The inducer can be a compound, environmental condition, or other stimulus. Thus, inducible/conditional promoters can be responsive to any suitable stimuli such as chemical, biological, or other molecular agents, temperature, light, and/or pH. Suitable inducible/conditional promoters include, but are not limited to, Tet-On, Tet-Off, Lac promoter, pBad, AlcA, LexA, Hsp70 promoter, Hsp90 promoter, pDawn, XVE/OlexA, GVG, and pOp/LhGR.

[0135] Where expression in a plant cell is desired, the components of the CRISPR-Cas system described herein are typically placed under control of a plant promoter, i.e., a promoter operable in plant cells. The use of different types of promoters is envisaged.

[0136] A constitutive plant promoter is a promoter that is able to express the open reading frame (ORF) that it controls in all or nearly all of the plant tissues during all or nearly all developmental stages of the plant (referred to as "constitutive expression"). One non-limiting example of a constitutive promoter is the cauliflower mosaic virus 35S promoter. Different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. In particular embodiments, one or more of the programmable pattern recognition protein or system components are expressed under the control of a constitutive promoter, such as the cauliflower mosaic virus 35S promoter issue-preferred promoters can be utilized to target enhanced expression in certain cell types within a particular plant tissue, for instance vascular cells in leaves or roots or in specific cells of the seed. Examples of particular promoters for use in the programmable pattern recognition protein or system are found in Kawamata et al., (1997) Plant Cell Physiol 38:792-803; Yamamoto et al., (1997) Plant J 12:255-65; Hire et al, (1992) Plant Mol Biol 20:207-18,Kuster et al, (1995) Plant Mol Biol 29:759-72, and Capana et al., (1994) Plant Mol Biol 25:681 -91.

[0137] Examples of promoters that are inducible and that can allow for spatiotemporal control of gene editing or gene expression may use a form of energy. The form of energy may include but is not limited to sound energy, electromagnetic radiation, chemical energy and/or thermal energy. Examples of inducible systems include tetracycline inducible promoters (Tet-On or Tet-Off), small molecule two-hybrid transcription activations systems (FKBP, ABA, etc.), or light inducible systems (Phytochrome, LOV domains, or cryptochrome)., such as a Light Inducible Transcriptional Effector (LITE) that direct changes in transcriptional activity in a sequence-specific manner. The components of a light inducible system may include one

or more elements of the programmable pattern recognition protein or system described herein, a light-responsive cytochrome heterodimer (e.g., from Arabidopsis thaliana), and a transcriptional activation/repression domain. In some embodiments, the vector can include one or more of the inducible DNA binding proteins provided in International Patent Publication No. WO 2014/018423 and US Patent Publication Nos., 2015/0291966, 2017/0166903, 2019/0203212, which describe e.g., embodiments of inducible DNA binding proteins and methods of use and can be adapted for use with the present invention.

[0138] In some embodiments, transient or inducible expression can be achieved by including, for example, chemical-regulated promotors, i.e., whereby the application of an exogenous chemical induces gene expression. Modulation of gene expression can also be obtained by including a chemical-repressible promoter, where application of the chemical represses gene expression. Chemical-inducible promoters include, but are not limited to, the maize ln2-2 promoter, activated by benzene sulfonamide herbicide safeners (De Veylder et al., (1997) Plant Cell Physiol 38:568-77), the maize GST promoter (GST-ll-27, WO93/01294), activated by hydrophobic electrophilic compounds used as pre-emergent herbicides, and the tobacco PR-1 a promoter (Ono et al., (2004) Biosci Biotechnol Biochem 68:803-7) activated by salicylic acid. Promoters which are regulated by antibiotics, such as tetracycline-inducible and tetracycline-repressible promoters (Gatz et al., (1991) Mol Gen Genet 227:229-37; U.S. Patent Nos. 5,814,618 and 5,789,156) can also be used herein.

[0139] In some embodiments, the polynucleotide, vector or system thereof can include one or more elements capable of translocating and/or expressing a programmable pattern recognition protein or system polynucleotide to/in a specific cell component or organelle. Such organelles can include, but are not limited to, nucleus, ribosome, endoplasmic reticulum, Golgi apparatus, chloroplast, mitochondria, vacuole, lysosome, cytoskeleton, plasma membrane, cell wall, peroxisome, centrioles, etc. Such regulatory elements can include, but are not limited to, nuclear localization signals (examples of which are described in greater detail elsewhere herein), any such as those that are annotated in the LocSigDB database (see e.g., http://genome.unmc.edu/LocSigDB/ and Negi et al., 2015. Database. 2015: bav003; doi: 10.1093/database/bav003), nuclear export signals (e.g., LXXXLXXLXL (SEQ ID NO: 10) and others described elsewhere herein), endoplasmic reticulum localization/retention signals (e.g., KDEL, KDXX, KKXX, KXX, and others described elsewhere herein; and see e.g., Liu et al. 2007 Mol. Biol. Cell. 18(3):1073-1082 and Gorleku et al., 2011. J. Biol. Chem. 286:39573-

39584), mitochondria (see e.g., Cell Reports. 22:2818-2826, particularly at Fig. 2; Doyle et al. 2013. PLoS ONE 8, e67938; Funes et al. 2002. J. Biol. Chem. 277:6051-6058; Matouschek et al. 1997. PNAS USA 85:2091-2095; Oca-Cossio et al., 2003. 165:707-720; Waltner et al., 1996. J. Biol. Chem. 271:21226-21230; Wilcox et al., 2005. PNAS USA 102:15435-15440; Galanis et al., 1991. FEBS Lett 282:425-430, peroxisome (e.g., (S/A/C)-(K/R/H)-(L/A), SLK, (R/K)-(L/V/I)-XXXXX-(H/Q)-(L/A/F). Suitable protein targeting motifs can also be designed or identified using any suitable database or prediction tool, including but not limited to Minimotif Miner (http:minimotifminer.org, http://mitominer.mrc-mbu.cam.ac.uk/release-4.0/embodiment.do?name=Protein%20MTS), LocDB (see above), PTSs predictor (), TargetP-2.0 (http://www.cbs.dtu.dk/services/TargetP/), ChloroP (http://www.cbs.dtu.dk/services/ChloroP/); NetNES (http://www.cbs.dtu.dk/services/NetNES/), Predotar (https://urgi.versailles.inra.fr/predotar/), and SignalP (http://www.cbs.dtu.dk/services/SignalP/).

Selectable Markers and Tags

[0140] One or more of the programmable pattern recognition protein or system polynucleotides can be operably linked, fused to, or otherwise modified to include a polynucleotide that encodes or is a selectable marker or tag, which can be a polynucleotide or polypeptide. In some embodiments, the polypeptide encoding a polypeptide selectable marker can be incorporated in the programmable pattern recognition protein or system polynucleotide such that the selectable marker polypeptide, when translated, is inserted between two amino acids between the N- and C- terminus of the programmable pattern recognition protein or system polypeptide or at the N- and/or C-terminus of the programmable pattern recognition protein or system polypeptide. In some embodiments, the selectable marker or tag is a polynucleotide barcode or unique molecular identifier (UMI).

[0141] It will be appreciated that the polynucleotide encoding such selectable markers or tags can be incorporated into a polynucleotide encoding one or more components of the programmable pattern recognition protein or system described herein in an appropriate manner to allow expression of the selectable marker or tag. Such techniques and methods are described elsewhere herein and will be instantly appreciated by one of ordinary skill in the art in view of this disclosure. Many such selectable markers and tags are generally known in the art and are intended to be within the scope of this disclosure.

[0142] Suitable selectable markers and tags include, but are not limited to, affinity tags, such as chitin binding protein (CBP), maltose binding protein (MBP), glutathione-S-transferase (GST), poly(His) tag; solubilization tags such as thioredoxin (TRX) and poly(NANP), MBP, and GST; chromatography tags such as those consisting of polyanionic amino acids, such as FLAG-tag; epitope tags such as V5-tag, Myc-tag, HA-tag and NE-tag; protein tags that can allow specific enzymatic modification (such as biotinylation by biotin ligase) or chemical modification (such as reaction with FlAsH-EDT2 for fluorescence imaging), DNA and/or RNA segments that contain restriction enzyme or other enzyme cleavage sites; DNA segments that encode products that provide resistance against otherwise toxic compounds including antibiotics, such as, spectinomycin, ampicillin, kanamycin, tetracycline, Basta, neomycin phosphotransferase II (NEO), hygromycin phosphotransferase (HPT)) and the like; DNA and/or RNA segments that encode products that are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); DNA and/or RNA segments that encode products which can be readily identified (e.g., phenotypic markers such as β-galactosidase, GUS; fluorescent proteins such as green fluorescent protein (GFP), cyan (CFP), yellow (YFP), red (RFP), luciferase, and cell surface proteins); polynucleotides that can generate one or more new primer sites for PCR (e.g., the juxtaposition of two DNA sequences not previously juxtaposed), DNA sequences not acted upon or acted upon by a restriction endonuclease or other DNA modifying enzyme, chemical, etc.; epitope tags (e.g., GFP, FLAG- and His-tags), and, DNA sequences that make a molecular barcode or unique molecular identifier (UMI), DNA sequences required for a specific modification (e.g., methylation) that allows its identification. Other suitable markers will be appreciated by those of skill in the art.

[0143] Selectable markers and tags can be operably linked to one or more components of the CRISPR-Cas system described herein via suitable linker, such as a glycine or glycine serine linkers as short as GS or GG up to $(GGGGG)_3$ (SEQ ID NO: 11) or $(GGGGS)_3$ (SEQ ID NO: 12). Other suitable linkers are described elsewhere herein.

[0144] The vector or vector system can include one or more polynucleotides encoding one or more targeting moieties. In some embodiments, the targeting moiety encoding polynucleotides can be included in the vector or vector system, such as a viral vector system, such that they are expressed within and/or on the virus particle(s) produced such that the virus particles can be targeted to specific cells, tissues, organs, etc. In some embodiments, the targeting moiety encoding polynucleotides can be included in the vector or vector system such

that the programmable pattern recognition protein or system polynucleotide(s) and/or products expressed therefrom include the targeting moiety and can be targeted to specific cells, tissues, organs, etc. In some embodiments, such as non-viral carriers, the targeting moiety can be attached to the carrier (e.g., polymer, lipid, inorganic molecule etc.) and can be capable of targeting the carrier and any attached or associated programmable pattern recognition protein or system polynucleotide(s) to specific cells, tissues, organs, etc.

Vector Construction

[0145] The vectors described herein can be constructed using any suitable process or technique. In some embodiments, one or more suitable recombination and/or cloning methods or techniques can be used to the vector(s) described herein. Suitable recombination and/or cloning techniques and/or methods can include, but not limited to, those described in U.S. Patent Publication No. US 2004/0171156 A1. Other suitable methods and techniques are described elsewhere herein.

[0146] Construction of recombinant AAV vectors is 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). Any of the techniques and/or methods can be used and/or adapted for constructing an AAV or other vector described herein. nAAV vectors are discussed elsewhere herein.

[0147] In some embodiments, a vector comprises one or more insertion sites, such as a restriction endonuclease recognition sequence (also referred to as a “cloning site”). In some embodiments, one or more insertion sites (e.g., about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more insertion sites) are located upstream and/or downstream of one or more sequence elements of one or more vectors. When multiple different guide polynucleotides are used, a single expression construct may be used to target nucleic acid-targeting activity to multiple different, corresponding target sequences within a cell. For example, a single vector may comprise about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or more guide s polynucleotides. In some embodiments, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more such guide-polynucleotide-containing vectors may be provided, and optionally delivered to a cell.

[0148] Delivery vehicles, vectors, particles, nanoparticles, formulations and components thereof for expression of one or more elements of a programmable pattern recognition

composition or system described herein are as used in the foregoing documents, such as International Patent Publication No. WO 2014/093622 (PCT/US2013/074667) and are discussed in greater detail herein.

Viral Vectors

[0149] In some embodiments, the vector is a viral vector. The term of art "viral vector" and as used herein in this context refers to polynucleotide based vectors that contain one or more elements from or based upon one or more elements of a virus that can be capable of expressing and packaging a polynucleotide, such as a programmable pattern recognition polynucleotide of the present invention, into a virus particle and producing said virus particle when used alone or with one or more other viral vectors (such as in a viral vector system). Viral vectors and systems thereof can be used for producing viral particles for delivery of and/or expression of one or more components of the programmable pattern recognition composition or system described herein. The viral vector can be part of a viral vector system involving multiple vectors. In some embodiments, systems incorporating multiple viral vectors can increase the safety of these systems. Suitable viral vectors can include retroviral-based vectors, lentiviral-based vectors, adenoviral-based vectors, adeno associated vectors, helper-dependent adenoviral (HdAd) vectors, hybrid adenoviral vectors, herpes simplex virus-based vectors, poxvirus-based vectors, and Epstein-Barr virus-based vectors. Other embodiments of viral vectors and viral particles produce therefrom are described elsewhere herein. In some embodiments, the viral vectors are configured to produce replication incompetent viral particles for improved safety of these systems.

[0150] In certain embodiments, the virus structural component, which can be encoded by one or more polynucleotides in a viral vector or vector system, comprises one or more capsid proteins including an entire capsid. In certain embodiments, such as wherein a viral capsid comprises multiple copies of different proteins, the delivery system can provide one or more of the same protein or a mixture of such proteins. For example, AAV comprises 3 capsid proteins, VP1, VP2, and VP3, thus delivery systems of the invention can comprise one or more of VP1, and/or one or more of VP2, and/or one or more of VP3. Accordingly, the present invention is applicable to a virus within the family Adenoviridae, such as Atadenovirus, e.g., Ovine atadenovirus D, Aviadenovirus, e.g., Fowl aviadenovirus A, Ichtadenovirus, e.g., Sturgeon ichtadenovirus A, Mastadenovirus (which includes adenoviruses such as all human adenoviruses), e.g., Human mastadenovirus C, and Siadenovirus, e.g., Frog siadenovirus A.

Thus, a virus of within the family Adenoviridae is contemplated as within the invention with discussion herein as to adenovirus applicable to other family members. Target-specific AAV capsid variants can be used or selected. Non-limiting examples include capsid variants selected to bind to chronic myelogenous leukemia cells, human CD34 PBPC cells, breast cancer cells, cells of lung, heart, dermal fibroblasts, melanoma cells, stem cell, glioblastoma cells, coronary artery endothelial cells and keratinocytes. See, e.g., Buning et al, 2015, Current Opinion in Pharmacology 24, 94-104. From teachings herein and knowledge in the art as to modifications of adenovirus (see, e.g., US Patents 9,410,129, 7,344,872, 7,256,036, 6,911,199, 6,740,525; Matthews, "Capsid-Incorporation of Antigens into Adenovirus Capsid Proteins for a Vaccine Approach," Mol Pharm, 8(1): 3-11 (2011)), as well as regarding modifications of AAV, the skilled person can readily obtain a modified adenovirus that has a large payload protein or a CRISPR-protein, despite that heretofore it was not expected that such a large protein could be provided on an adenovirus. And as to the viruses related to adenovirus mentioned herein, as well as to the viruses related to AAV mentioned elsewhere herein, the teachings herein as to modifying adenovirus and AAV, respectively, can be applied to those viruses without undue experimentation from this disclosure and the knowledge in the art.

[0151] In some embodiments, the viral vector is configured such that when the cargo is packaged the cargo(s) (e.g., one or more components of the programmable pattern recognition composition or system, including but not limited to a STAND NTPase and/or optional effector, is external to the capsid or virus particle. In the sense that it is not inside the capsid (enveloped or encompassed with the capsid) but is externally exposed so that it can contact the target genomic DNA. In some embodiments, the viral vector is configured such that all the carog(s) are contained within the capsid after packaging.

Split Viral Vector Systems

[0152] When the programmable pattern recognition composition or system viral vector or vector system (be it a retroviral (e.g., AAV) or lentiviral vector) is designed so as to position the cargo(s) (e.g., one or more programmable pattern recognition composition or system components) at the internal surface of the capsid once formed, the cargo(s) will fill most or all of internal volume of the capsid. In other embodiments, the effector protein may be modified or divided so as to occupy a less of the capsid internal volume. Accordingly, in certain embodiments, the programmable pattern recognition composition or system or component thereof (can be divided in two portions, one portion comprises in one viral particle or capsid

and the second portion comprised in a second viral particle or capsid. In certain embodiments, by splitting the programmable pattern recognition composition or system or component thereof in two portions, space is made available to link one or more heterologous domains to one or both programmable pattern recognition composition or system component portions. Such systems can be referred to as "split vector systems" or in the context of the present disclosure a "split programmable pattern recognition composition or system" a "split programmable pattern recognition composition or system polypeptide", a "split STAND NTPase protein" and the like. This split protein approach is also described elsewhere herein. When the concept is applied to a vector system, it thus describes putting pieces of the split proteins on different vectors thus reducing the payload of any one vector. This approach can facilitate delivery of systems where the total system size is close to or exceeds the packaging capacity of the vector. This is independent of any regulation of the programmable pattern recognition composition or system that can be achieved with a split system or split protein design.

[0153] Split programmable pattern recognition composition or system polypeptides that can be incorporated into the AAV or other vectors described herein are set forth elsewhere herein and in documents incorporated herein by reference in further detail herein. In certain embodiments, each part of a split programmable pattern recognition composition or system polypeptides are attached to a member of a specific binding pair, and when bound with each other, the members of the specific binding pair maintain the parts of the programmable pattern recognition composition or system polypeptide in proximity. In certain embodiments, each part of a split programmable pattern recognition composition or system polypeptide is associated with an inducible binding pair. An inducible binding pair is one which is capable of being switched "on" or "off" by a protein or small molecule that binds to both members of the inducible binding pair. In general, according to the invention, programmable pattern recognition composition or system polypeptides may preferably split between domains, leaving domains intact. Preferred, non-limiting examples of such programmable pattern recognition composition or system polypeptides include, without limitation, STAND NTPase polypeptides, effector polypeptides, and orthologues.

[0154] In some embodiments, any AAV serotype is preferred. In some embodiments, the VP2 domain associated with the programmable pattern recognition composition or system polypeptide is an AAV serotype 2 VP2 domain. In some embodiments, the VP2 domain

associated with the programmable pattern recognition composition or system polypeptide is an AAV serotype 8 VP2 domain. The serotype can be a mixed serotype as is known in the art.

Retroviral and Lentiviral Vectors

[0155] Retroviral vectors can be composed 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. Suitable retroviral vectors for the CRISPR-Cas systems can include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian immunodeficiency virus (SIV), human immunodeficiency 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). Selection of a retroviral gene transfer system may therefore depend on the target tissue.

[0156] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and are described in greater detail elsewhere herein. A retrovirus can also be engineered to allow for conditional expression of the inserted transgene, such that only certain cell types are infected by the lentivirus.

[0157] Lentiviruses are complex retroviruses that have the ability to infect and express their genes in both mitotic and post-mitotic cells. Advantages of using a lentiviral approach can include the ability to transduce or infect non-dividing cells and their ability to typically produce high viral titers, which can increase efficiency or efficacy of production and delivery. Suitable lentiviral vectors include, but are not limited to, human immunodeficiency virus (HIV)-based lentiviral vectors, feline immunodeficiency virus (FIV)-based lentiviral vectors, simian immunodeficiency virus (SIV)-based lentiviral vectors, Moloney Murine Leukaemia Virus (Mo-MLV), Visna.maedi virus (VMV)-based lentiviral vector, carpine arthritis-encephalitis virus (CAEV)-based lentiviral vector, bovine immune deficiency virus (BIV)-based lentiviral vector, and Equine infectious anemia (EIAV)-based lentiviral vector. In some embodiments, an HIV-based lentiviral vector system can be used. In some embodiments, a FIV-based lentiviral vector system can be used.

[0158] In some embodiments, the lentiviral vector is an EIAV-based lentiviral vector or vector system. EIAV vectors have been used to mediate expression, packaging, and/or delivery in other contexts, such as for ocular gene therapy (see, e.g., Balagaan, J Gene Med 2006; 8: 275 – 285). In another embodiment, RetinoStat®, (see, e.g., Binley et al., HUMAN GENE THERAPY 23:980–991 (September 2012)), which describes RetinoStat®, an equine infectious anemia virus-based lentiviral gene therapy vector that expresses angiostatic proteins endostatin and angiostatin that is delivered via a subretinal injection for the treatment of the wet form of age-related macular degeneration. Any of these vectors described in these publications can be modified for the elements of the programmable pattern recognition composition or system described herein.

[0159] In some embodiments, the lentiviral vector or vector system thereof can be a first-generation lentiviral vector or vector system thereof. First-generation lentiviral vectors can contain a large portion of the lentivirus genome, including the gag and pol genes, other additional viral proteins (e.g., VSV-G) and other accessory genes (e.g., vif, vprm vpu, nef, and combinations thereof), regulatory genes (e.g., tat and/or rev) as well as the gene of interest between the LTRs. First generation lentiviral vectors can result in the production of virus particles that can be capable of replication in vivo, which may not be appropriate for some instances or applications.

[0160] In some embodiments, the lentiviral vector or vector system thereof can be a second-generation lentiviral vector or vector system thereof. Second-generation lentiviral vectors do not contain one or more accessory virulence factors and do not contain all components necessary for virus particle production on the same lentiviral vector. This can result in the production of a replication-incompetent virus particle and thus increase the safety of these systems over first-generation lentiviral vectors. In some embodiments, the second-generation vector lacks one or more accessory virulence factors (e.g., vif, vprm, vpu, nef, and combinations thereof). Unlike the first-generation lentiviral vectors, no single second generation lentiviral vector includes all features necessary to express and package a polynucleotide into a virus particle. In some embodiments, the envelope and packaging components are split between two different vectors with the gag, pol, rev, and tat genes being contained on one vector and the envelope protein (e.g., VSV-G) are contained on a second vector. The gene of interest, its promoter, and LTRs can be included on a third vector that can

be used in conjunction with the other two vectors (packaging and envelope vectors) to generate a replication-incompetent virus particle.

[0161] In some embodiments, the lentiviral vector or vector system thereof can be a third-generation lentiviral vector or vector system thereof. Third-generation lentiviral vectors and vector systems thereof have increased safety over first- and second-generation lentiviral vectors and systems thereof because, for example, the various components of the viral genome are split between two or more different vectors but used together in vitro to make virus particles, they can lack the tat gene (when a constitutively active promoter is included up-stream of the LTRs), and they can include one or more deletions in the 3'LTR to create self-inactivating (SIN) vectors having disrupted promoter/enhancer activity of the LTR. In some embodiments, a third-generation lentiviral vector system can include (i) a vector plasmid that contains the polynucleotide of interest and upstream promoter that are flanked by the 5' and 3' LTRs, which can optionally include one or more deletions present in one or both of the LTRs to render the vector self-inactivating; (ii) a "packaging vector(s)" that can contain one or more genes involved in packaging a polynucleotide into a virus particle that is produced by the system (e.g., gag, pol, and rev) and upstream regulatory sequences (e.g., promoter(s)) to drive expression of the features present on the packaging vector, and (iii) an "envelope vector" that contains one or more envelope protein genes and upstream promoters. In certain embodiments, the third-generation lentiviral vector system can include at least two packaging vectors, with the gag-pol being present on a different vector than the rev gene.

[0162] In some embodiments, self-inactivating lentiviral vectors with an siRNA targeting a common exon shared by HIV tat/rev, a nucleolar-localizing TAR decoy, and an anti–CCR5-specific hammerhead ribozyme (see, e.g., DiGiusto et al. (2010) Sci Transl Med 2:36ra43) can be used/and or adapted to the programmable pattern recognition composition or system of the present invention.

[0163] In some embodiments, the pseudotype and infectivity or tropisim of a lentivirus particle can be tuned by altering the type of envelope protein(s) included in the lentiviral vector or system thereof. As used herein, an "envelope protein" or "outer protein" means a protein exposed at the surface of a viral particle that is not a capsid protein. For example, envelope or outer proteins typically comprise proteins embedded in the envelope of the virus. In some embodiments, a lentiviral vector or vector system thereof can include a VSV-G envelope protein. VSV-G mediates viral attachment to an LDL receptor (LDLR) or an LDLR family

member present on a host cell, which triggers endocytosis of the viral particle by the host cell. Because LDLR is expressed by a wide variety of cells, viral particles expressing the VSV-G envelope protein can infect or transduce a wide variety of cell types. Other suitable envelope proteins can be incorporated based on the host cell that a user desires to be infected by a virus particle produced from a lentiviral vector or system thereof described herein and can include, but are not limited to, feline endogenous virus envelope protein (RD114) (see e.g., Hanawa et al. Molec. Ther. 2002 5(3) 242-251), modified Sindbis virus envelope proteins (see e.g., Morizono et al. 2010. J. Virol. 84(14) 6923-6934; Morizono et al. 2001. J. Virol. 75:8016-8020; Morizono et al. 2009. J. Gene Med. 11:549-558; Morizono et al. 2006 Virology 355:71-81; Morizono et al J. Gene Med. 11:655-663, Morizono et al. 2005 Nat. Med. 11:346-352), baboon retroviral envelope protein (see e.g., Girard-Gagnepain et al. 2014. Blood. 124: 1221-1231); Tupaia paramyxovirus glycoproteins (see e.g., Enkirch T. et al., 2013. Gene Ther. 20:16-23); measles virus glycoproteins (see e.g., Funke et al. 2008. Molec. Ther. 16(8): 1427-1436), rabies virus envelope proteins, MLV envelope proteins, Ebola envelope proteins, baculovirus envelope proteins, filovirus envelope proteins, hepatitis E1 and E2 envelope proteins, gp41 and gp120 of HIV, hemagglutinin, neuraminidase, M2 proteins of influenza virus, and combinations thereof.

[0164] In some embodiments, the tropism of the resulting lentiviral particle can be tuned by incorporating cell targeting peptides into a lentiviral vector such that the cell targeting peptides are expressed on the surface of the resulting lentiviral particle. In some embodiments, a lentiviral vector can contain an envelope protein that is fused to a cell targeting protein (see e.g., Buchholz et al. 2015. Trends Biotechnol. 33:777-790; Bender et al. 2016. PLoS Pathog. 12(e1005461); and Friedrich et al. 2013. Mol. Ther. 2013. 21: 849-859.

[0165] In some embodiments, a split-intein-mediated approach to target lentiviral particles to a specific cell type can be used (see e.g., Chamoun-Emaneulli et al. 2015. Biotechnol. Bioeng. 112:2611-2617, Ramirez et al. 2013. Protein. Eng. Des. Sel. 26:215-233. In these embodiments, a lentiviral vector can contain one half of a splicing-deficient variant of the naturally split intein from Nostoc punctiforme fused to a cell targeting peptide and the same or different lentiviral vector can contain the other half of the split intein fused to an envelope protein, such as a binding-deficient, fusion-competent virus envelope protein. This can result in production of a virus particle from the lentiviral vector or vector system that includes a split intein that can function as a molecular Velcro linker to link the cell-binding protein to the

pseudotyped lentivirus particle. This approach can be advantageous for use where surface-incompatibilities can restrict the use of, e.g., cell targeting peptides.

[0166] In some embodiments, a covalent-bond-forming protein-peptide pair can be incorporated into one or more of the lentiviral vectors described herein to conjugate a cell targeting peptide to the virus particle (see e.g., Kasaraneni et al. 2018. Sci. Reports (8) No. 10990). In some embodiments, a lentiviral vector can include an N-terminal PDZ domain of InaD protein (PDZ1) and its pentapeptide ligand (TEFCA (SEQ ID NO: 13)) from NorpA, which can conjugate the cell targeting peptide to the virus particle via a covalent bond (e.g., a disulfide bond). In some embodiments, the PDZ1 protein can be fused to an envelope protein, which can optionally be binding deficient and/or fusion competent virus envelope protein and included in a lentiviral vector. In some embodiments, the TEFCA (SEQ ID NO: 13) can be fused to a cell targeting peptide and the TEFCA-CPT (SEQ ID NO: 13)fusion construct can be incorporated into the same or a different lentiviral vector as the PDZ1-envenlope protein construct. During virus production, specific interaction between the PDZ1 and TEFCA (SEQ ID NO: 13) facilitates producing virus particles covalently functionalized with the cell targeting peptide and thus capable of targeting a specific cell-type based upon a specific interaction between the cell targeting peptide and cells expressing its binding partner. This approach can be advantageous for use where surface-incompatibilities can restrict the use of, e.g., cell targeting peptides.

[0167] Lentiviral vectors have been disclosed as in the treatment for Parkinson's Disease, see, e.g., US Patent Publication No. 20120295960 and US Patent Nos. 7303910 and 7351585. Lentiviral vectors have also been disclosed for the treatment of ocular diseases, see e.g., US Patent Publication Nos. 20060281180, 20090007284, US20110117189; US20090017543; US20070054961, US20100317109. Lentiviral vectors have also been disclosed for delivery to the brain, see, e.g., US Patent Publication Nos. US20110293571; US20110293571, US20040013648, US20070025970, US20090111106 and US Patent No. US7259015. Any of these systems or a variant thereof can be used to deliver a programmable pattern recognition composition or system polynucleotide described herein to a cell.

[0168] In some embodiments, a lentiviral vector system can include one or more transfer plasmids. Transfer plasmids can be generated from various other vector backbones and can include one or more features that can work with other retroviral and/or lentiviral vectors in the system that can, for example, improve safety of the vector and/or vector system, increase virial

titers, and/or increase or otherwise enhance expression of the desired insert to be expressed and/or packaged into the viral particle. Suitable features that can be included in a transfer plasmid can include, but are not limited to, 5'LTR, 3'LTR, SIN/LTR, origin of replication (Ori), selectable marker genes (e.g., antibiotic resistance genes), Psi (Ψ), RRE (rev response element), cPPT (central polypurine tract), promoters, WPRE (woodchuck hepatitis post-transcriptional regulatory element), SV40 polyadenylation signal, pUC origin, SV40 origin, F1 origin, and combinations thereof.

[0169] In another embodiment, Cocal vesiculovirus envelope pseudotyped retroviral or lentiviral vector particles are contemplated (see, e.g., US Patent Publication No. 20120164118 assigned to the Fred Hutchinson Cancer Research Center). Cocal virus is in the Vesiculovirus genus and is a causative agent of vesicular stomatitis in mammals. Cocal virus was originally isolated from mites in Trinidad (Jonkers et al., Am. J. Vet. Res. 25:236-242 (1964)), and infections have been identified in Trinidad, Brazil, and Argentina from insects, cattle, and horses. Many of the vesiculoviruses that infect mammals have been isolated from naturally infected arthropods, suggesting that they are vector-borne. Antibodies to vesiculoviruses are common among people living in rural areas where the viruses are endemic and laboratory-acquired; infections in humans usually result in influenza-like symptoms. The Cocal virus envelope glycoprotein shares 71.5% identity at the amino acid level with VSV-G Indiana, and phylogenetic comparison of the envelope gene of vesiculoviruses shows that Cocal virus is serologically distinct from, but most closely related to, VSV-G Indiana strains among the vesiculoviruses. Jonkers et al., Am. J. Vet. Res. 25:236-242 (1964) and Travassos da Rosa et al., Am. J. Tropical Med. & Hygiene 33:999-1006 (1984). The Cocal vesiculovirus envelope pseudotyped retroviral vector particles may include for example, lentiviral, alpharetroviral, betaretroviral, gammaretroviral, deltaretroviral, and epsilonretroviral vector particles that may comprise retroviral Gag, Pol, and/or one or more accessory protein(s) and a Cocal vesiculovirus envelope protein. In certain embodiments of these embodiments, the Gag, Pol, and accessory proteins are lentiviral and/or gammaretroviral. In some embodiments, a retroviral vector can contain encoding polypeptides for one or more Cocal vesiculovirus envelope proteins such that the resulting viral or pseudoviral particles are Cocal vesiculovirus envelope pseudotyped.

Adenoviral vectors, Helper-dependent Adenoviral vectors, and Hybrid Adenoviral Vectors

[0170] In some embodiments, the vector can be an adenoviral vector. In some embodiments, the adenoviral vector can include elements such that the virus particle produced

using the vector or system thereof can be serotype 2 or serotype 5. In some embodiments, the polynucleotide to be delivered via the adenoviral particle can be up to about 8 kb. Thus, in some embodiments, an adenoviral vector can include a DNA polynucleotide to be delivered that can range in size from about 0.001 kb to about 8 kb. Adenoviral vectors have been used successfully in several contexts (see e.g., Teramato et al. 2000. Lancet. 355:1911-1912; Lai et al. 2002. DNA Cell. Biol. 21:895-913; Flotte et al., 1996. Hum. Gene. Ther. 7:1145-1159; and Kay et al. 2000. Nat. Genet. 24:257-261.

[0171] In some embodiments the vector can be a helper-dependent adenoviral vector or system thereof. These are also referred to in the art as "gutless" or "gutted" vectors and are a modified generation of adenoviral vectors (see e.g., Thrasher et al. 2006. Nature. 443:E5-7). In certain embodiments of the helper-dependent adenoviral vector system one vector (the helper) can contain all the viral genes required for replication but contains a conditional gene defect in the packaging domain. The second vector of the system can contain only the ends of the viral genome, one or more CRISPR-Cas polynucleotides, and the native packaging recognition signal, which can allow selective packaged release from the cells (see e.g., Cideciyan et al. 2009. N Engl J Med. 361:725-727). Helper-dependent adenoviral vector systems have been successful for gene delivery in several contexts (see e.g., Simonelli et al. 2010. J Am Soc Gene Ther. 18:643-650; Cideciyan et al. 2009. N Engl J Med. 361:725-727; Crane et al. 2012. Gene Ther. 19(4):443-452; Alba et al. 2005. Gene Ther. 12:18-S27; Croyle et al. 2005. Gene Ther. 12:579-587; Amalfitano et al. 1998. J. Virol. 72:926-933; and Morral et al. 1999. PNAS. 96:12816-12821). The techniques and vectors described in these publications can be adapted for inclusion and delivery of the programmable pattern recognition composition or system polynucleotides described herein. In some embodiments, the polynucleotide to be delivered via the viral particle produced from a helper-dependent adenoviral vector or system thereof can be up to about 37 kb. Thus, in some embodiments, an adenoviral vector can include a DNA polynucleotide to be delivered that can range in size from about 0.001 kb to about 37 kb (see e.g., Rosewell et al. 2011. J. Genet. Syndr. Gene Ther. Suppl. 5:001).

[0172] In some embodiments, the vector is a hybrid-adenoviral vector or system thereof. Hybrid adenoviral vectors are composed of the high transduction efficiency of a gene-deleted adenoviral vector and the long-term genome-integrating potential of adeno-associated, retroviruses, lentivirus, and transposon based-gene transfer. In some embodiments, such hybrid vector systems can result in stable transduction and limited integration site. See e.g., Balague

et al. 2000. Blood. 95:820-828; Morral et al. 1998. Hum. Gene Ther. 9:2709-2716; Kubo and Mitani. 2003. J. Virol. 77(5): 2964-2971; Zhang et al. 2013. PloS One. 8(10) e76771; and Cooney et al. 2015. Mol. Ther. 23(4):667-674), whose techniques and vectors described therein can be modified and adapted for use in the programmable pattern recognition composition or system of the present invention. In some embodiments, a hybrid-adenoviral vector can include one or more features of a retrovirus and/or an adeno-associated virus. In some embodiments the hybrid-adenoviral vector can include one or more features of a spuma retrovirus or foamy virus (FV). See e.g., Ehrhardt et al. 2007. Mol. Ther. 15:146-156 and Liu et al. 2007. Mol. Ther. 15:1834-1841, whose techniques and vectors described therein can be modified and adapted for use in the programmable pattern recognition composition or system of the present invention. Advantages of using one or more features from the FVs in the hybrid-adenoviral vector or system thereof can include the ability of the viral particles produced therefrom to infect a broad range of cells, a large packaging capacity as compared to other retroviruses, and the ability to persist in quiescent (non-dividing) cells. See also e.g., Ehrhardt et al. 2007. Mol. Ther. 156:146-156 and Shuji et al. 2011. Mol. Ther. 19:76-82, whose techniques and vectors described therein can be modified and adapted for use in the programmable pattern recognition composition or system of the present invention.

Adeno Associated Viral (AAV) Vectors

[0173] In an embodiment, the vector can be an adeno-associated virus (AAV) vector. 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); and Muzyczka, J. Clin. Invest. 94:1351 (1994). Although similar to adenoviral vectors in some of their features, AAVs have some deficiency in their replication and/or pathogenicity and thus can be safer that adenoviral vectors. In some embodiments the AAV can integrate into a specific site on chromosome 19 of a human cell with no observable side effects. In some embodiments, the capacity of the AAV vector, system thereof, and/or AAV particles can be up to about 4.7 kb.

[0174] The AAV vector or system thereof can include one or more regulatory molecules. In some embodiments the regulatory molecules can be promoters, enhancers, repressors and the like, which are described in greater detail elsewhere herein. In some embodiments, the AAV vector or system thereof can include one or more polynucleotides that can encode one or more regulatory proteins. In some embodiments, the one or more regulatory proteins can be selected from Rep78, Rep68, Rep52, Rep40, variants thereof, and combinations thereof.

[0175] The AAV vector or system thereof can include one or more polynucleotides that can encode one or more capsid proteins. The capsid proteins can be selected from VP1, VP2, VP3, and combinations thereof. The capsid proteins can be capable of assembling into a protein shell of the AAV virus particle. In some embodiments, the AAV capsid can contain 60 capsid proteins. In some embodiments, the ratio of VP1:VP2:VP3 in a capsid can be about 1:1:10.

[0176] In some embodiments, the AAV vector or system thereof can include one or more adenovirus helper factors or polynucleotides that can encode one or more adenovirus helper factors. Such adenovirus helper factors can include, but are not limited, E1A, E1B, E2A, E4ORF6, and VA RNAs. In some embodiments, a producing host cell line expresses one or more of the adenovirus helper factors.

[0177] The AAV vector or system thereof can be configured to produce AAV particles having a specific serotype. In some embodiments, the serotype can be AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-8, AAV-9 or any combinations thereof. In some embodiments, the AAV can be AAV1, AAV-2, AAV-5 or any combination thereof. One can select the AAV of the AAV with regard to the cells to be targeted; e.g., one can select AAV serotypes 1, 2, 5 or a hybrid capsid AAV-1, AAV-2, AAV-5 or any combination thereof for targeting brain and/or neuronal cells; and one can select AAV-4 for targeting cardiac tissue; and one can select AAV8 for delivery to the liver. Thus, in some embodiments, an AAV vector or system thereof capable of producing AAV particles capable of targeting the brain and/or neuronal cells can be configured to generate AAV particles having serotypes 1, 2, 5 or a hybrid capsid AAV-1, AAV-2, AAV-5 or any combination thereof. In some embodiments, an AAV vector or system thereof capable of producing AAV particles capable of targeting cardiac tissue can be configured to generate an AAV particle having an AAV-4 serotype. In some embodiments, an AAV vector or system thereof capable of producing AAV particles capable of targeting the liver can be configured to generate an AAV having an AAV-8 serotype. In some embodiments, the AAV vector is a hybrid AAV vector or system thereof. Hybrid AAVs are AAVs that include genomes with elements from one serotype that are packaged into a capsid derived from at least one different serotype. For example, if it is the rAAV2/5 that is to be produced, and if the production method is based on the helper-free, transient transfection method discussed above, the 1st plasmid and the 3rd plasmid (the adeno helper plasmid) will be the same as discussed for rAAV2 production. However, the second plasmid, the pRepCap will be different. In this plasmid, called pRep2/Cap5, the Rep gene is still derived from AAV2,

while the Cap gene is derived from AAV5. The production scheme is the same as the above-mentioned approach for AAV2 production. The resulting rAAV is called rAAV2/5, in which the genome is based on recombinant AAV2, while the capsid is based on AAV5. It is assumed the cell or tissue-tropism displayed by this AAV2/5 hybrid virus should be the same as that of AAV5.

[0178] A tabulation of certain AAV serotypes as to these cells can be found in Grimm, D. et al, J. Virol. 82: 5887-5911 (2008).

[0179] In some embodiments, the AAV vector or system thereof is configured as a "gutless" vector, similar to that described in connection with a retroviral vector. In some embodiments, the "gutless" AAV vector or system thereof can have the cis-acting viral DNA elements involved in genome amplification and packaging in linkage with the heterologous sequences of interest (e.g., the programmable pattern recognition composition or system polynucleotide(s)).

[0180] In some embodiments, the AAV vectors are produced in in insect cells, e.g., Spodoptera frugiperda Sf9 insect cells, grown in serum-free suspension culture. Serum-free insect cells can be purchased from commercial vendors, e.g., Sigma Aldrich (EX-CELL 405).

[0181] In another embodiment, the invention provides a non-naturally occurring or engineered programmable pattern recognition composition or system protein associated with Adeno Associated Virus (AAV), e.g., an AAV comprising a programmable pattern recognition composition or system protein as a fusion, with or without a linker, to or with an AAV capsid protein such as VP1, VP2, and/or VP3; and, for shorthand purposes, such a non-naturally occurring or engineered programmable pattern recognition composition or system protein is herein termed a "AAV- programmable pattern recognition composition or system protein" More in particular, modifying the knowledge in the art, e.g., Rybniker et al., "Incorporation of Antigens into Viral Capsids Augments Immunogenicity of Adeno-Associated Virus Vector-Based Vaccines," J Virol. Dec 2012; 86(24): 13800–13804, Lux K, et al. 2005. Green fluorescent protein-tagged adeno-associated virus particles allow the study of cytosolic and nuclear trafficking. J. Virol. 79:11776–11787, Munch RC, et al. 2012. "Displaying high-affinity ligands on adeno-associated viral vectors enables tumor cell-specific and safe gene transfer." Mol. Ther. [Epub ahead of print.] doi:10.1038/mt.2012.186 and Warrington KH, Jr, et al. 2004. Adeno-associated virus type 2 VP2 capsid protein is nonessential and can tolerate large peptide insertions at its N terminus. J. Virol. 78:6595–6609, each incorporated herein by

reference, one can obtain a modified AAV capsid of the invention. It will be understood by those skilled in the art that the modifications described herein if inserted into the AAV cap gene may result in modifications in the VP1, VP2 and/or VP3 capsid subunits. Alternatively, the capsid subunits can be expressed independently to achieve modification in only one or two of the capsid subunits (VP1, VP2, VP3, VP1+VP2, VP1+VP3, or VP2+VP3). One can modify the cap gene to have expressed at a desired location a non-capsid protein advantageously a large payload protein, such as a programmable pattern recognition composition or system - protein. Likewise, these can be fusions, with the protein, e.g., large payload protein such as a programmable pattern recognition composition or system-protein fused in a manner analogous to prior art fusions. *See, e.g.*, US Patent Publication 20090215879; Nance et al., "Perspective on Adeno-Associated Virus Capsid Modification for Duchenne Muscular Dystrophy Gene Therapy," Hum Gene Ther. 26(12):786–800 (2015) and documents cited therein, incorporated herein by reference. The skilled person, from this disclosure and the knowledge in the art can make and use modified AAV or AAV capsid as in the herein invention, and through this disclosure one knows now that large payload proteins can be fused to the AAV capsid. Applicants provide AAV capsid programmable pattern recognition composition or system R protein fusions and those AAV-capsid programmable pattern recognition composition or system protein fusions can be a recombinant AAV that contains nucleic acid molecule(s) encoding or providing programmable pattern recognition composition or system or complex RNA guide(s), whereby the programmable pattern recognition composition or system protein fusion delivers a programmable pattern recognition composition or system complex by the fusion, e.g., VP1, VP2, or VP3 fusion, and the guide RNA is provided by the coding of the recombinant virus, whereby in vivo, in a cell, the programmable pattern recognition composition or system is assembled from the nucleic acid molecule(s) of the recombinant providing the guide RNA and the outer surface of the virus providing the programmable pattern recognition composition or system polypeptide. Accordingly, the instant invention is also applicable to a virus in the genus Dependoparvovirus or in the family Parvoviridae, for instance, AAV, or a virus of Amdoparvovirus, e.g., Carnivore amdoparvovirus 1, a virus of Aveparvovirus, e.g., Galliform aveparvovirus 1, a virus of Bocaparvovirus, e.g., Ungulate bocaparvovirus 1, a virus of Copiparvovirus, e.g., Ungulate copiparvovirus 1, a virus of Dependoparvovirus, e.g., Adeno-associated dependoparvovirus A, a virus of Erythroparvovirus, e.g., Primate erythroparvovirus 1, a virus of Protoparvovirus, e.g., Rodent

protoparvovirus 1, a virus of Tetraparvovirus, e.g., Primate tetraparvovirus 1. Thus, a virus of within the family Parvoviridae or the genus Dependoparvovirus or any of the other foregoing genera within Parvoviridae is contemplated as within the invention with discussion herein as to AAV applicable to such other viruses.

[0182] In some embodiments, the programmable pattern recognition composition or system polypeptide is external to the capsid or virus particle. In the sense that it is not inside the capsid (enveloped or encompassed with the capsid) but is externally exposed so that it can contact the target genomic DNA). In some embodiments, the programmable pattern recognition composition or system polypeptide is associated with the AAV VP2 domain by way of a fusion protein. In some embodiments, the association may be considered to be a modification of the VP2 domain. Where reference is made herein to a modified VP2 domain, then this will be understood to include any association discussed herein of the VP2 domain and the programmable pattern recognition composition or system polypeptide. In some embodiments, the AAV VP2 domain may be associated (or tethered) to the programmable pattern recognition composition or system polypeptide via a connector protein, for example using a system such as the streptavidin-biotin system. In an embodiment, the present invention provides a polynucleotide encoding the present programmable pattern recognition composition or system polypeptide and associated AAV VP2 domain. In one embodiment, the invention provides a non-naturally occurring modified AAV having a VP2-programmable pattern recognition composition or system polypeptide capsid protein, wherein the programmable pattern recognition composition or system polypeptide is part of or tethered to the VP2 domain. In some preferred embodiments, the programmable pattern recognition composition or system polypeptide is fused to the VP2 domain so that, in another embodiment, the invention provides a non-naturally occurring modified AAV having a VP2- programmable pattern recognition composition or system polypeptide fusion capsid protein. Thus, reference herein to a VP2-programmable pattern recognition composition or system polypeptide capsid protein may also include a VP2-programmable pattern recognition composition or system polypeptide fusion capsid protein. In some embodiments, the VP2-programmable pattern recognition composition or system polypeptide capsid protein further comprises a linker, whereby the VP2-programmable pattern recognition composition or system polypeptide is distanced from the remainder of the AAV. In some embodiments, the VP2-programmable pattern recognition composition or system polypeptide capsid protein further comprises at least one protein

complex, e.g., programmable pattern recognition composition or system polypeptide complex, such as a programmable pattern recognition composition or system polypeptide complex guide RNA that targets a particular DNA, TALE, etc. A programmable pattern recognition composition or system polypeptide complex, such as programmable pattern recognition composition or system comprising the VP2- programmable pattern recognition composition or system polypeptide capsid protein and at least one programmable pattern recognition composition or system polypeptide complex, such as a programmable pattern recognition composition or system polypeptide complex guide RNA that targets a particular DNA, is also provided in one embodiment.

[0183] In one embodiment, the invention provides a non-naturally occurring or engineered composition comprising a programmable pattern recognition composition or system polypeptide which is part of or tethered to an AAV capsid domain, i.e., VP1, VP2, or VP3 domain of Adeno-Associated Virus (AAV) capsid. In some embodiments, part of or tethered to an AAV capsid domain includes associated with associated with a AAV capsid domain. In some embodiments, the programmable pattern recognition composition or system polypeptide may be fused to the AAV capsid domain. In some embodiments, the fusion may be to the N-terminal end of the AAV capsid domain. As such, in some embodiments, the C- terminal end of the programmable pattern recognition composition or system polypeptide is fused to the N-terminal end of the AAV capsid domain. In some embodiments, an NLS and/or a linker (such as a GlySer linker) may be positioned between the C- terminal end of the programmable pattern recognition composition or system polypeptide and the N- terminal end of the AAV capsid domain. In some embodiments, the fusion may be to the C-terminal end of the AAV capsid domain. In some embodiments, this is not preferred due to the fact that the VP1, VP2 and VP3 domains of AAV are alternative splices of the same RNA and so a C- terminal fusion may affect all three domains. In some embodiments, the AAV capsid domain is truncated. In some embodiments, some or all of the AAV capsid domain is removed. In some embodiments, some of the AAV capsid domain is removed and replaced with a linker (such as a GlySer linker), typically leaving the N- terminal and C- terminal ends of the AAV capsid domain intact, such as the first 2, 5 or 10 amino acids. In this way, the internal (non-terminal) portion of the VP3 domain may be replaced with a linker. It is particularly preferred that the linker is fused to the CRISPR protein. A branched linker may be used, with the programmable pattern recognition composition or system polypeptide fused to the end of one of the branches. This allows for

some degree of spatial separation between the capsid and the programmable pattern recognition composition or system polypeptide. In this way, the programmable pattern recognition composition or system polypeptide is part of (or fused to) the AAV capsid domain.

[0184] In other embodiments, the CRISPR enzyme may be fused in frame within, i.e. internal to, the AAV capsid domain. Thus, in some embodiments, the AAV capsid domain again preferably retains its N- terminal and C- terminal ends. In this case, a linker is preferred, in some embodiments, either at one or both ends of the programmable pattern recognition composition or system polypeptide. In this way, the programmable pattern recognition composition or system polypeptide is again part of (or fused to) the AAV capsid domain. In certain embodiments, the positioning of the programmable pattern recognition composition or system polypeptide is such that the programmable pattern recognition composition or system polypeptide is at the external surface of the viral capsid once formed. In one embodiment, the invention provides a non-naturally occurring or engineered composition comprising a programmable pattern recognition composition or system polypeptide associated with a AAV capsid domain of Adeno-Associated Virus (AAV) capsid. Here, associated may mean in some embodiments fused, or in some embodiments bound to, or in some embodiments tethered to. The programmable pattern recognition composition or system polypeptide may, in some embodiments, be tethered to the VP1, VP2, or VP3 domain. This may be via a connector protein or tethering system such as the biotin-streptavidin system. In one example, a biotinylation sequence (15 amino acids) could therefore be fused to the programmable pattern recognition composition or system polypeptide. When a fusion of the AAV capsid domain, especially the N- terminus of the AAV AAV capsid domain, with streptavidin is also provided, the two will therefore associate with very high affinity. Thus, in some embodiments, provided is a composition or system comprising a programmable pattern recognition composition or system polypeptide-biotin fusion and a streptavidin- AAV capsid domain arrangement, such as a fusion. The programmable pattern recognition composition or system polypeptide-biotin and streptavidin- AAV capsid domain forms a single complex when the two parts are brought together. NLSs may also be incorporated between the programmable pattern recognition composition or system polypeptide and the biotin; and/or between the streptavidin and the AAV capsid domain.

[0185] As such, provided is a fusion of a programmable pattern recognition composition or system polypeptide with a connector protein specific for a high affinity ligand for that

connector, whereas the AAV VP2 domain is bound to said high affinity ligand. For example, streptavidin may be the connector fused to the programmable pattern recognition composition or system polypeptide, while biotin may be bound to the AAV VP2 domain. Upon co-localization, the streptavidin will bind to the biotin, thus connecting the programmable pattern recognition composition or system polypeptide to the AAV VP2 domain. The reverse arrangement is also possible. In some embodiments, a biotinylation sequence (15 amino acids) could therefore be fused to the AAV VP2 domain, especially the N- terminus of the AAV VP2 domain. A fusion of the programmable pattern recognition composition or system polypeptide with streptavidin is also preferred, in some embodiments. In some embodiments, the biotinylated AAV capsids with streptavidin-programmable pattern recognition composition or system polypeptide are assembled *in vitro*. This way the AAV capsids should assemble in a straightforward manner and the programmable pattern recognition composition or system polypeptide-streptavidin fusion can be added after assembly of the capsid. In other embodiments a biotinylation sequence (15 amino acids) could therefore be fused to the programmable pattern recognition composition or system polypeptide, together with a fusion of the AAV VP2 domain, especially the N- terminus of the AAV VP2 domain, with streptavidin. For simplicity, a fusion of the programmable pattern recognition composition or system polypeptide and the AAV VP2 domain is preferred in some embodiments. In some embodiments, the fusion may be to the N- terminal end of the programmable pattern recognition composition or system polypeptide. In other words, in some embodiments, the AAV and programmable pattern recognition composition or system polypeptide are associated via fusion. In some embodiments, the AAV and programmable pattern recognition composition or system polypeptide are associated via fusion including a linker. Suitable linkers are discussed herein but include Gly Ser linkers. Fusion to the N- term of AAV VP2 domain is preferred, in some embodiments. In some embodiments, the programmable pattern recognition composition or system polypeptide comprises at least one Nuclear Localization Signal (NLS). In a further embodiment, the present invention provides compositions comprising the programmable pattern recognition composition or system polypeptide and associated AAV VP2 domain or the polynucleotides or vectors described herein. Such compositions and formulations are discussed elsewhere herein.

[0186] An alternative tether may be to fuse or otherwise associate the AAV capsid domain to an adaptor protein which binds to or recognizes to a corresponding RNA sequence or motif.

In some embodiments, the adaptor is or comprises a binding protein which recognizes and binds (or is bound by) an RNA sequence specific for said binding protein. In some embodiments, a preferred example is the MS2 (*see* Konermann et al. Dec 2014, cited *infra*, incorporated herein by reference) binding protein which recognizes and binds (or is bound by) an RNA sequence specific for the MS2 protein.

[0187] With the AAV capsid domain associated with the adaptor protein, the CRISPR protein may, in some embodiments, be tethered to the adaptor protein of the AAV capsid domain. The programmable pattern recognition composition or system polypeptide may, in some embodiments, be tethered to the adaptor protein of the AAV capsid domain via the CRISPR enzyme being in a complex with a modified guide, see Konermann *et al.* The modified guide is, in some embodiments, a sgRNA. In some embodiments, the modified guide comprises a distinct RNA sequence; see, e.g., International Patent Application No. PCT/US14/70175, incorporated herein by reference.

[0188] In some embodiments, distinct RNA sequence is an aptamer. Thus, corresponding aptamer- adaptor protein systems are preferred. One or more functional domains may also be associated with the adaptor protein. An example of a preferred arrangement would be: [AAV AAV capsid domain - adaptor protein] - [modified guide - programmable pattern recognition composition or system polypeptide].

[0189] In certain embodiments, the positioning of the programmable pattern recognition composition or system polypeptide is such that the programmable pattern recognition composition or system polypeptide is at the internal surface of the viral capsid once formed. In one embodiment, the invention provides a non-naturally occurring or engineered composition comprising a programmable pattern recognition composition or system polypeptide associated with an internal surface of an AAV capsid domain. Here again, associated may mean in some embodiments fused, or in some embodiments bound to, or in some embodiments tethered to. The programmable pattern recognition composition or system polypeptide may, in some embodiments, be tethered to the VP1, VP2, or VP3 domain such that it locates to the internal surface of the viral capsid once formed. This may be via a connector protein or tethering system such as the biotin-streptavidin system as described above and/or elsewhere herein.

Herpes Simplex Viral Vectors

[0190] In some embodiments, the vector can be a Herpes Simplex Viral (HSV)-based vector or system thereof. HSV systems can include the disabled infections single copy (DISC) viruses, which are composed of a glycoprotein H defective mutant HSV genome. When the defective HSV is propagated in complementing cells, virus particles can be generated that are capable of infecting subsequent cells permanently replicating their own genome but are not capable of producing more infectious particles. See e.g., 2009. Trobridge. Exp. Opin. Biol. Ther. 9:1427-1436, whose techniques and vectors described therein can be modified and adapted for use in the CRISPR-Cas system of the present invention. In some embodiments where an HSV vector or system thereof is utilized, the host cell can be a complementing cell. In some embodiments, HSV vector or system thereof can be capable of producing virus particles capable of delivering a polynucleotide cargo of up to 150 kb. Thus, in some embodiment the programmable pattern recognition composition or system polynucleotide(s) included in the HSV-based viral vector or system thereof can sum from about 0.001 to about 150 kb. HSV-based vectors and systems thereof have been successfully used in several contexts including various models of neurologic disorders. See e.g., Cockrell et al. 2007. Mol. Biotechnol. 36:184-204; Kafri T. 2004. Mol. Biol. 246:367-390; Balaggan and Ali. 2012. Gene Ther. 19:145-153; Wong et al. 2006. Hum. Gen. Ther. 2002. 17:1-9; Azzouz et al. J. Neruosci. 22L10302-10312; and Betchen and Kaplitt. 2003. Curr. Opin. Neurol. 16:487-493, whose techniques and vectors described therein can be modified and adapted for use in the CRISPR-Cas system of the present invention.

Poxvirus Vectors

[0191] In some embodiments, the vector can be a poxvirus vector or system thereof. In some embodiments, the poxvirus vector can result in cytoplasmic expression of one or more programmable pattern recognition composition or system polynucleotides of the present invention. In some embodiments the capacity of a poxvirus vector or system thereof can be about 25 kb or more. In some embodiments, a poxvirus vector or system thereof can include one or more programmable pattern recognition composition or system polynucleotides described herein.

Viral Vectors for delivery to plants

[0192] The systems and compositions may be delivered to plant cells using viral vehicles. In particular embodiments, the compositions and systems may be introduced in the plant cells

using a plant viral vector (e.g., as described in Scholthof et al. 1996, Annu Rev Phytopathol. 1996;34:299-323). Such viral vector may be a vector from a DNA virus, e.g., geminivirus (e.g., cabbage leaf curl virus, bean yellow dwarf virus, wheat dwarf virus, tomato leaf curl virus, maize streak virus, tobacco leaf curl virus, or tomato golden mosaic virus) or nanovirus (e.g., Faba bean necrotic yellow virus). The viral vector may be a vector from an RNA virus, e.g., tobravirus (e.g., tobacco rattle virus, tobacco mosaic virus), potexvirus (e.g., potato virus X), or hordeivirus (e.g., barley stripe mosaic virus). The replicating genomes of plant viruses may be non-integrative vectors.

Virus Particle Production from Viral Vectors

Retroviral Production

[0193] In some embodiments, one or more viral vectors and/or system thereof can be delivered to a suitable cell line for production of virus particles containing the polynucleotide or other payload to be delivered to a host cell. Suitable host cells for virus production from viral vectors and systems thereof described herein are known in the art and are commercially available. For example, suitable host cells include HEK 293 cells and its variants (HEK 293T and HEK 293TN cells). In some embodiments, the suitable host cell for virus production from viral vectors and systems thereof described herein can stably express one or more genes involved in packaging (e.g., pol, gag, and/or VSV-G) and/or other supporting genes.

[0194] In some embodiments, after delivery of one or more viral vectors to the suitable host cells for or virus production from viral vectors and systems thereof, the cells are incubated for an appropriate length of time to allow for viral gene expression from the vectors, packaging of the polynucleotide to be delivered (e.g., a programmable pattern recognition composition or system polynucleotide), and virus particle assembly, and secretion of mature virus particles into the culture media. Various other methods and techniques are generally known to those of ordinary skill in the art.

[0195] Mature virus particles can be collected from the culture media by a suitable method. In some embodiments, this can involve centrifugation to concentrate the virus. The titer of the composition containing the collected virus particles can be obtained using a suitable method. Such methods can include transducing a suitable cell line (e.g., NIH 3T3 cells) and determining transduction efficiency, infectivity in that cell line by a suitable method. Suitable methods include PCR-based methods, flow cytometry, and antibiotic selection-based methods. Various other methods and techniques are generally known to those of ordinary skill in the art. The

concentration of virus particle can be adjusted as needed. In some embodiments, the resulting composition containing virus particles can contain 1 $X10^1$ -1 X 10^{20} particles/mL.

[0196] Lentiviruses may be prepared from any lentiviral vector or vector system described herein. In one example embodiment, after cloning pCasES10 (which contains a lentiviral transfer plasmid backbone), HEK293FT at low passage (p=5) can be seeded in a T-75 flask to 50% confluence the day before transfection in DMEM with 10% fetal bovine serum and without antibiotics. After 20 hours, the media can be changed to OptiMEM (serum-free) media and transfection of the lentiviral vectors can done 4 hours later. Cells can be transfected with 10 µg of lentiviral transfer plasmid (pCasES10) and the appropriate packaging plasmids (e.g., 5 µg of pMD2.G (VSV-g pseudotype), and 7.5ug of psPAX2 (gag/pol/rev/tat)). Transfection can be carried out in 4mL OptiMEM with a cationic lipid delivery agent (50uL Lipofectamine 2000 and 100ul Plus reagent). After 6 hours, the media can be changed to antibiotic-free DMEM with 10% fetal bovine serum. These methods can use serum during cell culture, but serum-free methods are preferred.

[0197] Following transfection and allowing the producing cells (also referred to as packaging cells) to package and produce virus particles with packaged cargo, the lentiviral particles can be purified. In an exemplary embodiment, virus-containing supernatants can be harvested after 48 hours. Collected virus-containing supernatants can first be cleared of debris and filtered through a 0.45um low protein binding (PVDF) filter. They can then be spun in an ultracentrifuge for 2 hours at 24,000 rpm. The resulting virus-containing pellets can be resuspended in 50ul of DMEM overnight at 4 degrees C. They can be then aliquoted and used immediately or immediately frozen at -80 degrees C for storage.

AAV Particle Production

[0198] There are two main strategies for producing AAV particles from AAV vectors and systems thereof, such as those described herein, which depend on how the adenovirus helper factors are provided (helper v. helper free). In some embodiments, a method of producing AAV particles from AAV vectors and systems thereof can include adenovirus infection into cell lines that stably harbor AAV replication and capsid encoding polynucleotides along with AAV vector containing the polynucleotide to be packaged and delivered by the resulting AAV particle (e.g., the CRISPR-Cas system polynucleotide(s)). In some embodiments, a method of producing AAV particles from AAV vectors and systems thereof can be a "helper free" method, which includes co-transfection of an appropriate producing cell line with three vectors

(e.g., plasmid vectors): (1) an AAV vector that contains a polynucleotide of interest (e.g., the CRISPR-Cas system polynucleotide(s)) between 2 ITRs; (2) a vector that carries the AAV Rep-Cap encoding polynucleotides; and (3) helper polynucleotides. One of skill in the art will appreciate various methods and variations thereof that are both helper and -helper free and as well as the different advantages of each system.

Non-Viral Vectors

[0199] In some embodiments, the vector is a non-viral vector or vector system. The term of art "Non-viral vector" and as used herein in this context refers to molecules and/or compositions that are vectors but that are not based on one or more component of a virus or virus genome (excluding any nucleotide to be delivered and/or expressed by the non-viral vector) that can be capable of incorporating programmable pattern recognition composition or system polynucleotide(s) and delivering said programmable pattern recognition composition or system polynucleotide(s) to a cell and/or expressing the polynucleotide in the cell. It will be appreciated that this does not exclude vectors containing a polynucleotide designed to target a virus-based polynucleotide that is to be delivered. For example, if a gRNA to be delivered is directed against a virus component and it is inserted or otherwise coupled to an otherwise non-viral vector or carrier, this would not make said vector a "viral vector". Non-viral vectors can include, without limitation, naked polynucleotides and polynucleotide (non-viral) based vector and vector systems.

Naked Polynucleotides

[0200] In some embodiments one or more programmable pattern recognition composition or system polynucleotides described elsewhere herein can be included in a naked polynucleotide. The term of art "naked polynucleotide" as used herein refers to polynucleotides that are not associated with another molecule (e.g., proteins, lipids, and/or other molecules) that can often help protect it from environmental factors and/or degradation. As used herein, associated with includes, but is not limited to, linked to, adhered to, adsorbed to, enclosed in, enclosed in or within, mixed with, and the like. Naked polynucleotides that include one or more of the programmable pattern recognition composition or system polynucleotides described herein can be delivered directly to a host cell and optionally expressed therein. The naked polynucleotides can have any suitable two- and three-dimensional configurations. By way of non-limiting examples, naked polynucleotides can be single-stranded molecules, double stranded molecules, circular molecules (e.g., plasmids and artificial chromosomes), molecules

that contain portions that are single stranded and portions that are double stranded (e.g., ribozymes), and the like. In some embodiments, the naked polynucleotide contains only the programmable pattern recognition composition or system polynucleotide(s) of the present invention. In some embodiments, the naked polynucleotide can contain other nucleic acids and/or polynucleotides in addition to the programmable pattern recognition composition or system polynucleotide(s) of the present invention. The naked polynucleotides can include one or more elements of a transposon system. Transposons and system thereof are described in greater detail elsewhere herein.

Non-Viral Polynucleotide Vectors

[0201] In some embodiments, one or more of the programmable pattern recognition composition or system polynucleotides can be included in a non-viral polynucleotide vector. Suitable non-viral polynucleotide vectors include, but are not limited to, transposon vectors and vector systems, plasmids, bacterial artificial chromosomes, yeast artificial chromosomes, AR(antibiotic resistance)-free plasmids and miniplasmids, circular covalently closed vectors (e.g. minicircles, minivectors, miniknots,), linear covalently closed vectors ("dumbbell shaped"), MIDGE (minimalistic immunologically defined gene expression) vectors, MiLV (micro-linear vector) vectors, Ministrings, mini-intronic plasmids, PSK systems (post-segregationally killing systems), ORT (operator repressor titration) plasmids, and the like. See e.g., Hardee et al. 2017. Genes. 8(2):65.

[0202] In some embodiments, the non-viral polynucleotide vector can have a conditional origin of replication. In some embodiments, the non-viral polynucleotide vector can be an ORT plasmid. In some embodiments, the non-viral polynucleotide vector can have a minimalistic immunologically defined gene expression. In some embodiments, the non-viral polynucleotide vector can have one or more post-segregationally killing system genes. In some embodiments, the non-viral polynucleotide vector is AR-free. In some embodiments, the non-viral polynucleotide vector is a minivector. In some embodiments, the non-viral polynucleotide vector includes a nuclear localization signal. In some embodiments, the non-viral polynucleotide vector can include one or more CpG motifs. In some embodiments, the non-viral polynucleotide vectors can include one or more scaffold/matrix attachment regions (S/MARs). See e.g., Mirkovitch et al. 1984. Cell. 39:223-232, Wong et al. 2015. Adv. Genet. 89:113-152, whose techniques and vectors can be adapted for use in the present invention. S/MARs are AT-rich sequences that play a role in the spatial organization of chromosomes

through DNA loop base attachment to the nuclear matrix. S/MARs are often found close to regulatory elements such as promoters, enhancers, and origins of DNA replication. Inclusion of one or S/MARs can facilitate a once-per-cell-cycle replication to maintain the non-viral polynucleotide vector as an episome in daughter cells. In certain embodiments, the S/MAR sequence is located downstream of an actively transcribed polynucleotide (e.g., one or more CRISPR-Cas system polynucleotides of the present invention) included in the non-viral polynucleotide vector. In some embodiments, the S/MAR can be a S/MAR from the beta-interferon gene cluster. See e.g. Verghese et al. 2014. Nucleic Acid Res. 42:e53; Xu et al. 2016. Sci. China Life Sci. 59:1024-1033; Jin et al. 2016. 8:702-711; Koirala et al. 2014. Adv. Exp. Med. Biol. 801:703-709; and Nehlsen et al. 2006. Gene Ther. Mol. Biol. 10:233-244, whose techniques and vectors can be adapted for use in the present invention.

[0203] In some embodiments, the non-viral vector is a transposon vector or system thereof. As used herein, "transposon" (also referred to as transposable element) refers to a polynucleotide sequence that is capable of moving form location in a genome to another. There are several classes of transposons. Transposons include retrotransposons and DNA transposons. Retrotransposons require the transcription of the polynucleotide that is moved (or transposed) in order to transpose the polynucleotide to a new genome or polynucleotide. DNA transposons are those that do not require reverse transcription of the polynucleotide that is moved (or transposed) in order to transpose the polynucleotide to a new genome or polynucleotide. In some embodiments, the non-viral polynucleotide vector can be a retrotransposon vector. In some embodiments, the retrotransposon vector includes long terminal repeats. In some embodiments, the retrotransposon vector does not include long terminal repeats. In some embodiments, the non-viral polynucleotide vector can be a DNA transposon vector. DNA transposon vectors can include a polynucleotide sequence encoding a transposase. In some embodiments, the transposon vector is configured as a non-autonomous transposon vector, meaning that the transposition does not occur spontaneously on its own. In some of these embodiments, the transposon vector lacks one or more polynucleotide sequences encoding proteins required for transposition. In some embodiments, the non-autonomous transposon vectors lack one or more Ac elements.

[0204] In some embodiments a non-viral polynucleotide transposon vector system can include a first polynucleotide vector that contains the programmable pattern recognition composition or system polynucleotide(s) of the present invention flanked on the 5' and 3' ends

by transposon terminal inverted repeats (TIRs) and a second polynucleotide vector that includes a polynucleotide capable of encoding a transposase coupled to a promoter to drive expression of the transposase. When both are expressed in the same cell the transposase can be expressed from the second vector and can transpose the material between the TIRs on the first vector (e.g., the programmable pattern recognition composition or system polynucleotide(s) of the present invention) and integrate it into one or more positions in the host cell's genome. In some embodiments the transposon vector or system thereof can be configured as a gene trap. In some embodiments, the TIRs can be configured to flank a strong splice acceptor site followed by a reporter and/or other gene (e.g., one or more of the programmable pattern recognition composition or system polynucleotide(s) of the present invention) and a strong poly A tail. When transposition occurs while using this vector or system thereof, the transposon can insert into an intron of a gene and the inserted reporter or other gene can provoke a mis-splicing process and as a result it in activates the trapped gene.

[0205] Any suitable transposon system can be used. Suitable transposon and systems thereof can include, Sleeping Beauty transposon system (Tc1/mariner superfamily) (see e.g. Ivics et al. 1997. Cell. 91(4): 501-510), piggyBac (piggyBac superfamily) (see e.g., Li et al. 2013 110(25): E2279-E2287 and Yusa et al. 2011. PNAS. 108(4): 1531-1536), Tol2 (superfamily hAT), Frog Prince (Tc1/mariner superfamily) (see e.g., Miskey et al. 2003 Nucleic Acid Res. 31(23):6873-6881) and variants thereof.

Delivery of the Polynucleotides, Vectors, and Vector Systems

[0206] The polynucleotides, vectors, and/or vector systems can be delivered, such as to a cell or cells, by any suitable method or technique. In some embodiments, delivery can include association or otherwise incorporating the polynucleotides, vectors and/or vector systems with one or more delivery vehicles. Exemplary delivery methods and vehicles are discussed in greater detail below.

Physical Delivery

[0207] In some embodiments, the polynucleotides, vectors, and vector systems or any delivery vehicle containing the same may be introduced to cells by physical delivery methods. Examples of physical methods include microinjection, electroporation, and hydrodynamic delivery. Both nucleic acid and proteins may be delivered using such methods. For example, proteins of the present invention may be prepared in vitro, isolated, (refolded, purified if needed), and introduced to cells.

Microinjection

[0208] Microinjection of the cargo directly to cells can achieve high efficiency, e.g., above 90% or about 100%. In some embodiments, microinjection may be performed using a microscope and a needle (e.g., with 0.5–5.0 μm in diameter) to pierce a cell membrane and deliver the cargo directly to a target site within the cell. Microinjection may be used for *in vitro* and *ex vivo* delivery.

[0209] Plasmids comprising coding sequences for proteins of the programmable pattern recognition composition or system and/or guide RNAs, mRNAs, and/or guide RNAs, may be microinjected. In some cases, microinjection may be used i) to deliver DNA directly to a cell nucleus, and/or ii) to deliver mRNA (e.g., in vitro transcribed) to a cell nucleus or cytoplasm. In certain examples, microinjection may be used to delivery sgRNA directly to the nucleus and programmable pattern recognition composition or system polypeptide-encoding mRNA to the cytoplasm, e.g., facilitating translation and shuttling of said polypeptides or polynucleotides to the nucleus.

[0210] Microinjection may be used to generate genetically modified animals. For example, gene editing cargos may be injected into zygotes to allow for efficient germline modification. Such approach can yield normal embryos and full-term mouse pups harboring the desired modification(s). Microinjection can also be used to provide transiently up- or down- regulate a specific gene within the genome of a cell, e.g., using CRISPRa and CRISPRi.

Electroporation

[0211] In some embodiments, the programmable pattern recognition composition or system polypeptide or polynucleoitdes and/or delivery vehicles may be delivered by electroporation. Electroporation may use pulsed high-voltage electrical currents to transiently open nanometer-sized pores within the cellular membrane of cells suspended in buffer, allowing for components with hydrodynamic diameters of tens of nanometers to flow into the cell. In some cases, electroporation may be used on various cell types and efficiently transfer cargo into cells. Electroporation may be used for *in vitro* and *ex vivo* delivery.

[0212] Electroporation may also be used to deliver the cargo to into the nuclei of mammalian cells by applying specific voltage and reagents, e.g., by nucleofection. Such approaches include those described in Wu Y, et al. (2015). Cell Res 25:67–79; Ye L, et al. (2014). Proc Natl Acad Sci USA 111:9591–6; Choi PS, Meyerson M. (2014). Nat Commun 5:3728; Wang J, Quake SR. (2014). Proc Natl Acad Sci 111:13157–62. Electroporation may

also be used to deliver the cargo in vivo, e.g., with methods described in Zuckermann M, et al. (2015). Nat Commun 6:7391.

Hydrodynamic Delivery

[0213] Hydrodynamic delivery may also be used for delivering the programmable pattern recognition composition or system polypeptides and/or polynucleotides, e.g., for *in vivo* delivery. In some examples, hydrodynamic delivery may be performed by rapidly pushing a large volume (8–10% body weight) solution containing the gene editing cargo into the bloodstream of a subject (e.g., an animal or human), e.g., for mice, via the tail vein. As blood is incompressible, the large bolus of liquid may result in an increase in hydrodynamic pressure that temporarily enhances permeability into endothelial and parenchymal cells, allowing for cargo not normally capable of crossing a cellular membrane to pass into cells. This approach may be used for delivering naked DNA plasmids and proteins. The delivered cargos may be enriched in liver, kidney, lung, muscle, and/or heart.

Transfection

[0214] The programmable pattern recognition composition or system polypeptides and/or polynucleotides, may be introduced to cells by transfection methods for introducing nucleic acids into cells. Examples of transfection methods include calcium phosphate-mediated transfection, cationic transfection, liposome transfection, dendrimer transfection, heat shock transfection, magnetofection, lipofection, impalefection, optical transfection, proprietary agent-enhanced uptake of nucleic acid.

Transduction

[0215] The programmable pattern recognition composition or system polypeptides and/or polynucleotides can be introduced to cells by transduction by a viral or pseudoviral particle. Methods of packaging the cargos in viral particles can be accomplished using any suitable viral vector or vector systems. Such viral vector and vector systems are described in greater detail elsewhere herein. As used in this context herein “transduction” refers to the process by which foreign nucleic acids and/or proteins are introduced to a cell (prokaryote or eukaryote) by a viral or pseudo viral particle. After packaging in a viral particle or pseudo viral particle, the viral particles can be exposed to cells (e.g., in vitro, ex vivo, or in vivo) where the viral or pseudoviral particle infects the cell and delivers the cargo to the cell via transduction. Viral and pseudoviral particles can be optionally concentrated prior to exposure to target cells. In some

embodiments, the virus titer of a composition containing viral and/or pseudoviral particles can be obtained and a specific titer be used to transduce cells.

Biolistics

[0216] The programmable pattern recognition composition or system polypeptides and/or polynucleotides can be introduced to cells using a biolistic method or technique. The term of art "biolistic", as used herein refers to the delivery of nucleic acids to cells by high-speed particle bombardment. In some embodiments, the cargo(s) can be attached, associated with, or otherwise coupled to particles, which than can be delivered to the cell via a gene-gun (see e.g., Liang et al. 2018. Nat. Protocol. 13:413-430; Svitashev et al. 2016. Nat. Comm. 7:13274; Ortega-Escalante et al., 2019. Plant. J. 97:661-672). In some embodiments, the particles can be gold, tungsten, palladium, rhodium, platinum, or iridium particles.

Implantable Devices

[0217] In some embodiments, the delivery system includes an implantable device that incorporates or is coated with a programmable pattern recognition composition or system polypeptides and/or polynucleotides described herein. Various implantable devices are described in the art, and include any device, graft, or other composition that can be implanted into a subject.

Delivery Vehicles

[0218] The delivery systems may comprise one or more delivery vehicles. The delivery vehicles may deliver the cargo into cells, tissues, organs, or organisms (e.g., animals or plants). The cargos may be packaged, carried, or otherwise associated with the delivery vehicles. The delivery vehicles may be selected based on the types of cargo to be delivered, and/or the delivery is *in vitro* and/or *in vivo*. Examples of delivery vehicles include vectors, viruses (e.g., virus particles), non-viral vehicles, and other delivery reagents described herein.

[0219] The delivery vehicles in accordance with the present invention may a greatest dimension (e.g., diameter) of less than 100 microns (μm). In some embodiments, the delivery vehicles have a greatest dimension of less than 10 μm. In some embodiments, the delivery vehicles may have a greatest dimension of less than 2000 nanometers (nm). In some embodiments, the delivery vehicles may have a greatest dimension of less than 1000 nanometers (nm). In some embodiments, the delivery vehicles may have a greatest dimension (e.g., diameter) of less than 900 nm, less than 800 nm, less than 700 nm, less than 600 nm, less than 500 nm, less than 400 nm, less than 300 nm, less than 200 nm, less than 150nm, or less

than 100nm, less than 50nm. In some embodiments, the delivery vehicles may have a greatest dimension ranging between 25 nm and 200 nm.

[0220] In some embodiments, the delivery vehicles may be or comprise particles. For example, the delivery vehicle may be or comprise nanoparticles (e.g., particles with a greatest dimension (e.g., diameter) no greater than 1000 nm. The particles may be provided in different forms, e.g., as solid particles (e.g., metal such as silver, gold, iron, titanium), non-metal, lipid-based solids, polymers), suspensions of particles, or combinations thereof. Metal, dielectric, and semiconductor particles may be prepared, as well as hybrid structures (e.g., core–shell particles).

[0221] Nanoparticles may also be used to deliver the compositions and systems to plant cells, e.g., as described in WO 2008042156, US 20130185823, and WO2015089419. In general, a "nanoparticle" refers to any particle having a diameter of less than 1000 nm. In certain preferred embodiments, nanoparticles of the invention have a greatest dimension (e.g., diameter) of 500 nm or less. In other preferred embodiments, nanoparticles of the invention have a greatest dimension ranging between 25 nm and 200 nm. In other preferred embodiments, nanoparticles of the invention have a greatest dimension of 100 nm or less. In other preferred embodiments, nanoparticles of the invention have a greatest dimension ranging between 35 nm and 60 nm. It will be appreciated that reference made herein to particles or nanoparticles can be interchangeable, where appropriate. Nanoparticles made of semiconducting material may also be labeled quantum dots if they are small enough (typically sub 10 nm) that quantization of electronic energy levels occurs. Such nanoscale particles are used in biomedical applications as drug carriers or imaging agents and may be adapted for similar purposes in the present invention. Semi-solid and soft nanoparticles have been manufactured and are within the scope of the present invention. Nanoparticles with one half hydrophilic and the other half hydrophobic are termed Janus particles and are particularly effective for stabilizing emulsions. They can self-assemble at water/oil interfaces and act as solid surfactants.

[0222] Particle characterization (including e.g., characterizing morphology, dimension, etc.) is done using a variety of different techniques. Common techniques are electron microscopy (TEM, SEM), atomic force microscopy (AFM), dynamic light scattering (DLS), X-ray photoelectron spectroscopy (XPS), powder X-ray diffraction (XRD), Fourier transform infrared spectroscopy (FTIR), matrix-assisted laser desorption/ionization time-of-flight mass spectrometry(MALDI-TOF), ultraviolet-visible spectroscopy, dual polarization interferometry

and nuclear magnetic resonance (NMR). Characterization (dimension measurements) may be made as to native particles (i.e., preloading) or after loading of the cargo (herein cargo refers to e.g., one or more components of CRISPR-Cas system e.g., CRISPR enzyme or mRNA or guide RNA, or any combination thereof, and may include additional carriers and/or excipients) to provide particles of an optimal size for delivery for any *in vitro*, *ex vivo* and/or in *vivo* application of the present invention. In certain preferred embodiments, particle dimension (e.g., diameter) characterization is based on measurements using dynamic laser scattering (DLS). Mention is made of US Patent No. 8,709,843; US Patent No. 6,007,845; US Patent No. 5,855,913; US Patent No. 5,985,309; US. Patent No. 5,543,158; and the publication by James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84, describing particles, methods of making and using them and measurements thereof.

Vector Based Delivery Vehicles

[0223] Vectors and Vector systems that can be used to deliver programmable pattern recognition composition or system polypeptides and/or polynucleotides are described in greater detail elsewhere herein.

Non-Vector Delivery Vehicles

[0224] The delivery vehicles may comprise non-viral vehicles. In general, methods and vehicles capable of delivering nucleic acids and/or proteins may be used for delivering the systems compositions herein. Examples of non-viral vehicles include lipid nanoparticles, cell-penetrating peptides (CPPs), DNA nanoclews, metal nanoparticles, streptolysin O, multifunctional envelope-type nanodevices (MENDs), lipid-coated mesoporous silica particles, and other inorganic nanoparticles.

Lipid Particles

[0225] The delivery vehicles may comprise lipid particles, e.g., lipid nanoparticles (LNPs) and liposomes. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, International Patent Publication Nos. WO 91/17424 and WO 91/16024. 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).

Lipid nanoparticles (LNPs)

[0226] LNPs may encapsulate nucleic acids within cationic lipid particles (e.g., liposomes), and may be delivered to cells with relative ease. In some examples, lipid nanoparticles do not contain any viral components, which helps minimize safety and immunogenicity concerns. Lipid particles may be used for *in vitro*, *ex vivo*, and *in vivo* deliveries. Lipid particles may be used for various scales of cell populations.

[0227] In some examples. LNPs may be used for delivering DNA molecules (e.g., those comprising coding sequences of Cas and/or gRNA) and/or RNA molecules (e.g., mRNA of Cas, gRNAs). In certain cases, LNPs may be use for delivering RNP complexes of Cas/gRNA.

[0228] Components in LNPs may comprise cationic lipids 1,2- dilineoyl-3-dimethylammonium-propane (DLinDAP), l,2-dilinoleyloxy-3-N,N- dimethylaminopropane (DLinDMA), l,2-dilinoleyloxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), l,2-dilinoleyl-4-(2-dimethylaminoethyl)-[l,3]-dioxolane (DLinKC2-DMA), (3- o-[2"-(methoxypolyethyleneglycol 2000) succinoyl]-l,2-dimyristoyl-sn-glycol (PEG-S-DMG), R-3-[(ro-methoxy-poly(ethylene glycol)2000) carbamoyl]-l,2-dimyristyloxlpropyl-3-amine (PEG-C-DOMG, and any combination thereof. Preparation of LNPs and encapsulation may be adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011).

[0229] In some embodiments, an LNP delivery vehicle can be used to deliver a virus particle containing a CRISPR-Cas system and/or component(s) thereof. In some embodiments, the virus particle(s) can be adsorbed to the lipid particle, such as through electrostatic interactions, and/or can be attached to the liposomes via a linker.

[0230] In some embodiments, the LNP contains a nucleic acid, wherein the charge ratio of nucleic acid backbone phosphates to cationic lipid nitrogen atoms is about 1: 1.5 – 7 or about 1:4.

[0231] In some embodiments, the LNP also includes a shielding compound, which is removable from the lipid composition under in vivo conditions. In some embodiments, the shielding compound is a biologically inert compound. In some embodiments, the shielding compound does not carry any charge on its surface or on the molecule as such. In some

embodiments, the shielding compounds are polyethylenglycoles (PEGs), hydroxyethylglucose (HEG) based polymers, polyhydroxyethyl starch (polyHES) and polypropylene. In some embodiments, the PEG, HEG, polyHES, and a polypropylene weight between about 500 to 10,000 Da or between about 2000 to 5000 Da. In some embodiments, the shielding compound is PEG2000 or PEG5000.

[0232] In some embodiments, the LNP can include one or more helper lipids. In some embodiments, the helper lipid can be a phosphor lipid or a steroid. In some embodiments, the helper lipid is between about 20 mol % to 80 mol % of the total lipid content of the composition. In some embodiments, the helper lipid component is between about 35 mol % to 65 mol % of the total lipid content of the LNP. In some embodiments, the LNP includes lipids at 50 mol% and the helper lipid at 50 mol% of the total lipid content of the LNP.

[0233] Other non-limiting, exemplary LNP delivery vehicles are described in U.S. Patent Publication Nos. US 20160174546, US 20140301951, US 20150105538, US 20150250725, Wang et al., J. Control Release, 2017 Jan 31. pii: S0168-3659(17)30038-X. doi: 10.1016/j.jconrel.2017.01.037. [Epub ahead of print]; Altınoğlu et al., Biomater Sci., 4(12):1773-80, Nov. 15, 2016; Wang et al., PNAS, 113(11):2868-73 March 15, 2016; Wang et al., PloS One, 10(11): e0141860. doi: 10.1371/journal.pone.0141860. eCollection 2015, Nov. 3, 2015; Takeda et al., Neural Regen Res. 10(5):689-90, May 2015; Wang et al., Adv. Healthc Mater., 3(9):1398-403, Sep. 2014; and Wang et al., Agnew Chem Int Ed Engl., 53(11):2893-8, Mar. 10, 2014; James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84; Coelho et al., N Engl J Med 2013; 369:819-29; Aleku *et al.*, *Cancer Res.*, 68(23): 9788-98 (Dec. 1, 2008), Strumberg *et al.*, *Int. J. Clin. Pharmacol. Ther.*, 50(1): 76-8 (Jan. 2012), Schultheis *et al.*, *J. Clin. Oncol.*, 32(36): 4141-48 (Dec. 20, 2014), and Fehring *et al.*, *Mol. Ther.*, 22(4): 811-20 (Apr. 22, 2014); Novobrantseva, Molecular Therapy–Nucleic Acids (2012) 1, e4; doi:10.1038/mtna.2011.3; WO2012135025; US 20140348900; US 20140328759; US 20140308304; WO 2005/105152; WO 2006/069782; WO 2007/121947; US 2015/082080; US 20120251618; 7,982,027; 7,799,565; 8,058,069; 8,283,333; 7,901,708; 7,745,651; 7,803,397; 8,101,741; 8,188,263; 7,915,399; 8,236,943 and 7,838,658 and European Pat. Nos 1766035; 1519714; 1781593 and 1664316.

Liposomes

[0234] In some embodiments, a lipid particle may be liposome. Liposomes are spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. In some embodiments, liposomes are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes and the blood brain barrier (BBB).

[0235] Liposomes can be made from several different types of lipids, e.g., phospholipids. A liposome may comprise natural phospholipids and lipids such as l,2-distearoryl-sn-glycero-3 -phosphatidyl choline (DSPC), sphingomyelin, egg phosphatidylcholines, monosialoganglioside, or any combination thereof.

[0236] Several other additives may be added to liposomes in order to modify their structure and properties. For instance, liposomes may further comprise cholesterol, sphingomyelin, and/or l,2-dioleoyl-sn-glycero-3- phosphoethanolamine (DOPE), e.g., to increase stability and/or to prevent the leakage of the liposomal inner cargo.

[0237] In some embodiments, a liposome delivery vehicle can be used to deliver a virus particle containing a CRISPR-Cas system and/or component(s) thereof. In some embodiments, the virus particle(s) can be adsorbed to the liposome, such as through electrostatic interactions, and/or can be attached to the liposomes via a linker.

[0238] In some embodiments, the liposome can be a Trojan Horse liposome (also known in the art as Molecular Trojan Horses), see e.g. http://cshprotocols.cshlp.org/content/2010/4/pdb.prot5407.long, the teachings of which can be applied and/or adapted to generated and/or deliver the CRISPR-Cas systems described herein.

[0239] Other non-limiting, exemplary liposomes can be those as set forth in Wang et al., ACS Synthetic Biology, 1, 403-07 (2012); Wang et al., PNAS, 113(11) 2868-2873 (2016); Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679; WO 2008/042973; US Pat. No. 8,071,082; WO 2014/186366; 20160257951; US20160129120; US 20160244761; 20120251618; WO2013/093648; Lipofectin (a combination of DOTMA and DOPE), Lipofectase, LIPOFECTAMINE.RTM. (e.g., LIPOFECTAMINE.RTM. 2000, LIPOFECTAMINE.RTM. 3000, LIPOFECTAMINE.RTM. RNAiMAX, LIPOFECTAMINE.RTM. LTX), SAINT-RED

(Synvolux Therapeutics, Groningen Netherlands), DOPE, Cytofectin (Gilead Sciences, Foster City, Calif.), and Eufectins (JBL, San Luis Obispo, Calif.).

Stable nucleic-acid-lipid particles (SNALPs)

[0240] In some embodiments, the lipid particles may be stable nucleic acid lipid particles (SNALPs). SNALPs may comprise an ionizable lipid (DLinDMA) (e.g., cationic at low pH), a neutral helper lipid, cholesterol, a diffusible polyethylene glycol (PEG)-lipid, or any combination thereof. In some examples, SNALPs may comprise synthetic cholesterol, dipalmitoylphosphatidylcholine, 3-N-[(w-methoxy polyethylene glycol)2000)carbamoyl]-l,2-dimyrestyloxypropylamine, and cationic 1,2-dilinoleyloxy-3-N,Ndimethylaminopropane. In some examples, SNALPs may comprise synthetic cholesterol, 1,2-distearoyl-sn-glycero-3-phosphocholine, PEG- cDMA, and 1,2-dilinoleyloxy-3-(N;N-dimethyl)aminopropane (DLinDMAo).

[0241] Other non-limiting, exemplary SNALPs that can be used to deliver the CRISPR-Cas systems described herein can be any such SNALPs as described in Morrissey et al., Nature Biotechnology, Vol. 23, No. 8, August 2005, Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006; Geisbert et al., Lancet 2010; 375: 1896-905; Judge, J. Clin. Invest. 119:661-673 (2009); and Semple et al., Nature Niotechnology, Volume 28 Number 2 February 2010, pp. 172-177.

Other Lipids

[0242] The lipid particles may also comprise one or more other types of lipids, e.g., cationic lipids, such as amino lipid 2,2-dilinoleyl-4-dimethylaminoethyl-[l,3]- dioxolane (DLin-KC2-DMA), DLin-KC2-DMA4, C12- 200 and colipids disteroylphosphatidyl choline, cholesterol, and PEG-DMG.

[0243] In some embodiments, the delivery vehicle can be or include a lipidoid, such as any of those set forth in, for example, US 20110293703.

[0244] In some embodiments, the delivery vehicle can be or include an amino lipid, such as any of those set forth in, for example, Jayaraman, Angew. Chem. Int. Ed. 2012, 51, 8529 – 8533.

[0245] In some embodiments, the delivery vehicle can be or include a lipid envelope, such as any of those set forth in, for example, Korman et al., 2011. Nat. Biotech. 29:154-157.

Lipoplexes/polyplexes

[0246] In some embodiments, the delivery vehicles comprise lipoplexes and/or polyplexes. Lipoplexes may bind to negatively charged cell membrane and induce endocytosis into the cells. Examples of lipoplexes may be complexes comprising lipid(s) and non-lipid components. Examples of lipoplexes and polyplexes include FuGENE-6 reagent, a non-liposomal solution containing lipids and other components, zwitterionic amino lipids (ZALs), Ca2þ (e.g., forming DNA/Ca^{2+} microcomplexes), polyethenimine (PEI) (e.g., branched PEI), and poly(L-lysine) (PLL).

Sugar-Based Particles

[0247] In some embodiments, the delivery vehicle can be a sugar-based particle. In some embodiments, the sugar-based particles can be or include GalNAc, such as any of those described in WO2014118272; US 20020150626; Nair, JK et al., 2014, Journal of the American Chemical Society 136 (49), 16958-16961; Østergaard et al., Bioconjugate Chem., 2015, 26 (8), pp 1451–1455;

Cell Penetrating Peptides

[0248] In some embodiments, the delivery vehicles comprise cell penetrating peptides (CPPs). CPPs are short peptides that facilitate cellular uptake of various molecular cargo (e.g., from nanosized particles to small chemical molecules and large fragments of DNA).

[0249] CPPs may be of different sizes, amino acid sequences, and charges. In some examples, CPPs can translocate the plasma membrane and facilitate the delivery of various molecular cargoes to the cytoplasm or an organelle. CPPs may be introduced into cells via different mechanisms, e.g., direct penetration in the membrane, endocytosis-mediated entry, and translocation through the formation of a transitory structure.

[0250] CPPs may have an amino acid composition that either contains a high relative abundance of positively charged amino acids such as lysine or arginine or has sequences that contain an alternating pattern of polar/charged amino acids and non-polar, hydrophobic amino acids. These two types of structures are referred to as polycationic or amphipathic, respectively. A third class of CPPs are the hydrophobic peptides, containing only apolar residues, with low net charge or have hydrophobic amino acid groups that are crucial for cellular uptake. Another type of CPPs is the trans-activating transcriptional activator (Tat) from Human Immunodeficiency Virus 1 (HIV-1). Examples of CPPs include to Penetratin, Tat (48-60), Transportan, and (R-AhX-R4) (Ahx refers to aminohexanoyl), Kaposi fibroblast growth factor

(FGF) signal peptide sequence, integrin β3 signal peptide sequence, polyarginine peptide Args sequence, Guanine rich-molecular transporters, and sweet arrow peptide. Examples of CPPs and related applications also include those described in US Patent 8,372,951.

[0251] CPPs can be used for *in vitro* and *ex vivo* work quite readily, and extensive optimization for each cargo and cell type is usually required. In some examples, CPPs may be covalently attached to the Cas protein directly, which is then complexed with the gRNA and delivered to cells. In some examples, separate delivery of CPP–Cas and CPP–gRNA to multiple cells may be performed. CPP may also be used to delivery RNPs.

[0252] CPPs may be used to deliver the compositions and systems to plants. In some examples, CPPs may be used to deliver the components to plant protoplasts, which are then regenerated to plant cells and further to plants.

DNA Nanoclews

[0253] In some embodiments, the delivery vehicles comprise DNA nanoclews. A DNA nanoclew refers to a sphere-like structure of DNA (e.g., with a shape of a ball of yarn). The nanoclew may be synthesized by rolling circle amplification with palindromic sequences that aide in the self-assembly of the structure. The sphere may then be loaded with a payload. An example of DNA nanoclew is described in Sun W et al, J Am Chem Soc. 2014 Oct 22;136(42):14722-5; and Sun W et al, Angew Chem Int Ed Engl. 2015 Oct 5;54(41):12029-33. DNA nanoclew may have a palindromic sequences to be partially complementary to the gRNA within the Cas:gRNA ribonucleoprotein complex. A DNA nanoclew may be coated, e.g., coated with PEI to induce endosomal escape.

Metal Nanoparticles

[0254] In some embodiments, the delivery vehicles comprise gold nanoparticles (also referred to AuNPs or colloidal gold). Gold nanoparticles may form complex with cargos, e.g., Cas:gRNA RNP. Gold nanoparticles may be coated, e.g., coated in a silicate and an endosomal disruptive polymer, PAsp(DET). Examples of gold nanoparticles include AuraSense Therapeutics' Spherical Nucleic Acid (SNA™) constructs, and those described in Mout R, et al. (2017). ACS Nano 11:2452–8; Lee K, et al. (2017). Nat Biomed Eng 1:889–901. Other metal nanoparticles can also be complexed with cargo(s). Such metal particles include, tungsten, palladium, rhodium, platinum, and iridium particles. Other non-limiting, exemplary metal nanoparticles are described in US 20100129793.

iTOP

[0255] In some embodiments, the delivery vehicles comprise iTOP. iTOP refers to a combination of small molecules drives the highly efficient intracellular delivery of native proteins, independent of any transduction peptide. iTOP may be used for induced transduction by osmocytosis and propanebetaine, using NaCl-mediated hyperosmolality together with a transduction compound (propanebetaine) to trigger macropinocytotic uptake into cells of extracellular macromolecules. Examples of iTOP methods and reagents include those described in D'Astolfo DS, Pagliero RJ, Pras A, et al. (2015). Cell 161:674–690.

Polymer-based Particles

[0256] In some embodiments, the delivery vehicles may comprise polymer-based particles (e.g., nanoparticles). In some embodiments, the polymer-based particles may mimic a viral mechanism of membrane fusion. The polymer-based particles may be a synthetic copy of Influenza virus machinery and form transfection complexes with various types of nucleic acids (siRNA, miRNA, plasmid DNA or shRNA, mRNA) that cells take up via the endocytosis pathway, a process that involves the formation of an acidic compartment. The low pH in late endosomes acts as a chemical switch that renders the particle surface hydrophobic and facilitates membrane crossing. Once in the cytosol, the particle releases its payload for cellular action. This Active Endosome Escape technology is safe and maximizes transfection efficiency as it is using a natural uptake pathway. In some embodiments, the polymer-based particles may comprise alkylated and carboxyalkylated branched polyethylenimine. In some examples, the polymer-based particles are VIROMER, e.g., VIROMER RNAi, VIROMER RED, VIROMER mRNA, VIROMER CRISPR. Example methods of delivering the systems and compositions herein include those described in Bawage SS et al., Synthetic mRNA expressed Cas13a mitigates RNA virus infections, www.biorxiv.org/content/10.1101/370460v1.full doi: doi.org/10.1101/370460, Viromer® RED, a powerful tool for transfection of keratinocytes. doi: 10.13140/RG.2.2.16993.61281, Viromer® Transfection - Factbook 2018: technology, product overview, users' data., doi:10.13140/RG.2.2.23912.16642. Other exemplary and non-limiting polymeric particles are described in US 20170079916, US 20160367686, US 20110212179, US 20130302401, 6,007,845, 5,855,913, 5,985,309, 5,543,158, WO2012135025, US 20130252281, US 20130245107, US 20130244279; US 20050019923, 20080267903;

Streptolysin O (SLO)

[0257] The delivery vehicles may be streptolysin O (SLO). SLO is a toxin produced by Group A streptococci that works by creating pores in mammalian cell membranes. SLO may act in a reversible manner, which allows for the delivery of proteins (e.g., up to 100 kDa) to the cytosol of cells without compromising overall viability. Examples of SLO include those described in Sierig G, et al. (2003). Infect Immun 71:446–55; Walev I, et al. (2001). Proc Natl Acad Sci U S A 98:3185–90; Teng KW, et al. (2017). Elife 6:e25460.

Multifunctional Envelope-Type Nanodevice (MEND)

[0258] The delivery vehicles may comprise multifunctional envelope-type nanodevice (MENDs). MENDs may comprise condensed plasmid DNA, a PLL core, and a lipid film shell. A MEND may further comprise cell-penetrating peptide (e.g., stearyl octaarginine). The cell penetrating peptide may be in the lipid shell. The lipid envelope may be modified with one or more functional components, e.g., one or more of: polyethylene glycol (e.g., to increase vascular circulation time), ligands for targeting of specific tissues/cells, additional cell-penetrating peptides (e.g., for greater cellular delivery), lipids to enhance endosomal escape, and nuclear delivery tags. In some examples, the MEND may be a tetra-lamellar MEND (T-MEND), which may target the cellular nucleus and mitochondria. In certain examples, a MEND may be a PEG-peptide-DOPE-conjugated MEND (PPD-MEND), which may target bladder cancer cells. Examples of MENDs include those described in Kogure K, et al. (2004). J Control Release 98:317–23; Nakamura T, et al. (2012). Acc Chem Res 45:1113–21.

Lipid-coated mesoporous silica particles

[0259] The delivery vehicles may comprise lipid-coated mesoporous silica particles. Lipid-coated mesoporous silica particles may comprise a mesoporous silica nanoparticle core and a lipid membrane shell. The silica core may have a large internal surface area, leading to high cargo loading capacities. In some embodiments, pore sizes, pore chemistry, and overall particle sizes may be modified for loading different types of cargos. The lipid coating of the particle may also be modified to maximize cargo loading, increase circulation times, and provide precise targeting and cargo release. Examples of lipid-coated mesoporous silica particles include those described in Du X, et al. (2014). Biomaterials 35:5580–90; Durfee PN, et al. (2016). ACS Nano 10:8325–45.

Inorganic nanoparticles

[0260] The delivery vehicles may comprise inorganic nanoparticles. Examples of inorganic nanoparticles include carbon nanotubes (CNTs) (e.g., as described in Bates K and Kostarelos K. (2013). Adv Drug Deliv Rev 65:2023–33.), bare mesoporous silica nanoparticles (MSNPs) (e.g., as described in Luo GF, et al. (2014). Sci Rep 4:6064), and dense silica nanoparticles (SiNPs) (as described in Luo D and Saltzman WM. (2000). Nat Biotechnol 18:893–5).

Exosomes

[0261] The delivery vehicles may comprise exosomes. Exosomes include membrane bound extracellular vesicles, which can be used to contain and delivery various types of biomolecules, such as proteins, carbohydrates, lipids, and nucleic acids, and complexes thereof (e.g., RNPs). Examples of exosomes include those described in Schroeder A, et al., J Intern Med. 2010 Jan;267(1):9-21; El-Andaloussi S, et al., Nat Protoc. 2012 Dec;7(12):2112-26; Uno Y, et al., Hum Gene Ther. 2011 Jun;22(6):711-9; Zou W, et al., Hum Gene Ther. 2011 Apr;22(4):465-75.

[0262] In some examples, the exosome may form a complex (e.g., by binding directly or indirectly) to one or more components of the cargo. In certain examples, a molecule of an exosome may be fused with first adapter protein and a component of the cargo may be fused with a second adapter protein. The first and the second adapter protein may specifically bind each other, thus associating the cargo with the exosome. Examples of such exosomes include those described in Ye Y, et al., Biomater Sci. 2020 Apr 28. doi: 10.1039/d0bm00427h.

[0263] Other non-limiting, exemplary exosomes include any of those set forth in Alvarez-Erviti et al. 2011, Nat Biotechnol 29: 341; [1401] El-Andaloussi et al. (Nature Protocols 7:2112–2126(2012); and Wahlgren et al. (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130).

Spherical Nucleic Acids (SNAs)

[0264] In some embodiments, the delivery vehicle can be a SNA. SNAs are three dimensional nanostructures that can be composed of densely functionalized and highly oriented nucleic acids that can be covalently attached to the surface of spherical nanoparticle cores. The core of the spherical nucleic acid can impart the conjugate with specific chemical and physical properties, and it can act as a scaffold for assembling and orienting the oligonucleotides into a dense spherical arrangement that gives rise to many of their functional properties,

distinguishing them from all other forms of matter. In some embodiments, the core is a crosslinked polymer. Non-limiting, exemplary SNAs can be any of those set forth in Cutler et al., J. Am. Chem. Soc. 2011 133:9254-9257, Hao et al., Small. 2011 7:3158-3162, Zhang et al., ACS Nano. 2011 5:6962-6970, Cutler et al., J. Am. Chem. Soc. 2012 134:1376-1391, Young et al., Nano Lett. 2012 12:3867-71, Zheng et al., Proc. Natl. Acad. Sci. USA. 2012 109:11975-80, Mirkin, Nanomedicine 2012 7:635-638 Zhang et al., J. Am. Chem. Soc. 2012 134:16488-1691, Weintraub, Nature 2013 495:S14-S16, Choi et al., Proc. Natl. Acad. Sci. USA. 2013 110(19):7625-7630, Jensen et al., Sci. Transl. Med. 5, 209ra152 (2013) and Mirkin, et al., and Small, 10:186-192.

Self-Assembling Nanoparticles

[0265] In some embodiments, the delivery vehicle is a self-assembling nanoparticle. The self-assembling nanoparticles can contain one or more polymers. The self-assembling nanoparticles can be PEGylated. Self-assembling nanoparticles are known in the art. Non-limiting, exemplary self-assembling nanoparticles can any as set forth in Schiffelers et al., Nucleic Acids Research, 2004, Vol. 32, No. 19, Bartlett et al. (PNAS, September 25, 2007,vol. 104, no. 39; Davis et al., Nature, Vol 464, 15 April 2010.

Supercharged Proteins

[0266] In some embodiments, the delivery vehicle can be a supercharged protein. As used herein "Supercharged proteins" are a class of engineered or naturally occurring proteins with unusually high positive or negative net theoretical charge. Non-limiting, exemplary supercharged proteins can be any of those set forth in Lawrence et al., 2007, Journal of the American Chemical Society 129, 10110–10112.

Targeted Delivery

[0267] In some embodiments, the delivery vehicle can allow for targeted delivery to a specific cell, tissue, organ, or system. In such embodiments, the delivery vehicle can include one or more targeting moieties that can direct targeted delivery of the cargo(s). In an embodiment, the delivery vehicle comprises a targeting moiety, such as active targeting of a lipid entity of the invention, e.g., lipid particle or nanoparticle or liposome or lipid bilayer of the invention comprising a targeting moiety for active targeting.

[0268] With regard to targeting moieties, mention is made of Deshpande et al, "Current trends in the use of liposomes for tumor targeting," Nanomedicine (Lond). 8(9), doi:10.2217/nnm.13.118 (2013), and the documents it cites, all of which are incorporated

herein by reference and the teachings of which can be applied and/or adapted for targeted delivery of one or more CRISPR-Cas molecules described herein. Mention is also made of International Patent Publication No. WO 2016/027264, and the documents it cites, all of which are incorporated herein by reference, the teachings of which can be applied and/or adapted for targeted delivery of one or more CRISPR-Cas molecules described herein. And mention is made of Lorenzer et al, "Going beyond the liver: Progress and challenges of targeted delivery of siRNA therapeutics," Journal of Controlled Release, 203: 1–15 (2015), , and the documents it cites, all of which are incorporated herein by reference, the teachings of which can be applied and/or adapted for targeted delivery of one or more CRISPR-Cas molecules described herein.

[0269] An actively targeting lipid particle or nanoparticle or liposome or lipid bilayer delivery system (generally as to embodiments of the invention, "lipid entity of the invention" delivery systems) are prepared by conjugating targeting moieties, including small molecule ligands, peptides and monoclonal antibodies, on the lipid or liposomal surface; for example, certain receptors, such as folate and transferrin (Tf) receptors (TfR), are overexpressed on many cancer cells and have been used to make liposomes tumor cell specific. Liposomes that accumulate in the tumor microenvironment can be subsequently endocytosed into the cells by interacting with specific cell surface receptors. To efficiently target liposomes to cells, such as cancer cells, it is useful that the targeting moiety have an affinity for a cell surface receptor and to link the targeting moiety in sufficient quantities to have optimum affinity for the cell surface receptors; and determining these embodiments are within the ambit of the skilled artisan. In the field of active targeting, there are a number of cell-, e.g., tumor-, specific targeting ligands.

[0270] Also, as to active targeting, with regard to targeting cell surface receptors such as cancer cell surface receptors, targeting ligands on liposomes can provide attachment of liposomes to cells, e.g., vascular cells, via a noninternalizing epitope; and this can increase the extracellular concentration of that which is being delivered, thereby increasing the amount delivered to the target cells. A strategy to target cell surface receptors, such as cell surface receptors on cancer cells, such as overexpressed cell surface receptors on cancer cells, is to use receptor-specific ligands or antibodies. Many cancer cell types display upregulation of tumor-specific receptors. For example, TfRs and folate receptors (FRs) are greatly overexpressed by many tumor cell types in response to their increased metabolic demand. Folic acid can be used as a targeting ligand for specialized delivery owing to its ease of conjugation to nanocarriers, its high affinity for FRs and the relatively low frequency of FRs, in normal tissues as compared

with their overexpression in activated macrophages and cancer cells, e.g., certain ovarian, breast, lung, colon, kidney and brain tumors. Overexpression of FR on macrophages is an indication of inflammatory diseases, such as psoriasis, Crohn's disease, rheumatoid arthritis and atherosclerosis; accordingly, folate-mediated targeting of the invention can also be used for studying, addressing or treating inflammatory disorders, as well as cancers. Folate-linked lipid particles or nanoparticles or liposomes or lipid bylayers of the invention ("lipid entity of the invention") deliver their cargo intracellularly through receptor-mediated endocytosis. Intracellular trafficking can be directed to acidic compartments that facilitate cargo release, and, most importantly, release of the cargo can be altered or delayed until it reaches the cytoplasm or vicinity of target organelles. Delivery of cargo using a lipid entity of the invention having a targeting moiety, such as a folate-linked lipid entity of the invention, can be superior to nontargeted lipid entity of the invention. The attachment of folate directly to the lipid head groups may not be favorable for intracellular delivery of folate-conjugated lipid entity of the invention, since they may not bind as efficiently to cells as folate attached to the lipid entity of the invention surface by a spacer, which may can enter cancer cells more efficiently. A lipid entity of the invention coupled to folate can be used for the delivery of complexes of lipid, e.g., liposome, e.g., anionic liposome and virus or capsid or envelope or virus outer protein, such as those herein discussed such as adenovirous or AAV. Tf is a monomeric serum glycoprotein of approximately 80 KDa involved in the transport of iron throughout the body. Tf binds to the TfR and translocates into cells via receptor-mediated endocytosis. The expression of TfR can be higher in certain cells, such as tumor cells (as compared with normal cells and is associated with the increased iron demand in rapidly proliferating cancer cells. Accordingly, the invention comprehends a TfR-targeted lipid entity of the invention, e.g., as to liver cells, liver cancer, breast cells such as breast cancer cells, colon such as colon cancer cells, ovarian cells such as ovarian cancer cells, head, neck and lung cells, such as head, neck and non-small-cell lung cancer cells, cells of the mouth such as oral tumor cells.

[0271] Also, as to active targeting, a lipid entity of the invention can be multifunctional, i.e., employ more than one targeting moiety such as CPP, along with Tf; a bifunctional system; e.g., a combination of Tf and poly-L-arginine which can provide transport across the endothelium of the blood–brain barrier. EGFR, is a tyrosine kinase receptor belonging to the ErbB family of receptors that mediates cell growth, differentiation and repair in cells, especially non-cancerous cells, but EGF is overexpressed in certain cells such as many solid

tumors, including colorectal, non-small-cell lung cancer, squamous cell carcinoma of the ovary, kidney, head, pancreas, neck and prostate, and especially breast cancer. The invention comprehends EGFR-targeted monoclonal antibody(ies) linked to a lipid entity of the invention. HER-2 is often overexpressed in patients with breast cancer, and is also associated with lung, bladder, prostate, brain and stomach cancers. HER-2, encoded by the ERBB2 gene. The invention comprehends a HER-2-targeting lipid entity of the invention, e.g., an anti-HER-2-antibody(or binding fragment thereof)-lipid entity of the invention, a HER-2-targeting-PEGylated lipid entity of the invention (e.g., having an anti-HER-2-antibody or binding fragment thereof), a HER-2-targeting-maleimide-PEG polymer- lipid entity of the invention (e.g., having an anti-HER-2-antibody or binding fragment thereof). Upon cellular association, the receptor-antibody complex can be internalized by formation of an endosome for delivery to the cytoplasm.

[0272] With respect to receptor-mediated targeting, the skilled artisan takes into consideration ligand/target affinity and the quantity of receptors on the cell surface, and that PEGylation can act as a barrier against interaction with receptors. The use of antibody-lipid entity of the invention targeting can be advantageous. Multivalent presentation of targeting moieties can also increase the uptake and signaling properties of antibody fragments. In practice of the invention, the skilled person takes into account ligand density (e.g., high ligand densities on a lipid entity of the invention may be advantageous for increased binding to target cells). Preventing early by macrophages can be addressed with a sterically stabilized lipid entity of the invention and linking ligands to the terminus of molecules such as PEG, which is anchored in the lipid entity of the invention (e.g., lipid particle or nanoparticle or liposome or lipid bilayer). The microenvironment of a cell mass such as a tumor microenvironment can be targeted; for instance, it may be advantageous to target cell mass vasculature, such as the tumor vasculature microenvironment. Thus, the invention comprehends targeting VEGF. VEGF and its receptors are well-known proangiogenic molecules and are well-characterized targets for antiangiogenic therapy. Many small-molecule inhibitors of receptor tyrosine kinases, such as VEGFRs or basic FGFRs, have been developed as anticancer agents and the invention comprehends coupling any one or more of these peptides to a lipid entity of the invention, e.g., phage IVO peptide(s) (e.g., via or with a PEG terminus), tumor-homing peptide APRPG (SEQ ID NO: 14) such as APRPG-PEG-modified (SEQ ID NO: 14). VCAM, the vascular endothelium plays a key role in the pathogenesis of inflammation, thrombosis and

atherosclerosis. CAMs are involved in inflammatory disorders, including cancer, and are a logical target, E- and P-selectins, VCAM-1 and ICAMs. Can be used to target a lipid entity of the invention., e.g., with PEGylation.

[0273] Matrix metalloproteases (MMPs) belong to the family of zinc-dependent endopeptidases. They are involved in tissue remodeling, tumor invasiveness, resistance to apoptosis and metastasis. There are four MMP inhibitors called TIMP1–4, which determine the balance between tumor growth inhibition and metastasis; a protein involved in the angiogenesis of tumor vessels is MT1-MMP, expressed on newly formed vessels and tumor tissues. The proteolytic activity of MT1-MMP cleaves proteins, such as fibronectin, elastin, collagen and laminin, at the plasma membrane and activates soluble MMPs, such as MMP-2, which degrades the matrix. An antibody or fragment thereof such as a Fab′ fragment can be used in the practice of the invention such as for an antihuman MT1-MMP monoclonal antibody linked to a lipid entity of the invention, e.g., via a spacer such as a PEG spacer. αβ-integrins or integrins are a group of transmembrane glycoprotein receptors that mediate attachment between a cell and its surrounding tissues or extracellular matrix.

[0274] Integrins contain two distinct chains (heterodimers) called α- and β-subunits. The tumor tissue-specific expression of integrin receptors can be utilized for targeted delivery in the invention, e.g., whereby the targeting moiety can be an RGD peptide such as a cyclic RGD.

[0275] Aptamers are ssDNA or RNA oligonucleotides that impart high affinity and specific recognition of the target molecules by electrostatic interactions, hydrogen bonding and hydrophobic interactions as opposed to the Watson–Crick base pairing, which is typical for the bonding interactions of oligonucleotides. Aptamers as a targeting moiety can have advantages over antibodies: aptamers can demonstrate higher target antigen recognition as compared with antibodies; aptamers can be more stable and smaller in size as compared with antibodies; aptamers can be easily synthesized and chemically modified for molecular conjugation; and aptamers can be changed in sequence for improved selectivity and can be developed to recognize poorly immunogenic targets. Such moieties as a sgc8 aptamer can be used as a targeting moiety (e.g., via covalent linking to the lipid entity of the invention, e.g., via a spacer, such as a PEG spacer).

[0276] Also, as to active targeting, the invention also comprehends intracellular delivery. Since liposomes follow the endocytic pathway, they are entrapped in the endosomes (pH 6.5–6) and subsequently fuse with lysosomes (pH <5), where they undergo degradation that results

in a lower therapeutic potential. The low endosomal pH can be taken advantage of to escape degradation. Fusogenic lipids or peptides, which destabilize the endosomal membrane after the conformational transition/activation at a lowered pH. Amines are protonated at an acidic pH and cause endosomal swelling and rupture by a buffer effect Unsaturated dioleoylphosphatidylethanolamine (DOPE) readily adopts an inverted hexagonal shape at a low pH, which causes fusion of liposomes to the endosomal membrane. This process destabilizes a lipid entity containing DOPE and releases the cargo into the cytoplasm; fusogenic lipid GALA (SEQ ID NO: 15), cholesteryl-GALA (SEQ ID NO: 15) and PEG-GALA (SEQ ID NO: 15) may show a highly efficient endosomal release; a pore-forming protein listeriolysin O may provide an endosomal escape mechanism; and histidine-rich peptides have the ability to fuse with the endosomal membrane, resulting in pore formation, and can buffer the proton pump causing membrane lysis.

[0277] The invention comprehends a lipid entity of the invention modified with CPP(s), for intracellular delivery that may proceed via energy dependent macropinocytosis followed by endosomal escape. The invention further comprehends organelle-specific targeting. A lipid entity of the invention surface-functionalized with the triphenylphosphonium (TPP) moiety or a lipid entity of the invention with a lipophilic cation, rhodamine 123 can be effective in delivery of cargo to mitochondria. DOPE/sphingomyelin/stearyl-octa-arginine can delivers cargos to the mitochondrial interior via membrane fusion. A lipid entity of the invention surface modified with a lysosomotropic ligand, octadecyl rhodamine B can deliver cargo to lysosomes. Ceramides are useful in inducing lysosomal membrane permeabilization; the invention comprehends intracellular delivery of a lipid entity of the invention having a ceramide. The invention further comprehends a lipid entity of the invention targeting the nucleus, e.g., via a DNA-intercalating moiety. The invention also comprehends multifunctional liposomes for targeting, i.e., attaching more than one functional group to the surface of the lipid entity of the invention, for instance to enhances accumulation in a desired site and/or promotes organelle-specific delivery and/or target a particular type of cell and/or respond to the local stimuli such as temperature (e.g., elevated), pH (e.g., decreased), respond to externally applied stimuli such as a magnetic field, light, energy, heat or ultrasound and/or promote intracellular delivery of the cargo. All of these are considered actively targeting moieties.

[0278] It should be understood that as to each possible targeting or active targeting moiety herein discussed, there is an embodiment of the invention wherein the delivery system

comprises such a targeting or active targeting moiety. Likewise, **Table 8** provides exemplary targeting moieties that can be used in the practice of the invention an as to each an embodiment of the invention provides a delivery system that comprises such a targeting moiety.

Table 8. Exemplary, non-limiting, Targeting Moieties		
Targeting Moiety	Target Molecule	Target Cell or Tissue
folate	folate receptor	cancer cells
transferrin	transferrin receptor	cancer cells
Antibody CC52	rat CC531	rat colon adenocarcinoma CC531
anti- *HER2* antibody	*HER2*	*HER2* -overexpressing tumors
anti-*GD2*	*GD2*	neuroblastoma, melanoma
anti-EGFR	*EGFR*	tumor cells overexpressing EGFR
pH-dependent fusogenic peptide diINF-7		ovarian carcinoma
anti-VEGFR	VEGF Receptor	tumor vasculature
anti-CD19	CD19 (B cell marker)	leukemia, lymphoma
cell-penetrating peptide		blood-brain barrier
cyclic arginine-glycine-aspartic acid-tyrosine-cysteine peptide (c(RGDyC)-LP)	avβ3	glioblastoma cells, human umbilical vein endothelial cells, tumor angiogenesis
ASSHN peptide		endothelial progenitor cells; anti-cancer
PR_b peptide	$\alpha_5\beta_1$ integrin	cancer cells
AG86 peptide	$\alpha_6\beta_4$ integrin	cancer cells
KCCYSL (P6.1 peptide)	HER-2 receptor	cancer cells
affinity peptide LN (YEVGHRC)	Aminopeptidase N (APN/CD13)	APN-positive tumor
synthetic somatostatin analogue	Somatostatin receptor 2 (SSTR2)	breast cancer
anti-CD20 monoclonal antibody	B-lymphocytes	B cell lymphoma

[0279] Thus, in an embodiment of the delivery system, the targeting moiety comprises a receptor ligand, such as, for example, hyaluronic acid for CD44 receptor, galactose for hepatocytes, or antibody or fragment thereof such as a binding antibody fragment against a desired surface receptor, and as to each of a targeting moiety comprising a receptor ligand, or

an antibody or fragment thereof such as a binding fragment thereof, such as against a desired surface receptor, there is an embodiment of the invention wherein the delivery system comprises a targeting moiety comprising a receptor ligand, or an antibody or fragment thereof such as a binding fragment thereof, such as against a desired surface receptor, or hyaluronic acid for CD44 receptor, galactose for hepatocytes (*see, e.g.*, Surace et al, "Lipoplexes targeting the CD44 hyaluronic acid receptor for efficient transfection of breast cancer cells," J. Mol Pharm 6(4):1062-73; doi: 10.1021/mp800215d (2009); Sonoke et al, "Galactose-modified cationic liposomes as a liver-targeting delivery system for small interfering RNA," Biol Pharm Bull. 34(8):1338-42 (2011); Torchilin, "Antibody-modified liposomes for cancer chemotherapy," Expert Opin. Drug Deliv. 5 (9), 1003-1025 (2008); Manjappa et al, "Antibody derivatization and conjugation strategies: application in preparation of stealth immunoliposome to target chemotherapeutics to tumor," J. Control. Release 150 (1), 2-22 (2011); Sofou S "Antibody-targeted liposomes in cancer therapy and imaging," Expert Opin. Drug Deliv. 5 (2): 189-204 (2008); Gao J et al, "Antibody-targeted immunoliposomes for cancer treatment," Mini. Rev. Med. Chem. 13(14): 2026-2035 (2013); Molavi et al, "Anti-CD30 antibody conjugated liposomal doxorubicin with significantly improved therapeutic efficacy against anaplastic large cell lymphoma," Biomaterials 34(34):8718-25 (2013), each of which and the documents cited therein are hereby incorporated herein by reference), the teachings of which can be applied and/or adapted for targeted delivery of one or more CRISPR-Cas molecules described herein.

[0280] Other exemplary targeting moieties are described elsewhere herein, such as epitope tags and the like.

Responsive Delivery

[0281] In some embodiments, the delivery vehicle can allow for responsive delivery of the cargo(s). Responsive delivery, as used in this context herein, refers to delivery of cargo(s) by the delivery vehicle in response to an external stimuli. Examples of suitable stimuli include, without limitation, an energy (light, heat, cold, and the like), a chemical stimuli (e.g., chemical composition, etc.), and a biologic or physiologic stimuli (e.g., environmental pH, osmolarity, salinity, biologic molecule, etc.). In some embodiments, the targeting moiety can be responsive to an external stimuli and facilitate responsive delivery. In other embodiments, responsiveness is determined by a non-targeting moiety component of the delivery vehicle.

[0282] The delivery vehicle can be stimuli-sensitive, e.g., sensitive to an externally applied stimuli, such as magnetic fields, ultrasound or light; and pH-triggering can also be used, e.g., a labile linkage can be used between a hydrophilic moiety such as PEG and a hydrophobic moiety such as a lipid entity of the invention, which is cleaved only upon exposure to the relatively acidic conditions characteristic of the a particular environment or microenvironment such as an endocytic vacuole or the acidotic tumor mass. pH-sensitive copolymers can also be incorporated in embodiments of the invention can provide shielding; diortho esters, vinyl esters, cysteine-cleavable lipopolymers, double esters and hydrazones are a few examples of pH-sensitive bonds that are quite stable at pH 7.5, but are hydrolyzed relatively rapidly at pH 6 and below, e.g., a terminally alkylated copolymer of N-isopropylacrylamide and methacrylic acid that copolymer facilitates destabilization of a lipid entity of the invention and release in compartments with decreased pH value; or, the invention comprehends ionic polymers for generation of a pH-responsive lipid entity of the invention (e.g., poly(methacrylic acid), poly(diethylaminoethyl methacrylate), poly(acrylamide) and poly(acrylic acid)).

[0283] Temperature-triggered delivery is also within the ambit of the invention. Many pathological areas, such as inflamed tissues and tumors, show a distinctive hyperthermia compared with normal tissues. Utilizing this hyperthermia is an attractive strategy in cancer therapy since hyperthermia is associated with increased tumor permeability and enhanced uptake. This technique involves local heating of the site to increase microvascular pore size and blood flow, which, in turn, can result in an increased extravasation of embodiments of the invention. Temperature-sensitive lipid entity of the invention can be prepared from thermosensitive lipids or polymers with a low critical solution temperature. Above the low critical solution temperature (e.g., at site such as tumor site or inflamed tissue site), the polymer precipitates, disrupting the liposomes to release. Lipids with a specific gel-to-liquid phase transition temperature are used to prepare these lipid entities of the invention; and a lipid for a thermosensitive embodiment can be dipalmitoylphosphatidylcholine. Thermosensitive polymers can also facilitate destabilization followed by release, and a useful thermosensitive polymer is poly (N-isopropylacrylamide). Another temperature triggered system can employ lysolipid temperature-sensitive liposomes.

[0284] The invention also comprehends redox-triggered delivery. The difference in redox potential between normal and inflamed or tumor tissues, and between the intra- and extra-cellular environments has been exploited for delivery, e.g., GSH is a reducing agent abundant

in cells, especially in the cytosol, mitochondria and nucleus. The GSH concentrations in blood and extracellular matrix are just one out of 100 to one out of 1000 of the intracellular concentration, respectively. This high redox potential difference caused by GSH, cysteine and other reducing agents can break the reducible bonds, destabilize a lipid entity of the invention and result in release of payload. The disulfide bond can be used as the cleavable/reversible linker in a lipid entity of the invention, because it causes sensitivity to redox owing to the disulfideto-thiol reduction reaction; a lipid entity of the invention can be made reduction sensitive by using two (e.g., two forms of a disulfide-conjugated multifunctional lipid as cleavage of the disulfide bond (e.g., via tris(2-carboxyethyl)phosphine, dithiothreitol, L-cysteine or GSH), can cause removal of the hydrophilic head group of the conjugate and alter the membrane organization leading to release of payload. Calcein release from reduction-sensitive lipid entity of the invention containing a disulfide conjugate can be more useful than a reduction-insensitive embodiment.

[0285] Enzymes can also be used as a trigger to release payload. Enzymes, including MMPs (e.g. MMP2), phospholipase A2, alkaline phosphatase, transglutaminase or phosphatidylinositol-specific phospholipase C, have been found to be overexpressed in certain tissues, e.g., tumor tissues. In the presence of these enzymes, specially engineered enzyme-sensitive lipid entity of the invention can be disrupted and release the payload. an MMP2-cleavable octapeptide (Gly-Pro-Leu-Gly-Ile-Ala-Gly-Gln (SEQ ID NO: 16)) can be incorporated into a linker, and can have antibody targeting, e.g., antibody 2C5.

[0286] The invention also comprehends light-or energy-triggered delivery, e.g., the lipid entity of the invention can be light-sensitive, such that light or energy can facilitate structural and conformational changes, which lead to direct interaction of the lipid entity of the invention with the target cells via membrane fusion, photo-isomerism, photofragmentation or photopolymerization; such a moiety therefor can be benzoporphyrin photosensitizer. Ultrasound can be a form of energy to trigger delivery; a lipid entity of the invention with a small quantity of particular gas, including air or perfluorated hydrocarbon can be triggered to release with ultrasound, e.g., low-frequency ultrasound (LFUS). Magnetic delivery: A lipid entity of the invention can be magnetized by incorporation of magnetites, such as Fe_3O_4 or γ-Fe_2O_3, e.g., those that are less than 10 nm in size. Targeted delivery can be then by exposure to a magnetic field.

ENGINEERED CELLS AND CELL POPULATIONS

[0287] Described herein are various aspects of engineered cells or cell populations that can include one or more of the programmable pattern recognition composition or system polynucleotides, polypeptides, vectors, and/or vector systems, and/or programmable pattern recognition composition or system particles (e.g., those particles, such as virus particles, produced from a programmable pattern recognition composition or system polynucleotide and/or vector(s)) described elsewhere herein. In some embodiments, the engineered cells can express one or more of the programmable pattern recognition composition or system polynucleotides and/or can produce one or more particles, such as virus particles or exosomes, containing a programmable pattern recognition composition or system, which are described in greater detail herein. Such cells are also referred to herein as "producer cells".

[0288] Described in certain example embodiments herein are engineered cells modified to express elements (i) and (iii) of the detection composition described herein. In certain example embodiments, where the engineered cells are further modified to express element (iv) of the detection composition described herein. In certain example embodiments, where the engineered cells are further modified to express element (ii) of the detection composition described herein.

[0289] In an embodiment, the invention provides a non-human eukaryotic organism; for example, a multicellular eukaryotic organism, including a eukaryotic host cell containing one or more components of an engineered delivery system described herein according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell containing one or more components of a programmable pattern recognition composition or system described herein according to any of the described embodiments. In some embodiments, the organism is a host of AAV.

[0290] The engineered cell can be any eukaryotic cell, including but not limited to, human, non-human animal, plant, algae, and the like.

[0291] The engineered cell can be a prokaryotic cell. The prokaryotic cell can be bacterial cell. The prokaryotic cell can be an archaea cell. The bacterial cell can be any suitable bacterial cell. Suitable bacterial cells can be from the genus Escherichia, Bacillus, Lactobacillus, Rhodococcus, Rodhobacter, Synechococcus, Synechoystis, Pseudomonas, Psedoaltermonas, Stenotrophamonas, and Streptomyces Suitable bacterial cells include, but are not limited to

Escherichia coli cells, Caulobacter crescentus cells, Rodhobacter sphaeroides cells, Psedoaltermonas haloplanktis cells. Suitable strains of bacterial include, but are not limited to BL21(DE3), DL21(DE3)-pLysS, BL21 Star-pLysS, BL21-SI, BL21-AI, Tuner, Tuner pLysS, Origami, Origami B pLysS, Rosetta, Rosetta pLysS, Rosetta-gami-pLysS, BL21 CodonPlus, AD494, BL2trxB, HMS174, NovaBlue(DE3), BLR, C41(DE3), C43(DE3), Lemo21(DE3), Shuffle T7, ArcticExpress and ArticExpress (DE3).

[0292] The engineered cell can be a eukaryotic cell. The eukaryotic cells may be those of or derived from a particular organism, such as a plant or a mammal, including but not limited to human, or non-human eukaryote or animal or mammal as herein discussed, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate. In some embodiments, the engineered cell can be a cell line. Examples of cell lines include, but are not limited to, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panc1, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/ 3T3 mouse embryo fibroblast, 3T3 Swiss, 3T3-L1, 132-d5 human fetal fibroblasts; 10.1 mouse fibroblasts, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr -/-, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, JY cells, K562 cells, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN / OPCT cell lines, Peer, PNT-1A / PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, Sf-9, SkBr3, T2, T-47D, T84, THP1 cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, and transgenic varieties thereof. Cell lines are available from a variety of sources known to those with skill in the art (see e.g., the American Type Culture Collection (ATCC) (Manassas, Va.)).

[0293] Further, the engineered cell may be a fungus cell. As used herein, a "fungal cell" refers to any type of eukaryotic cell within the kingdom of fungi. Phyla within the kingdom of fungi include Ascomycota, Basidiomycota, Blastocladiomycota, Chytridiomycota, Glomeromycota, Microsporidia, and Neocallimastomycota. fungal cells may include yeasts, molds, and filamentous fungi. In some embodiments, the fungal cell is a yeast cell.

[0294] As used herein, the term "yeast cell" refers to any fungal cell within the phyla Ascomycota and Basidiomycota. Yeast cells may include budding yeast cells, fission yeast cells, and mold cells. Without being limited to these organisms, many types of yeast used in laboratory and industrial settings are part of the phylum Ascomycota. In some embodiments, the yeast cell is an S. cerevisiae, Kluyveromyces marxianus, or Issatchenkia orientalis cell. Other yeast cells may include without limitation Candida spp. (e.g., Candida albicans), Yarrowia spp. (e.g., Yarrowia lipolytica), Pichia spp. (e.g., Pichia pastoris), Kluyveromyces spp. (e.g., Kluyveromyces lactis and Kluyveromyces marxianus), Neurospora spp. (e.g., Neurospora crassa), Fusarium spp. (e.g., Fusarium oxysporum), and Issatchenkia spp. (e.g., Issatchenkia orientali, a.k.a. Pichia kudriav”evii and Candida acidothermophilum). In some embodiments, the fungal cell is a filamentous fungal cell. As used herein, the term "filamentous fungal cell" refers to any type of fungal cell that grows in filaments, i.e., hyphae or mycelia. Examples of filamentous fungal cells may include without limitation Aspergillus (e.g., Aspergillus niger), Trichoderma spp. (e.g., Trichoderma reesei), Rhizopus spp. (e.g., Rhizopus oryza”), and Mortierella spp. (e.g., Mortierella isabellina).

[0295] In some embodiments, the fungal cell is an industrial strain. As used herein, "industrial strain" refers to any strain of fungal cell used in or isolated from an industrial process, e.g., production of a product on a commercial or industrial scale. Industrial strain may refer to a fungal species that is typically used in an industrial process, or it may refer to an isolate of a fungal species that may be also used for non-industrial purposes (e.g., laboratory research). Examples of industrial processes may include fermentation (e.g., in production of food or beverage products), distillation, biofuel production, production of a compound, and production of a polypeptide. Example“ of indus”rial strains can include, without limitation, JAY270 and ATCC4124.

[0296] In some embodiments, the fungal cell is a polyploid cell. As used herein, a "polyploid" cell may refer to any cell whose genome is present in more than one copy. A polyploid cell may refer to a type of cell that is naturally found in a polyploid state, or it may

refer to a cell that has been induced to exist in a polyploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A polyploid cell may refer to a cell whose entire genome is polyploid, or it may "refer" to a cell that is polyploid in a particular genomic locus of interest.

[0297] In some embodiments, the fungal cell is a diploid cell. As used herein, a "diploid" cell may refer to any cell whose genome is present in two copies. A diploid cell may refer to a type of cell that is naturally found in a diploid state, or it may refer to a cell that has been induced to exist in a diploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). For example, the S. cerevisiae strain S228C may be maintained in a haploid or diploid state. A diploid cell may refer to a cell whose entire genome is diploid, or it may refer to a cell that is diploid in a particular genomic locus of interest. In some embodiments, the fungal cell is a haploid cell. As used herein, a "haploid" cell may refer to any cell whose genome is present in one copy. A haploid cell may refer to a type of cell that is naturally found in a haploid state, or it may refer to a cell that has been induced to exist in a haploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). For example, the S. cerevisiae strain S228C may be maintained in a haploid or diploid state. A haploid cell may refer to a cell whose entire genome is haploid, or it may refer to a cell that is haploid in a particular genomic locus of interest.

[0298] In some embodiments, the engineered cell is a cell obtained from a subject. In some embodiments, the subject is a healthy or non-diseased subject. In some embodiments, the subject is a subject with a desired physiological and/or biological characteristic such that when an engineered delivery vesicle is produced it can package one or more molecules that are within the producer cell that can be related to the desired physiological and/or biological characteristic. In this context, the cargo molecules incorporated into the delivery vesicles can be capable of transferring the desired characteristic to a recipient cell.

[0299] In some embodiments, a cell can be obtained from a subject, modified such that it is an engineered delivery vesicle producer cell, and administered back to the subject from which it was obtained (autologous) or delivered to an allogenic subject. In other words, a producer cell described herein can be used in an autologous or allogenic context, such as in a cell therapy. In these embodiments, the cells can deliver a cargo, such as a therapeutic cargo or a cargo that can manipulate a cellular microenvironment within the subject.

[0300] Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids (e.g., such as one or more of the polynucleotides of the engineered delivery system described herein) in cells or target tissues. Such methods can be used to administer nucleic acids encoding components of a nucleic acid-targeting system to cells in culture, or in a host organism. In some embodiments, a delivery is via a polynucleotide molecule (e.g., a DNA or RNA molecule) not contained in a vector. In some embodiments, delivery is via a vector. In some embodiments, delivery is via viral particles. In aspects, delivery is via a particle, (e.g., a nanoparticle) carrying one or more engineered delivery system polynucleotides, vectors, or viral particles. Particles, including nanoparticles, are discussed in greater detail elsewhere herein.

[0301] Vector delivery can be appropriate in some embodiments, where in vivo expression is envisaged. It will be appreciated that the engineered cells can be generated in vitro, ex vivo, in situ, or in vivo by delivery of one or more components of the engineered delivery systems as described elsewhere herein.

Engineered Microbiomes

[0302] As described elsewhere herein the engineered protein compositions of the present invention can be configured to engineer a microbiome by targeting specific microbes within a microbiome, via target recognition specific to polypeptides, molecules, and/or molecular patterns, optionally a PAMP, on desired target microbes within the microbiome. The target cells can be acted upon by the effector functions of the engineered protein composition to kill the target cells or otherwise modify them so as to e.g., have an inhibited or stimulated growth or proliferation so as to influence their relative or absolute amount or abundance within the microbiome. By altering the microbe population within a microbiome, the structure of the microbiome can be engineered as desired. In some embodiments, the engineered microbiome has positive effects on the health or other functionality of the organ, environment, or organism in which the microbiome exists. Such engineered microbiomes are within the scope of the present invention.

[0303] Suitable conventional viral and non-viral based methods of engineering cells to contain and/or express the engineered delivery system polynucleotides and/or vectors described herein are generally known in the art and/or described elsewhere herein.

PHARMACEUTICAL FORMULATIONS

[0304] Also described herein are pharmaceutical formulations that can contain an amount, effective amount, and/or least effective amount, and/or therapeutically effective amount of one or more compounds, molecules, compositions, vectors, vector systems, cells, or a combination thereof (which are also referred to as the primary active agent or ingredient elsewhere herein) of the present invention described in greater detail elsewhere herein and a pharmaceutically acceptable carrier or excipient. As used herein, "pharmaceutical formulation" refers to the combination of an active agent, compound, or ingredient with a pharmaceutically acceptable carrier or excipient, making the composition suitable for diagnostic, therapeutic, or preventive use in vitro, in vivo, or ex vivo. As used herein, "pharmaceutically acceptable carrier or excipient" refers to a carrier or excipient that is useful in preparing a pharmaceutical formulation that is generally safe, non-toxic, and is neither biologically or otherwise undesirable, and includes a carrier or excipient that is acceptable for veterinary use as well as human pharmaceutical use. A "pharmaceutically acceptable carrier or excipient" as used in the specification and claims includes both one and more than one such carrier or excipient. When present, the compound can optionally be present in the pharmaceutical formulation as a pharmaceutically acceptable salt. In some embodiments, the pharmaceutical formulation can include, such as an active ingredient, a programmable pattern recognition composition or system or component thereof described in greater detail elsewhere herein.

[0305] In some embodiments, the active ingredient is present as a pharmaceutically acceptable salt of the active ingredient. As used herein, "pharmaceutically acceptable salt" refers to any acid or base addition salt whose counter-ions are non-toxic to the subject to which they are administered in pharmaceutical doses of the salts. Suitable salts include, hydrobromide, iodide, nitrate, bisulfate, phosphate, isonicotinate, lactate, salicylate, acid citrate, tartrate, oleate, tannate, pantothenate, bitartrate, ascorbate, succinate, maleate, gentisinate, fumarate, gluconate, glucaronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate, camphorsulfonate, napthalenesulfonate, propionate, malonate, mandelate, malate, phthalate, and pamoate.

[0306] The pharmaceutical formulations described herein can be administered to a subject in need thereof via any suitable method or route to a subject in need thereof. Suitable administration routes can include, but are not limited to auricular (otic), buccal, conjunctival, cutaneous, dental, electro-osmosis, endocervical, endosinusial, endotracheal, enteral, epidural,

extra-amniotic, extracorporeal, hemodialysis, infiltration, interstitial, intra-abdominal, intra-amniotic, intra-arterial, intra-articular, intrabiliary, intrabronchial, intrabursal, intracardiac, intracartilaginous, intracaudal, intracavernous, intracavitary, intracerebral, intracisternal, intracorneal, intracoronal (dental), intracoronary, intracorporus cavernosum, intradermal, intradiscal, intraductal, intraduodenal, intradural, intraepidermal, intraesophageal, intragastric, intragingival, intraileal, intralesional, intraluminal, intralymphatic, intramedullary, intrameningeal, intramuscular, intraocular, intraovarian, intrapericardial, intraperitoneal, intrapleural, intraprostatic, intrapulmonary, intrasinal, intraspinal, intrasynovial, intratendinous, intratesticular, intrathecal, intrathoracic, intratubular, intratumor, intratympanic, intrauterine, intravascular, intravenous, intravenous bolus, intravenous drip, intraventricular, intravesical, intravitreal, iontophoresis, irrigation, laryngeal, nasal, nasogastric, occlusive dressing technique, ophthalmic, oral, oropharyngeal, other, parenteral, percutaneous, periarticular, peridural, perineural, periodontal, rectal, respiratory (inhalation), retrobulbar, soft tissue, subarachnoid, subconjunctival, subcutaneous, sublingual, submucosal, topical, transdermal, transmucosal, transplacental, transtracheal, transtympanic, ureteral, urethral, and/or vaginal administration, and/or any combination of the above administration routes, which typically depends on the disease to be treated and/or the active ingredient(s).

[0307] Where appropriate, compounds, molecules, compositions, vectors, vector systems, cells, or a combination thereof described in greater detail elsewhere herein can be provided to a subject in need thereof as an ingredient, such as an active ingredient or agent, in a pharmaceutical formulation. As such, also described are pharmaceutical formulations containing one or more of the compounds and salts thereof, or pharmaceutically acceptable salts thereof described herein. Suitable salts include, hydrobromide, iodide, nitrate, bisulfate, phosphate, isonicotinate, lactate, salicylate, acid citrate, tartrate, oleate, tannate, pantothenate, bitartrate, ascorbate, succinate, maleate, gentisinate, fumarate, gluconate, glucaronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate, camphorsulfonate, napthalenesulfonate, propionate, malonate, mandelate, malate, phthalate, and pamoate.

[0308] As used herein, “agent” refers to any substance, compound, molecule, and the like, which can be biologically active or otherwise can induce a biological and/or physiological effect on a subject to which it is administered to. As used herein, “active agent” or “active ingredient” refers to a substance, compound, or molecule, which is biologically active or

otherwise, induces a biological or physiological effect on a subject to which it is administered to. In other words, "active agent" or "active ingredient" refers to a component or components of a composition to which the whole or part of the effect of the composition is attributed. An agent can be a primary active agent, or in other words, the component(s) of a composition to which the whole or part of the effect of the composition is attributed. An agent can be a secondary agent, or in other words, the component(s) of a composition to which an additional part and/or other effect of the composition is attributed.

Pharmaceutically Acceptable Carriers and Secondary Ingredients and Agents

[0309] The pharmaceutical formulation can include a pharmaceutically acceptable carrier. Suitable pharmaceutically acceptable carriers include, but are not limited to water, salt solutions, alcohols, gum arabic, vegetable oils, benzyl alcohols, polyethylene glycols, gelatin, carbohydrates such as lactose, amylose or starch, magnesium stearate, talc, silicic acid, viscous paraffin, perfume oil, fatty acid esters, hydroxy methylcellulose, and polyvinyl pyrrolidone, which do not deleteriously react with the active composition.

[0310] The pharmaceutical formulations can be sterilized, and if desired, mixed with agents, such as lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic pressure, buffers, coloring, flavoring and/or aromatic substances, and the like which do not deleteriously react with the active compound.

[0311] In some embodiments, the pharmaceutical formulation can also include an effective amount of secondary active agents, including but not limited to, biologic agents or molecules including, but not limited to, e.g. polynucleotides, amino acids, peptides, polypeptides, antibodies, aptamers, ribozymes, hormones, immunomodulators, antipyretics, anxiolytics, antipsychotics, analgesics, antispasmodics, anti-inflammatories, anti-histamines, anti-infectives, chemotherapeutics, imaging agents, radiation sensitizers, and combinations thereof.

Effective Amounts

[0312] In some embodiments, the amount of the primary active agent and/or optional secondary agent can be an effective amount, least effective amount, and/or therapeutically effective amount. As used herein, "effective amount" refers to the amount of the primary and/or optional secondary agent included in the pharmaceutical formulation that achieve one or more therapeutic effects or desired effect. As used herein, "least effective" amount refers to the lowest amount of the primary and/or optional secondary agent that achieves the one or more therapeutic or other desired effects. As used herein, "therapeutically effective amount" refers

to the amount of the primary and/or optional secondary agent included in the pharmaceutical formulation that achieves one or more therapeutic effects.

[0313] The effective amount, least effective amount, and/or therapeutically effective amount of the primary and optional secondary active agent described elsewhere herein contained in the pharmaceutical formulation can be any non-zero amount ranging from about 0 to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000 pg, ng, μg, mg, or g or be any numerical value or subrange within any of these ranges.

[0314] In some embodiments, the effective amount, least effective amount, and/or therapeutically effective amount can be an effective concentration, least effective concentration, and/or therapeutically effective concentration, which can each be any non-zero amount ranging from about 0 to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000 pM, nM, μM, mM, or M or be any numerical value or subrange within any of these ranges.

[0315] In other embodiments, the effective amount, least effective amount, and/or therapeutically effective amount of the primary and optional secondary active agent be any non-zero amount ranging from about 0 to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000 IU or be any numerical value or subrange within any of these ranges.

[0316] In some embodiments, the primary and/or the optional secondary active agent present in the pharmaceutical formulation can be any non-zero amount ranging from about 0 to 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, 0.2, 0.21, 0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.3, 0.31, 0.32, 0.33, 0.34, 0.35, 0.36, 0.37, 0.38, 0.39, 0.4, 0.41, 0.42, 0.43, 0.44, 0.45, 0.46, 0.47, 0.48, 0.49, 0.5, 0.51, 0.52, 0.53, 0.54, 0.55, 0.56, 0.57, 0.58, 0.59, 0.6, 0.61, 0.62, 0.63, 0.64, 0.65, 0.66, 0.67, 0.68, 0.69, 0.7, 0.71, 0.72, 0.73, 0.74, 0.75, 0.76, 0.77, 0.78, 0.79, 0.8, 0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88, 0.89, 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.9, to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, 99.9 % w/w, v/v, or w/v of the pharmaceutical formulation or be any numerical value or subrange within any of these ranges.

[0317] In some embodiments where a cell or cell population is present in the pharmaceutical formulation (e.g., as a primary and/or or secondary active agent), the effective amount of cells can be any amount ranging from about 1 or 2 cells to $1X10^{1}$/mL, $1X10^{20}$/mL or more, such as about $1X10^{1}$/mL, $1X10^{2}$/mL, $1X10^{3}$/mL, $1X10^{4}$/mL, $1X10^{5}$/mL, $1X10^{6}$/mL, $1X10^{7}$/mL, $1X10^{8}$/mL, $1X10^{9}$/mL, $1X10^{10}$/mL, $1X10^{11}$/mL, $1X10^{12}$/mL, $1X10^{13}$/mL, $1X10^{14}$/mL, $1X10^{15}$/mL, $1X10^{16}$/mL, $1X10^{17}$/mL, $1X10^{18}$/mL, $1X10^{19}$/mL, to/or about $1X10^{20}$/mL or any numerical value or subrange within any of these ranges.

[0318] In some embodiments, the amount or effective amount, particularly where an infective particle is being delivered (e.g., a virus particle having the primary or secondary agent as a cargo), the effective amount of virus particles can be expressed as a titer (plaque forming units per unit of volume) or as a MOI (multiplicity of infection). In some embodiments, the effective amount can be about $1X10^{1}$ particles per pL, nL, μL, mL, or L to $1X10^{20}$/ particles per pL, nL, μL, mL, or L or more, such as about $1X10^{1}$, $1X10^{2}$, $1X10^{3}$, $1X10^{4}$, $1X10^{5}$, $1X10^{6}$, $1X10^{7}$, $1X10^{8}$, $1X10^{9}$, $1X10^{10}$, $1X10^{11}$, $1X10^{12}$, $1X10^{13}$, $1X10^{14}$, $1X10^{15}$, $1X10^{16}$, $1X10^{17}$, $1X10^{18}$, $1X10^{19}$, to/or about $1X10^{20}$ particles per pL, nL, μL, mL, or L. In some embodiments, the effective titer can be about $1X10^{1}$ transforming units per pL, nL, μL, mL, or L to $1X10^{20}$/ transforming units per pL, nL, μL, mL, or L or more, such as about $1X10^{1}$, $1X10^{2}$, $1X10^{3}$,

$1X10^4$, $1X10^5$, $1X10^6$, $1X10^7$, $1X10^8$, $1X10^9$, $1X10^{10}$, $1X10^{11}$, $1X10^{12}$, $1X10^{13}$, $1X10^{14}$, $1X10^{15}$, $1X10^{16}$, $1X10^{17}$, $1X10^{18}$, $1X10^{19}$, to/or about $1X10^{20}$ transforming units per pL, nL, μL, mL, or L or any numerical value or subrange within these ranges. In some embodiments, the MOI of the pharmaceutical formulation can range from about 0.1 to 10 or more, such as 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 4, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 5, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 6, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 7, 7.1, 7.2, 7.3, 7.4, 7.5, 7.6, 7.7, 7.8, 7.9, 8, 8.1, 8.2, 8.3, 8.4, 8.5, 8.6, 8.7, 8.8, 8.9, 9, 9.1, 9.2, 9.3, 9.4, 9.5, 9.6, 9.7, 9.8, 9.9, 10 or more or any numerical value or subrange within these ranges.

[0319] In some embodiments, the amount or effective amount of the one or more of the active agent(s) described herein contained in the pharmaceutical formulation can range from about 1 pg/kg to about 10 mg/kg based upon the bodyweight of the subject in need thereof or average bodyweight of the specific patient population to which the pharmaceutical formulation can be administered.

[0320] In embodiments where there is a secondary agent contained in the pharmaceutical formulation, the effective amount of the secondary active agent will vary depending on the secondary agent, the primary agent, the administration route, subject age, disease, stage of disease, among other things, which will be one of ordinary skill in the art.

[0321] When optionally present in the pharmaceutical formulation, the secondary active agent can be included in the pharmaceutical formulation or can exist as a stand-alone compound or pharmaceutical formulation that can be administered contemporaneously or sequentially with the compound, derivative thereof, or pharmaceutical formulation thereof.

[0322] In some embodiments, the effective amount of the secondary active agent, when optionally present, is any non-zero amount ranging from about 0 to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, 99.9 % w/w, v/v, or w/v of the total active agents present in the pharmaceutical formulation or any numerical value or subrange within these ranges. In additional embodiments, the effective amount of the secondary active agent is any non-zero amount

ranging from about 0 to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, 99.9 % w/w, v/v, or w/v of the total pharmaceutical formulation or any numerical value or subrange within these ranges.

Dosage Forms

[0323] In some embodiments, the pharmaceutical formulations described herein can be provided in a dosage form. The dosage form can be administered to a subject in need thereof. The dosage form can be effective generate specific concentration, such as an effective concentration, at a given site in the subject in need thereof. As used herein, "dose," "unit dose," or "dosage" can refer to physically discrete units suitable for use in a subject, each unit containing a predetermined quantity of the primary active agent, and optionally present secondary active ingredient, and/or a pharmaceutical formulation thereof calculated to produce the desired response or responses in association with its administration. In some embodiments, the given site is proximal to the administration site. In some embodiments, the given site is distal to the administration site. In some cases, the dosage form contains a greater amount of one or more of the active ingredients present in the pharmaceutical formulation than the final intended amount needed to reach a specific region or location within the subject to account for loss of the active components such as via first and second pass metabolism.

[0324] The dosage forms can be adapted for administration by any appropriate route. Appropriate routes include, but are not limited to, oral (including buccal or sublingual), rectal, intraocular, inhaled, intranasal, topical (including buccal, sublingual, or transdermal), vaginal, parenteral, subcutaneous, intramuscular, intravenous, internasal, and intradermal. Other appropriate routes are described elsewhere herein. Such formulations can be prepared by any method known in the art.

[0325] Dosage forms adapted for oral administration can discrete dosage units such as capsules, pellets or tablets, powders or granules, solutions, or suspensions in aqueous or non-aqueous liquids; edible foams or whips, or in oil-in-water liquid emulsions or water-in-oil liquid emulsions. In some embodiments, the pharmaceutical formulations adapted for oral administration also include one or more agents which flavor, preserve, color, or help disperse the pharmaceutical formulation. Dosage forms prepared for oral administration can also be in

the form of a liquid solution that can be delivered as a foam, spray, or liquid solution. The oral dosage form can be administered to a subject in need thereof. Where appropriate, the dosage forms described herein can be microencapsulated.

[0326] The dosage form can also be prepared to prolong or sustain the release of any ingredient. In some embodiments, compounds, molecules, compositions, vectors, vector systems, cells, or a combination thereof described herein can be the ingredient whose release is delayed. In some embodiments the primary active agent is the ingredient whose release is delayed. In some embodiments, an optional secondary agent can be the ingredient whose release is delayed. Suitable methods for delaying the release of an ingredient include, but are not limited to, coating or embedding the ingredients in material in polymers, wax, gels, and the like. Delayed release dosage formulations can be prepared as described in standard references such as "Pharmaceutical dosage form tablets," eds. Liberman et. al. (New York, Marcel Dekker, Inc., 1989), "Remington - The science and practice of pharmacy", 20th ed., Lippincott Williams & Wilkins, Baltimore, MD, 2000, and "Pharmaceutical dosage forms and drug delivery systems", 6th Edition, Ansel et al., (Media, PA: Williams and Wilkins, 1995). These references provide information on excipients, materials, equipment, and processes for preparing tablets and capsules and delayed release dosage forms of tablets and pellets, capsules, and granules. The delayed release can be anywhere from about an hour to about 3 months or more.

[0327] Examples of suitable coating materials include, but are not limited to, cellulose polymers such as cellulose acetate phthalate, hydroxypropyl cellulose, hydroxypropyl methylcellulose, hydroxypropyl methylcellulose phthalate, and hydroxypropyl methylcellulose acetate succinate; polyvinyl acetate phthalate, acrylic acid polymers and copolymers, and methacrylic resins that are commercially available under the trade name EUDRAGIT® (Roth Pharma, Westerstadt, Germany), zein, shellac, and polysaccharides.

[0328] Coatings may be formed with a different ratio of water-soluble polymer, water insoluble polymers, and/or pH dependent polymers, with or without water insoluble/water soluble non-polymeric excipient, to produce the desired release profile. The coating is either performed on the dosage form (matrix or simple) which includes, but is not limited to, tablets (compressed with or without coated beads), capsules (with or without coated beads), beads, particle compositions, "ingredient as is" formulated as, but not limited to, suspension form or as a sprinkle dosage form.

[0329] Where appropriate, the dosage forms described herein can be a liposome. In these embodiments, primary active ingredient(s), and/or optional secondary active ingredient(s), and/or pharmaceutically acceptable salt thereof where appropriate are incorporated into a liposome. In embodiments where the dosage form is a liposome, the pharmaceutical formulation is thus a liposomal formulation. The liposomal formulation can be administered to a subject in need thereof.

[0330] Dosage forms adapted for topical administration can be formulated as ointments, creams, suspensions, lotions, powders, solutions, pastes, gels, sprays, aerosols, or oils. In some embodiments for treatments of the eye or other external tissues, for example the mouth or the skin, the pharmaceutical formulations are applied as a topical ointment or cream. When formulated in an ointment, a primary active ingredient, optional secondary active ingredient, and/or pharmaceutically acceptable salt thereof where appropriate can be formulated with a paraffinic or water-miscible ointment base. In other embodiments, the primary and/or secondary active ingredient can be formulated in a cream with an oil-in-water cream base or a water-in-oil base. Dosage forms adapted for topical administration in the mouth include lozenges, pastilles, and mouth washes.

[0331] Dosage forms adapted for nasal or inhalation administration include aerosols, solutions, suspension drops, gels, or dry powders. In some embodiments, a primary active ingredient, optional secondary active ingredient, and/or pharmaceutically acceptable salt thereof where appropriate can be in a dosage form adapted for inhalation is in a particle-size-reduced form that is obtained or obtainable by micronization. In some embodiments, the particle size of the size reduced (e.g., micronized) compound or salt or solvate thereof, is defined by a D_{50} value of about 0.5 to about 10 microns as measured by an appropriate method known in the art. Dosage forms adapted for administration by inhalation also include particle dusts or mists. Suitable dosage forms wherein the carrier or excipient is a liquid for administration as a nasal spray or drops include aqueous or oil solutions/suspensions of an active (primary and/or secondary) ingredient, which may be generated by various types of metered dose pressurized aerosols, nebulizers, or insufflators. The nasal/inhalation formulations can be administered to a subject in need thereof.

[0332] In some embodiments, the dosage forms are aerosol formulations suitable for administration by inhalation. In some of these embodiments, the aerosol formulation contains a solution or fine suspension of a primary active ingredient, secondary active ingredient, and/or

pharmaceutically acceptable salt thereof where appropriate and a pharmaceutically acceptable aqueous or non-aqueous solvent. Aerosol formulations can be presented in single or multi-dose quantities in sterile form in a sealed container. For some of these embodiments, the sealed container is a single dose or multi-dose nasal or an aerosol dispenser fitted with a metering valve (e.g., metered dose inhaler), which is intended for disposal once the contents of the container have been exhausted.

[0333] Where the aerosol dosage form is contained in an aerosol dispenser, the dispenser contains a suitable propellant under pressure, such as compressed air, carbon dioxide, or an organic propellant, including but not limited to a hydrofluorocarbon. The aerosol formulation dosage forms in other embodiments are contained in a pump-atomizer. The pressurized aerosol formulation can also contain a solution or a suspension of a primary active ingredient, optional secondary active ingredient, and/or pharmaceutically acceptable salt thereof. In further embodiments, the aerosol formulation also contains co-solvents and/or modifiers incorporated to improve, for example, the stability and/or taste and/or fine particle mass characteristics (amount and/or profile) of the formulation. Administration of the aerosol formulation can be once daily or several times daily, for example 2, 3, 4, or 8 times daily, in which 1, 2, 3 or more doses are delivered each time. The aerosol formulations can be administered to a subject in need thereof.

[0334] For some dosage forms suitable and/or adapted for inhaled administration, the pharmaceutical formulation is a dry powder inhalable-formulations. In addition to a primary active agent, optional secondary active ingredient, and/or pharmaceutically acceptable salt thereof where appropriate, such a dosage form can contain a powder base such as lactose, glucose, trehalose, mannitol, and/or starch. In some of these embodiments, a primary active agent, secondary active ingredient, and/or pharmaceutically acceptable salt thereof where appropriate is in a particle-size reduced form. In further embodiments, a performance modifier, such as L-leucine or another amino acid, cellobiose octaacetate, and/or metals salts of stearic acid, such as magnesium or calcium stearate. In some embodiments, the aerosol formulations are arranged so that each metered dose of aerosol contains a predetermined amount of an active ingredient, such as the one or more of the compositions, compounds, vector(s), molecules, cells, and combinations thereof described herein.

[0335] Dosage forms adapted for vaginal administration can be presented as pessaries, tampons, creams, gels, pastes, foams, or spray formulations. Dosage forms adapted for rectal

administration include suppositories or enemas. The vaginal formulations can be administered to a subject in need thereof.

[0336] Dosage forms adapted for parenteral administration and/or adapted for injection can include aqueous and/or non-aqueous sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, solutes that render the composition isotonic with the blood of the subject, and aqueous and non-aqueous sterile suspensions, which can include suspending agents and thickening agents. The dosage forms adapted for parenteral administration can be presented in a single-unit dose or multi-unit dose containers, including but not limited to sealed ampoules or vials. The doses can be lyophilized and re-suspended in a sterile carrier to reconstitute the dose prior to administration. Extemporaneous injection solutions and suspensions can be prepared in some embodiments, from sterile powders, granules, and tablets. The parenteral formulations can be administered to a subject in need thereof.

[0337] For some embodiments, the dosage form contains a predetermined amount of a primary active agent, secondary active ingredient, and/or pharmaceutically acceptable salt thereof where appropriate per unit dose. In an embodiment, the predetermined amount of primary active agent, secondary active ingredient, and/or pharmaceutically acceptable salt thereof where appropriate can be an effective amount, a least effect amount, and/or a therapeutically effective amount. In other embodiments, the predetermined amount of a primary active agent, secondary active agent, and/or pharmaceutically acceptable salt thereof where appropriate, can be an appropriate fraction of the effective amount of the active ingredient.

Co-Therapies and Combination Therapies

[0338] In some embodiments, the pharmaceutical formulation(s) described herein are part of a combination treatment or combination therapy. The combination treatment can include the pharmaceutical formulation described herein and an additional treatment modality. The additional treatment modality can be a chemotherapeutic, a biological therapeutic, surgery, radiation, diet modulation, environmental modulation, a physical activity modulation, and combinations thereof.

[0339] In some embodiments, the co-therapy or combination therapy can additionally include but not limited to, polynucleotides, amino acids, peptides, polypeptides, antibodies, aptamers, ribozymes, hormones, immunomodulators, antipyretics, anxiolytics, antipsychotics,

analgesics, antispasmodics, anti-inflammatories, anti-histamines, anti-infectives, chemotherapeutics, radiation sensitizer, and any combination thereof.

Administration of the Pharmaceutical Formulations

[0340] The pharmaceutical formulations or dosage forms thereof described herein can be administered one or more times hourly, daily, monthly, or yearly (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more times hourly, daily, monthly, or yearly). In some embodiments, the pharmaceutical formulations or dosage forms thereof described herein can be administered continuously over a period of time ranging from minutes to hours to days. Devices and dosages forms are known in the art and described herein that are effective to provide continuous administration of the pharmaceutical formulations described herein. In some embodiments, the first one or a few initial amount(s) administered can be a higher dose than subsequent doses. This is typically referred to in the art as a loading dose or doses and a maintenance dose, respectively. In some embodiments, the pharmaceutical formulations can be administered such that the doses over time are tapered (increased or decreased) overtime so as to wean a subject gradually off of a pharmaceutical formulation or gradually introduce a subject to the pharmaceutical formulation.

[0341] As previously discussed, the pharmaceutical formulation can contain a predetermined amount of a primary active agent, secondary active agent, and/or pharmaceutically acceptable salt thereof where appropriate. In some of these embodiments, the predetermined amount can be an appropriate fraction of the effective amount of the active ingredient. Such unit doses may therefore be administered once or more than once a day, month, or year (e.g., 1, 2, 3, 4, 5, 6, or more times per day, month, or year). Such pharmaceutical formulations may be prepared by any of the methods well known in the art.

[0342] Where co-therapies or multiple pharmaceutical formulations are to be delivered to a subject, the different therapies or formulations can be administered sequentially or simultaneously. Sequential administration is administration where an appreciable amount of time occurs between administrations, such as more than about 15, 20, 30, 45, 60 minutes or more. The time between administrations in sequential administration can be on the order of hours, days, months, or even years, depending on the active agent present in each administration. Simultaneous administration refers to administration of two or more formulations at the same time or substantially at the same time (e.g., within seconds or just a

few minutes apart), where the intent is that the formulations be administered together at the same time.

DEVICES

[0343] Described in various embodiments herein are devices that are configured to carry out e.g., one or more of the assays, such as a detection, labeling, or screening, assay described herein. The devices can contain one or more of the programmable pattern recognition compositions, detection compositions, and/or systems or one or more components thereof. The assays or component thereof can be carried out on a device, such as tube, capillary, lateral flow strip, chip, cartridge or another device. The systems and/or assays described herein can be embodied on diagnostic devices. Devices can include very simple devices such as tubes for containing a single sample that contains all the reagents necessary to carry out a programmable pattern recognition and/or CRISPR-Cas collateral activity reaction described herein and provide a result (such as a colometric, turbidity shift, or fluorescent signal) all within the single tube. Other devices can be complex fully automated devices that are capable of handling tens to thousands of samples at time. As is described in greater detail elsewhere herein, one or more compositions (e.g., sample preparation, target amplification reaction, and/or programmable pattern recognition and/or CRISPR-Cas collateral activity detection reagents) can be included in the device. In some embodiments, they are included in one or more compartments and/or locations within the device in a free-dried, lyophilized or some other form. Devices can contain or be configured for optical-based readouts, lateral flow readouts, electrical readouts or others that are described herein and will be appreciated in view of the description provided herein.

[0344] In some example, a device contains a detection composition that comprises an engineered protein of the present invention and a detection construct. Binding of a target polypeptide, target molecule, and/or target molecular pattern on said target polypeptide and/or target molecule to the recognition domain activates the effector domain and mediates effector domain modification of the detection construct resulting in generation of a detectable signal thereby allowing detection of a target polypeptide, target molecule, and/or target molecular pattern on said target polypeptide and/or target molecule.

Discrete Volumes

[0345] In some embodiments the devices can include individual discrete volumes. In certain embodiments, an effector protein of the compositions or systems of the present invention is bound to each discrete volume in the device. In some embodiments, a detection

composition or component thereof (e.g., an engineered protein of the present invention and/or a detection construct) of the present invention is contained or bound to one or more or each discrete volumes in the device. Each discrete volume may comprise a different guide RNA specific for a different target molecule. Each discrete volume may contain a different engineered protein of the present invention, each specific to a different target polypeptide, target molecule, and/or target molecular pattern.

[0346] In certain embodiments, a sample is exposed to the one or more individual discrete volumes. In some embodiments, a sample is exposed to a solid substrate that comprises the individual discrete volumes. In certain embodiments, a sample is exposed to a solid substrate comprising more than one discrete volume each comprising an engineered protein of the present invention that is specific for a target polypeptide, target molecule, and/or target molecular pattern. In certain embodiments, a sample is exposed to a solid substrate comprising more than one discrete volume each comprising a guide RNA specific for a target molecule. Not being bound by a theory, each engineered protein of the present invention and/or each guide RNA will capture its target molecule from the sample and the sample does not need to be divided into separate assays. Thus, a valuable sample may be preserved.

[0347] An effector protein in the device (e.g., an effector domain of the engineered protein composition or detection construct) may be a fusion protein comprising an affinity tag. Affinity tags are well known in the art (e.g., HA tag, Myc tag, Flag tag, His tag, biotin). The effector protein may be linked to a biotin molecule and the discrete volumes may comprise streptavidin. In other embodiments, an effector protein compositions or systems of the present invention is bound by an antibody specific for the effector protein compositions or systems of the present invention. Methods of binding a CRISPR enzyme has been described previously (see, e.g., US20140356867A1) and can be adapted for use with the present invention.

[0348] Several substrates and configurations of devices capable of defining multiple individual discrete volumes within the device may be used. As used herein "individual discrete volume" refers to a discrete space, such as a container, receptacle, or other arbitrary defined volume or space that can be defined by properties that prevent and/or inhibit migration of target molecules, for example a volume or space defined by physical properties such as walls, for example the walls of a well, tube, or a surface of a droplet, which may be impermeable or semipermeable, or as defined by other means such as chemical, diffusion rate limited, electro-magnetic, or light illumination, or any combination thereof that can contain a target molecule

and a indexable nucleic acid identifier (for example nucleic acid barcode). By "diffusion rate limited" (for example diffusion defined volumes) is meant spaces that are only accessible to certain molecules or reactions because diffusion constraints effectively defining a space or volume as would be the case for two parallel laminar streams where diffusion will limit the migration of a target molecule from one stream to the other. By "chemical" defined volume or space is meant spaces where only certain target molecules can exist because of their chemical or molecular properties, such as size, where for example gel beads may exclude certain species from entering the beads but not others, such as by surface charge, matrix size or other physical property of the bead that can allow selection of species that may enter the interior of the bead. By "electro-magnetically" defined volume or space is meant spaces where the electro-magnetic properties of the target molecules or their supports such as charge, or magnetic properties can be used to define certain regions in a space such as capturing magnetic particles within a magnetic field or directly on magnets. By "optically" defined volume is meant any region of space that may be defined by illuminating it with visible, ultraviolet, infrared, or other wavelengths of light such that only target molecules within the defined space or volume may be labeled. One advantage to the use of non-walled, or semipermeable discrete volumes is that some reagents, such as buffers, chemical activators, or other agents may be passed through the discrete volume, while other materials, such as target molecules, may be maintained in the discrete volume or space. Typically, a discrete volume will include a fluid medium, (for example, an aqueous solution, an oil, a buffer, and/or a media capable of supporting cell growth) suitable for labeling of the target molecule with the indexable nucleic acid identifier under conditions that permit labeling. Exemplary discrete volumes or spaces useful in the disclosed methods include droplets (for example, microfluidic droplets and/or emulsion droplets), hydrogel beads or other polymer structures (for example poly-ethylene glycol di-acrylate beads or agarose beads), tissue slides (for example, fixed formalin paraffin embedded tissue slides with particular regions, volumes, or spaces defined by chemical, optical, or physical means), microscope slides with regions defined by depositing reagents in ordered arrays or random patterns, tubes (such as, centrifuge tubes, microcentrifuge tubes, test tubes, cuvettes, conical tubes, and the like), bottles (such as glass bottles, plastic bottles, ceramic bottles, Erlenmeyer flasks, scintillation vials and the like), wells (such as wells in a plate), plates, pipettes, or pipette tips among others. In certain embodiments, the compartment is an aqueous droplet in a water-in-oil emulsion. In specific embodiments, any of the applications,

methods, or systems described herein requiring exact or uniform volumes may employ the use of an acoustic liquid dispenser.

Samples

[0349] The device can be configured to hold, store, collect, receive, process and/or otherwise manipulate a sample and/or detect a component thereof. In some embodiments, the sample is a solid, semisolid, or liquid. In some embodiments, the sample is a biological sample. In some embodiments, the sample is obtained from a subject. In some embodiments, the sample is a bodily fluid. In some embodiments, the bodily fluid is saliva or nasal secretions. In some embodiments, the sample is not a bodily fluid but contains one or more cells from the subject, such as hair cells, skin cells, solid tissue or tumor cells. In some embodiments, the sample is obtained from a plant. In some embodiments, the sample is an environmental sample, such as air, soil, water, or a sample of molecules, organisms, viruses, and other particles present on an object surface. In some embodiments, the sample is a feedstuff or foodstuff or component thereof. Other exemplary samples that may be analyzed using the systems and devices described herein include biological samples of a subject or environmental samples. Environmental samples may include surfaces or fluids. The biological samples may include, but are not limited to, saliva, blood, plasma, sera, stool, urine, sputum, mucous, lymph, synovial fluid, spinal fluid, cerebrospinal fluid, a swab from skin or a mucosal membrane, or combination thereof. In an example embodiment, the environmental sample is taken from a solid surface, such as a surface used in the preparation of food or other sensitive compositions and materials.

[0350] A sample for use with the invention may be a biological or environmental sample, such as a surface sample, a fluid sample, or a food sample (fresh fruits or vegetables, meats). Food samples may include a beverage sample, a paper surface, a fabric surface, a metal surface, a wood surface, a plastic surface, a soil sample, a freshwater sample, a wastewater sample, a saline water sample, exposure to atmospheric air or other gas sample, or a combination thereof. For example, household/commercial/industrial surfaces made of any materials including, but not limited to, metal, wood, plastic, rubber, or the like, may be swabbed and tested for contaminants. Soil samples may be tested for the presence of pathogenic bacteria or parasites, or other microbes, both for environmental purposes and/or for human, animal, or plant disease testing. Water samples such as freshwater samples, wastewater samples, or saline water samples can be evaluated for cleanliness and safety, and/or potability, to detect the presence

of, for example, *Cryptosporidium parvum, Giardia lamblia*, or other microbial contamination. In further embodiments, a biological sample may be obtained from a source including, but not limited to, a tissue sample, saliva, blood, plasma, sera, stool, urine, sputum, mucous, lymph, synovial fluid, spinal fluid, cerebrospinal fluid, ascites, pleural effusion, seroma, pus, bile, aqueous or vitreous humor, transudate, exudate, or swab of skin or a mucosal membrane surface. In some embodiments, the biological sample is a bodily fluid. In some particular embodiments, an environmental sample or biological samples may be crude samples and/or the one or more target molecules may not be purified or amplified from the sample prior to application of the method. Identification of microbes may be useful and/or needed for any number of applications, and thus any type of sample from any source deemed appropriate by one of skill in the art may be used in accordance with the invention.

[0351] In particular embodiments, the methods and systems can be utilized for direct detection from patient samples. In an aspect, the methods and systems can further allow for direct detection from patient samples with a visual readout to further facilitate field-deployability. In an aspect, a field deployable version can include, for example the lateral flow devices and systems as described herein, and/or colorimetric detection. The methods and systems can be utilized to distinguish multiple viral species and strains and identify clinically relevant mutations, important with viral outbreaks such as the coronavirus outbreak in Wuhan (2019-nCoV). In an aspect, the sample is from a nasophyringeal swab or a saliva sample. See., e.g., Wyllie et al., "Saliva is more sensitive for SARS-CoV-2 detection in COVID-19 patients than nasopharyngeal swabs," DOI: 10.1101/2020.04.16.20067835.

Flexible Substrates

[0352] In certain example embodiments, the device comprises a flexible material substrate on which a number of spots or discrete volumes may be defined. Flexible substrate materials suitable for use in diagnostics and biosensing are known within the art. The flexible substrate materials may be made of plant derived fibers, such as cellulosic fibers, or may be made from flexible polymers such as flexible polyester films and other polymer types. Within each defined spot, reagents of the system described herein are applied to the individual spots. Each spot may contain the same reagents except for a different engineered protein of the present invention, different guide RNA or set of guide RNAs, or where applicable, a different detection aptamer to screen for multiple targets at once. Thus, the systems and devices herein may be able to screen samples from multiple sources (e.g., multiple clinical samples from different

individuals) for the presence of the same target, or a limited number of target, or aliquots of a single sample (or multiple samples from the same source) for the presence of multiple different targets in the sample. In certain example embodiments, the elements of the systems described herein are freeze dried onto the paper or cloth substrate. Example flexible material based substrates that may be used in certain example devices are disclosed in Pardee *et al. Cell.* 2016, 165(5):1255-66 and Pardee *et al. Cell.* 2014, 159(4):950-54. Suitable flexible material-based substrates for use with biological fluids, including blood are disclosed in International Patent Application Publication No. WO/2013/071301 entitled "Paper based diagnostic test" to Shevkoplyas *et al.* U.S. Patent Application Publication No. 2011/0111517 entitled "Paper-based microfluidic systems" to Siegel *et al.* and Shafiee *et al.* "Paper and Flexible Substrates as Materials for Biosensing Platforms to Detect Multiple Biotargets" Scientific Reports 5:8719 (2015). Further flexible based materials, including those suitable for use in wearable diagnostic devices are disclosed in Wang *et al.* "Flexible Substrate-Based Devices for Point-of-Care Diagnostics" Cell 34(11):909-21 (2016). Further flexible based materials may include nitrocellulose, polycarbonate, methylethyl cellulose, polyvinylidene fluoride (PVDF), polystyrene, or glass (see e.g., US20120238008). In certain embodiments, discrete volumes are separated by a hydrophobic surface, such as but not limited to wax, photoresist, or solid ink.

[0353] In some embodiments, the substrate, such as a flexible substrate, is a single use substrate, such as swab, strip, or cloth that is used to swab a surface or sample fluid or is placed in a prepared sample for detection by an assay described herein. For example, the system could be used to test for the presence of a pathogen on a food by swabbing the surface of a food product, such as a fruit or vegetable. Similarly, the single use substrate may be used to swab other surfaces for detection of certain microbes or agents, such as for use in security screening. Single use substrates may also have applications in forensics, where the compositions and systems of the present invention are designed to detect, for example identifying DNA SNPs, proteins, and other biomarkers, that may be used to identify a suspect, or certain tissue or cell markers to determine the type of biological matter present in a sample. Likewise, the single use substrate could be used to collect a sample from a patient – such as a saliva sample from the mouth – or a swab of the skin. In other embodiments, a sample or swab may be taken of a meat product on order to detect the presence of absence of contaminants on or within the meat product.

Microfluidic Devices

[0354] In certain example embodiments, the device is configured as a microfluidic device. It will be appreciated that the microfluidic device can incorporate a chip, cartridge, flexible substrate, lateral flow strip, and/or other components described elsewhere herein. In some embodiments, the microfluidic device can be configured to drive a sample through the device such that it contacts one or more detection reaction reagents (such as those that may be present on a flexible substrate within the device) and thus carries out a polypeptide cleavage detection reaction. In some embodiments, the microfluidic device is configured to generate and/or merge different droplets (*i.e.*, individual discrete volumes). For example, a first set of droplets may be formed containing samples to be screened and a second set of droplets formed containing the elements of the systems described herein. The first and second set of droplets are then merged and then diagnostic methods as described herein are carried out on the merged droplet set. Microfluidic devices disclosed herein may be silicone-based chips and may be fabricated using a variety of techniques, including, but not limited to, hot embossing, molding of elastomers, injection molding, LIGA, soft lithography, silicon fabrication and related thin film processing techniques. Suitable materials for fabricating the microfluidic devices include, but are not limited to, cyclic olefin copolymer (COC), polycarbonate, poly(dimethylsiloxane) (PDMS), and poly(methylacrylate) (PMMA). In one embodiment, soft lithography in PDMS may be used to prepare the microfluidic devices. For example, a mold may be made using photolithography which defines the location of flow channels, valves, and filters within a substrate. The substrate material is poured into a mold and allowed to set to create a stamp. The stamp is then sealed to a solid support, such as but not limited to, glass. Due to the hydrophobic nature of some polymers, such as PDMS, which absorbs some proteins and may inhibit certain biological processes, a passivating agent may be necessary (Schoffner *et al. Nucleic Acids Research*, 1996, 24:375-379). Suitable passivating agents are known in the art and include, but are not limited to, silanes, parylene, n-Dodecyl-b-D-matoside (DDM), pluronic, Tween-20, other similar surfactants, polyethylene glycol (PEG), albumin, collagen, and other similar proteins and peptides.

[0355] In certain example embodiments, the system and/or device may be adapted for conversion to a flow-cytometry readout in or allow to sensitive and quantitative measurements of millions of cells in a single experiment and improve upon existing flow-based methods, such as the PrimeFlow assay. In certain example embodiments, cells may be cast in droplets

containing unpolymerized gel monomer, which can then be cast into single-cell droplets suitable for analysis by flow cytometry. A detection construct comprising a fluorescent detectable label may be cast into the droplet comprising unpolymerized gel monomer. Upon polymerization of the gel monomer to form a bead within a droplet. Because gel polymerization is through free-radical formation, the fluorescent reporter becomes covalently bound to the gel. The detection construct may be further modified to comprise a linker, such as an amine. A quencher may be added post-gel formation and will bind via the linker to the reporter construct. Thus, the quencher is not bound to the gel and is free to diffuse away when the reporter is cleaved by the CRISPR effector protein. Amplification of signal in droplet may be achieved by coupling the detection construct to a hybridization chain reaction (HCR initiators) amplification. DNA/RNA hybrid hairpins may be incorporated into the gel which may comprise a hairpin loop that has a RNase sensitive domain. By protecting a strand displacement toehold within a hairpin loop that has a RNase sensitive domain, HCR initiators may be selectively deprotected following cleavage of the hairpin loop by the CRISPR effector protein. Following deprotection of HCR initiators via toehold mediated strand displacement, fluorescent HCR monomers may be washed into the gel to enable signal amplification where the initiators are deprotected.

[0356] An example of microfluidic device that may be used in the context of the invention is described in Hou *et al.* "Direct Detection and drug-resistance profiling of bacteremias using inertial microfluidics" Lap Chip. 15(10):2297-2307 (2016). Further LOC embodiments are described elsewhere herein.

[0100] In one aspect, the embodiments disclosed herein are directed to a nucleic acid, polypeptide, cell, or other molecule detection system comprising a programmable pattern recognition composition or system of the present invention and/or one or more guide RNAs designed to bind to corresponding target molecules (e.g., a target nucleic acid), a reporter construct (also referred to herein as a detection construct in this context), and optional amplification reagents (discussed in greater detail elsewhere herein) to amplify target nucleic acid molecules and/or detectable signals in a sample. Detection compositions and detection constructs of the present invention are described in greater detail elsewhere herein.

Lateral Flow Devices

[0357] In certain embodiments, the device is a lateral flow device. In certain embodiments, the detection assay can be provided on a lateral flow device, as described in International

Publication WO 2019/071051, incorporated herein by reference. The lateral flow device can be adapted to detect one or more coronaviruses and/or other viruses in combination of the coronavirus. The lateral flow device may comprise a flexible substrate, such as a paper substrate or a flexible polymer-based substrate, which can include freeze-dried reagents for detection assays with a visual readout of the assay results. See, WO 2019/071051 at [0145]-[0151] and Example 2, specifically incorporated herein by reference. In an aspect, lyophilized reagents can include preferred excipients that aid in rate of reaction, specificity, or other variables. The excipients may comprise trehalose, histidine, and/or glycine. In certain embodiments, the coronavirus assay can be utilized with isothermal amplification reagents, allowing amplification without complex instrumentation that may be unavailable in the field, as described in WO 2019/071051. Accordingly, the assay can be adapted for field diagnostics, including use of visual readout on a lateral flow device, rapid, sensitive detection and can be deployed for early and direct detection. Colorimetric detection can be utilized and may be particularly suited for field deployable applications, as described in International Application PCT/US2019/015726, published as WO2019/148206. In particular, colorimetric detection can be as described in WO2019/148206 at Figures 102, 105, 107-111 and [00306]-[00324], incorporated herein by reference.

[0358] In one embodiment, the invention provides a lateral flow device comprising a substrate comprising a first end and a second end. The first end may comprise a sample loading portion, a first region comprising a detectable ligand, two or more effector systems of the present invention (e.g., programmable pattern recognition compositions), two or more detection constructs, and one or more first capture regions, each comprising a first binding agent. The substrate may also comprise two or more second capture regions between the first region of the first end and the second end, each second capture region comprising a different binding agent. Each of the two or more effector systems of the present invention may comprise one or more effector proteins and one or more guide sequences, each guide sequence configured to bind one or more target molecules.

[0359] The device may comprise a lateral flow substrate for detecting a polynucleotide and/or polypeptide cleavage, such as a collateral polynucleotide and/or polynucleotide detection reaction. Substrates suitable for use in lateral flow assays are known in the art. These may include but are not necessarily limited to membranes or pads made of cellulose and/or glass fiber, polyesters, nitrocellulose, or absorbent pads (J Saudi Chem Soc 19(6):689-705;

2015), and other embodiments further described herein. The detection system, i.e., one or more programmable pattern recognition compositions or systems and corresponding detection constructs are added to the lateral flow substrate at a defined reagent portion of the lateral flow substrate, typically on one end of the lateral flow substrate. Detection constructs used within the context of the present invention are described in greater detail elsewhere herein. The lateral flow substrate further comprises a sample portion. The sample portion may be equivalent to, continuous with, or adjacent to the reagent portion. In an aspect, the lateral flow substrate can be utilized for visual readout of a detectable signal in one-pot reactions, e.g., wherein steps of extracting nucleic acids, amplifying nucleic acids, and detecting are performed in the same or single individual discrete volume.

Lateral Flow Substrate

[0360] In some embodiments, the device is a lateral flow device. In some embodiments, the lateral flow device can be composed of a composition or system and detection construct of the present invention described elsewhere herein and a lateral flow substrate for carrying out the detection reaction and/or nucleic acid release from the sample.

[0361] In certain example embodiments, a lateral flow device comprises a lateral flow substrate on which detection can be performed. Substrates suitable for use in lateral flow assays are known in the art. These may include, but are not necessarily limited to, membranes or pads made of cellulose and/or glass fiber, polyesters, nitrocellulose, or absorbent pads (*J Saudi Chem Soc* 19(6):689-705; 2015).

[0362] Lateral support substrates comprise a first and second end, and one or more capture regions that each comprise binding agents. The first end may comprise a sample loading portion, a first region comprising a detectable ligand, two or more effector compositions or systems of the present invention, two or more detection constructs, and one or more first capture regions, each comprising a first binding agent. The substrate may also comprise two or more second capture regions between the first region of the first end and the second end, each second capture region comprising a different binding agent. Each of the two or more of the effector compositions or systems of the present invention may comprise one or more effector proteins and one or more guide sequences, each guide sequence configured to bind one or more target molecules. The lateral flow substrates may be configured to detect a reaction mediated by an effector domain of the engineered protein of the present invention, such as a nuclease, protease and/or peptidase.

[0363] Lateral support substrates may be located within a housing (see for example, "Rapid Lateral Flow Test Strips" Merck Millipore 2013). The housing may comprise at least one opening for loading samples and a second single opening or separate openings that allow for reading of detectable signal generated at the first and second capture regions.

[0364] The embodiments disclosed herein can be prepared in freeze-dried format for convenient distribution and point-of-care (POC) applications. Such embodiments are useful in multiple scenarios in human health including, for example, viral detection, bacterial strain typing, sensitive genotyping, and detection of disease-associated cell free DNA. Accordingly, the lateral substrate comprising one or more of the elements of the system, including detectable ligands, effector systems, detection constructs and binding agents may be freeze-dried to the lateral flow substrate and packaged as a ready to use device. Alternatively, all or a portion of the elements of the system may be added to the reagent portion of the lateral flow substrate at the time of using the device.

First End and Second End of the Substrate

[0365] The substrate of the lateral flow device comprises a first and second end. The effector composition or system of the present invention described herein (including any corresponding detection constructs) are added to the lateral flow substrate at a defined reagent portion of the lateral flow substrate, typically on a first end of the lateral flow substrate. Detection constructs used within the context of the present invention are described in greater detail elsewhere herein. The lateral flow substrate can further include a sample portion. The sample portion may be equivalent to, continuous with, or adjacent to the reagent portion.

[0366] In certain example embodiments, the first end comprises a first region. The first region comprises a detectable ligand, two or more effector systems of the present invention (e.g., one or more engineered proteins of the present invention), two or more detection constructs, and one or more first capture regions, each comprising a first binding agent.

Capture Regions

[0367] The lateral flow substrate can comprise one or more capture regions. In embodiments the first end of the lateral flow substrate comprises one or more first capture regions, with two or more second capture regions between the first region of the first end of the substrate and the second end of the substrate. The capture regions may be provided as a capture line, typically a horizontal line running across the device, but other configurations are

possible. The first capture region is proximate to and on the same end of the lateral flow substrate as the sample loading portion.

Binding Agents

[0368] Specific binding-integrating molecules comprise any members of binding pairs that can be used in the present invention. Such binding pairs are known to those skilled in the art and include, but are not limited to, antibody-antigen pairs, enzyme-substrate pairs, receptor-ligand pairs, and streptavidin-biotin. In addition to such known binding pairs, novel binding pairs may be specifically designed. A characteristic of binding pairs is the binding between the two members of the binding pair.

[0369] A first binding agent that specifically binds the first molecule of the reporter construct is fixed or otherwise immobilized to the first capture region. The second capture region is located towards the opposite end of the lateral flow substrate from the first capture region. A second binding agent is fixed or otherwise immobilized at the second capture region. The second binding agent specifically binds the second molecule of the reporter construct, or the second binding agent may bind a detectable ligand. For example, the detectable ligand may be a particle, such as a colloidal particle, that when it aggregates can be detected visually, and generates a detectable positive signal. The particle may be modified with an antibody that specifically binds the second molecule on the reporter construct. If the reporter construct is not cleaved it will facilitate accumulation of the detectable ligand at the first binding region. If the reporter construct is cleaved the detectable ligand is released to flow to the second binding region. In such an embodiment, the second binding region comprises a second binding agent capable of specifically or non-specifically binding the detectable ligand on the antibody of the detectable ligand. Binding agents can be, for example, antibodies, that recognize a particular affinity tag. Such binding agents can further contain, for example, detectable labels, such as isotope labels and/or nucleic acid barcodes. A barcode is a short sequence of nucleotides (for example, DNA, RNA, or combinations thereof) that is used as an identifier. A nucleic acid barcode may have a length of 4-100 nucleotides and be either single or double-stranded. Methods for identifying cells with barcodes are known in the art. Accordingly, guide RNAs of the effector compositions and systems of the present invention may be used to detect the barcode.

Detectable Ligands

[0370] The first region is loaded with a detectable ligand, such as those disclosed herein, for example a gold nanoparticle. The detectable ligand may be a particle, such as a colloidal particle, that when it aggregates can be detected visually. The particle may be modified with an antibody that specifically binds the second molecule on the reporter construct. If the reporter construct is not cleaved it will facilitate accumulation of the detectable ligand at the first binding region. If the reporter construct is cleaved the detectable ligand is released to flow to the second binding region. In such an embodiment, the second binding agent is an agent capable of specifically or non-specifically binding the detectable ligand on the antibody on the detectable ligand. Examples of suitable binding agents for such an embodiment include, but are not limited to, protein A and protein G. In some examples, the detectable ligand is a gold nanoparticle, which may be modified with a first antibody, such as an anti-FITC antibody.

Lateral Flow Detection Constructs

[0371] The first region also comprises a detection construct. In one example embodiment, and for purposes of further illustration, the detection construct may comprise a FAM molecule on a first end of the detection construction and a biotin on a second end of the detection construct. Upstream of the flow of solution from the first end of the lateral flow substrate is a first test band. The test band may comprise a biotin ligand. Accordingly, when the detection construct is present it its initial state, i.e., in the absence of target, the FAM molecule on the first end will bind the anti-FITC antibody on the gold nanoparticle, and the biotin on the second end of the construct will bind the biotin ligand allowing for the detectable ligand to accumulate at the first test, generating a detectable signal. Generation of a detectable signal at the first band indicates the absence of the target ligand. In the presence of target, an effector complex of the present invention forms and an effector protein is activated resulting in cleavage of the detection construct containing a target polypeptide. In the absence of an intact detection construct the colloidal gold will flow past the second strip. The lateral flow device may comprise a second band, upstream of the first band. The second band may comprise a molecule capable of binding the antibody-labeled colloidal gold molecule, for example an anti-rabbit antibody capable of binding a rabbit anti-FITC antibody on the colloidal gold. Therefore, in the presence of one or more targets, the detectable ligand will accumulate at the second band, indicating the presence of the one or more targets in the sample. Other detection constructs

besides the one utilizing colloidal gold may be used in connection with the lateral flow devices herein. Other detection construct are described elsewhere herein.

[0372] In some embodiments, the first end of the lateral flow device comprises two detection constructs and each of the two detection constructs comprises a target polypeptide, comprising a first molecule on a first end and a second molecule on a second end. The first molecule and the second molecule may be linked by a polypeptide linker, such as a target polypeptide.

[0373] In some embodiments, the first molecule on the first end of the first detection construct may be FAM (or a first detection molecule) and the second molecule on the second end of the first detection construct may be biotin (or second detection molecule), or vice versa. In some embodiments, the first molecule on the first end of the second detection construct may be FAM and the second molecule on the second end of the second detection construct may be Digoxigenin (DIG), or vice versa.

[0374] In some embodiments, the first end may comprise three detection constructs, wherein each of the three detection constructs comprises a target polypeptide, comprising a first molecule on a first end and a second molecule on a second end. In specific embodiments, the first and second molecules on the detection constructs comprise Tye 665 and Alexa 488; Tye 665 and FAM, and Tye 665 and Digoxigenin (DIG), respectively. Other detection molecules are described elsewhere herein and can be used in connection with the lateral flow device described herein in view of the guiding principles above.

[0375] In some embodiments, the first end of the lateral flow device comprises two or more effector compositions or systems of the present invention. In some embodiments, such an effector system may include a one or more effector proteins and one or more guide sequences configured to bind to one or more target sequences.

Samples

[0376] When utilizing the detection systems with a lateral flow substrate, samples to be screened are loaded at the sample loading portion of the lateral flow substrate. The samples must be liquid samples or samples dissolved in an appropriate solvent, usually aqueous. The liquid sample reconstitutes the detection reagents such that a detection reaction can occur. The liquid sample begins to flow from the sample portion of the substrate towards the first and second capture regions. Exemplary samples are described in greater detail elsewhere herein. See also WO 2019/071051, which is incorporated by reference herein.

Cartridges and Chips

[0377] The cartridge, also referred to herein as a chip, according to the present invention comprises a series of components of ampoules and chambers that are communicatively coupled with one or more other components on the cartridge. The coupling is typically a fluidic communication, for example, via channels. The cartridge may comprise a membrane that seals one or more of the chambers and/or ampoules. In an aspect, the membrane allows for storage of reagents, buffers and other solid or fluid components which cover and seal the cartridge. The membrane can be configured to be punctured, pierced or otherwise released from sealing or covering one or more components of the cartridge by a means for releasing reagents. In some embodiments, the cartridge contains one or more wells, substrates (e.g., a flexible substrate), or other discrete volumes.

[0378] In some embodiments, the device is configured as lab-on-chip (LOC) diagnostic system. In some embodiments, the LOC is configured as a wireless lab-on-chip (LOC) diagnostic sensor system (see e.g., US patent number 9,470,699). In certain embodiments, the pattern recognition based detection assay is performed in a LOC controlled and/or read by a wireless device (e.g., a cell phone, a personal digital assistant (PDA), a tablet) and results and/or reaction are reported to and/or measured by said device. In some embodiments, the LOC may be a microfluidic device. The LOC may be a passive chip, wherein the chip is powered and controlled through a wireless device. In certain embodiments, the LOC includes a microfluidic channel for holding reagents and a channel for introducing a sample. In certain embodiments, a signal from the wireless device delivers power to the LOC and activates mixing of the sample and assay reagents. Specifically, in the case of the present invention, the system may include a masking agent, effector protein of the composition or system of the present invention (e.g., an engineered protein of the present invention and/or detection composition), and optionally guide RNAs specific for a target molecule. Upon activation of the LOC, the microfluidic device may mix the sample and assay reagents. Upon mixing, a sensor detects a signal and transmits the results to the wireless device. In certain embodiments, the unmasking agent is a conductive RNA or polypeptide molecule. The conductive RNA or polypeptide molecule may be attached to the conductive material. Conductive molecules can be conductive nanoparticles, conductive proteins, metal particles that are attached to the protein or latex or other beads that are conductive. In certain embodiments, if DNA or RNA is used then the conductive molecules can be attached directly to the matching DNA or RNA strands.

The release of the conductive molecules may be detected across a sensor. The assay may be a one step process. Lab-on-the chip technology is well described in the scientific literature and consists of multiple microfluidic channels, input or chemical wells. Reactions in wells can be measured using radio frequency identification (RFID) tag technology since conductive leads from RFID electronic chip can be linked directly to each of the test wells. An antenna can be printed or mounted in another layer of the electronic chip or directly on the back of the device. Furthermore, the leads, the antenna and the electronic chip can be embedded into the LOC chip, thereby preventing shorting of the electrodes or electronics. Since LOC allows complex sample separation and analyses, this technology allows LOC tests to be done independently of a complex or expensive reader. Rather a simple wireless device such as a cell phone or a PDA can be used. In one embodiment, the wireless device also controls the separation and control of the microfluidics channels for more complex LOC analyses. In one embodiment, a LED and other electronic measuring or sensing devices are included in the LOC-RFID chip. Not being bound by a theory, this technology is disposable and allows complex tests that require separation and mixing to be performed outside of a laboratory.

[0101] As noted above, certain embodiments enable the use of nucleic acid binding beads to concentrate target nucleic acid but that do not require elution of the isolated nucleic acid. Thus, in certain example embodiments, the cartridge may further comprise an activatable magnet, such as an electro-magnet. A means for activating the magnet may be located on the device, or the means for supplying the magnet or activating the magnet on the cartridge may be provided by a second device, such as those disclosed in further detail below.

[0102] The overall size of the device may be between 10, 15, 20, 25, 30, 35, 40, 45, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 mm in width, and 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, or 200 mm. The sizing of ampoules, chambers, and channels can be selected to be in line with the reaction volumes discussed herein and to fit within the general size parameters of the overall cartridge.

Ampoules

[0379] The ampoules, also referred to as blisters, allow for storage and release of reagents throughout the cartridge. Ampoules can include liquid or solid reagents, for example, lysis reagents in one ampoule and reaction reagents in another ampoule. The reagents can be as described elsewhere herein and can be adapted for the use in the cartridge or microfluidic or

other device. The ampoule may be sealed by a film that allows for the bursting, puncture or other release of the contents of the ampoules. See, e.g., Becker, H. & Gärtner, C. Microfluidics-enabled diagnostic systems: markets, challenges, and examples. In Microchip Diagnostics: Methods and Protocols (eds Taly, V. *et al.*) (Springer, New York, 2017); Czurratis *et al.*, doi: 10.1088/0960-1317/25/4/045002. Considerations for ampoules can include as discussed in, for example, Smith, S., et al., Blister pouches for effective reagent storage on microfluidic chips for blood cell counting. Microfluid Nanofluid 20, 163 (2016). DOI:10.1007/s10404-016-1830-2. In an aspect, the seal is a frangible seal formed of a composite-layer film that is assembled to the cartridge main body or other part of the device. While referred to herein as an ampoule, the ampoule may comprise a cavity on a chip which comprises a sealed film that is opened by the release means.

Chambers

[0380] The chip, microfluidic device, and/or other device described herein can have one or more chambers. The chambers on the chip may located and sized for fluidic communication via channels or other communication means with ampoules and/or other chambers on the chip. A chamber for receiving a sample can be provided. The sample can be injected, placed in a receptacle into the chamber for receiving a sample, or otherwise transferred to the chamber. A lysis chamber may comprise, for example, capture beads, that may be used for concentration and/or extraction of the desired target material from the sample. Alternatively, the beads may be comprised in an ampoule comprising lysis reagents that are in fluidic communication with the lysis chamber. An amplification chamber may also be provided with, for example, one or more lyophilized components of the system in the amplification chamber and/or communicatively connected to an ampoule comprising one or more components of the amplification reaction.

[0381] When the cartridge comprises a magnet, it may be configured near one or more of the chambers. In an aspect, the magnet is near the lysis well, and may be configured such that the device has a means for activating the magnet. Embodiments comprising a magnet in the cartridge may be utilized with methodologies using magnetic beads for extraction of particular target molecules.

System for Detection Assays

[0382] A system configured for use with the cartridge and to perform an assay, also referred to as a sample analysis apparatus, detection system or detection device, is configured

system to receive the cartridge and conduct an assay comprising isothermal amplification of nucleic acids and detection of target nucleic acids on the cartridge. The system may comprise: a body; a door housing which may be provided in an opened state or a closed state and configured to be coupled to the body of the sample analysis apparatus by a hinge or other closure means; a cartridge accommodating unit included in the detection system and configured to accommodate the cartridge. The system may further comprise one or more means for releasing reagents for extractions, amplification and/or detection; one or more heating means for extractions, amplification and/or detection, a means for mixing reagents for extraction, amplification, and/or detections, and/or a means for reading the results of the assay. The device may further comprise a user interface for programming the device and/or readout of the results of the assay.

Means for Release of Reagents

[0383] The system may comprise means for releasing reagents for extraction, amplification and/or detection. Release of reagents can be performed by a crushing, puncturing, applying heat or pressure until burst, cutting, or other means for the opening of the ampoule and release of contents. e.g., Becker, H. & Gärtner, C. Microfluidics-enabled diagnostic systems: markets, challenges, and examples. In Microchip Diagnostics: Methods and Protocols (eds Taly, V. *et al.*) (Springer, New York, 2017); Czurratis *et al.*, doi: 10.1088/0960-1317/25/4/045002. Mechanical actuators

Heating Means

[0384] The heating means or heating element can be provided, for example, by electrical or chemical elements. One or more heating means can be utilized, or circuits providing regulation of temperature to one or more locations within the detection device can be utilized. In an embodiment, the device is configured to comprise a heating means for heating the lysis (extraction) chamber and at the amplification chamber of the cartridge, sample vessel or other part of the device. In an aspect, the heating element is disposed under the extraction well. The system can be designed with one or more heating means for extraction, amplification and/or detection. In some embodiments, the device does not include a power source. In some embodiments, the heating element provides heat to about 65, 60, 55, 50, 45, 40, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25 degrees C or less. In some embodiments, the device does not contain any heating element.

Power Sources

[0385] In some embodiments, the device can include a power source. The power source can be coupled to one or more of the components of the device. In some embodiments, the power source is electrically coupled to one or more components of the device so as to provide electrical energy to the cone or more components. Suitable power sources that can be incorporated with the device are batteries (single use and rechargeable), solar powered power sources and batteries. In some embodiments, the power source can be coupled to an outside power source (e.g., an electric power grid) so as to recharge the on-board power source. In some embodiments, the device does not include a power source.

Mixing means

[0386] A means for mixing reagents for extraction, amplification and/or detection can be provided. A means for mixing reagents may comprise a means for mixing one or more fluids, or a fluid with a solid or lyophilized reaction mixture can also be provided. Means for mixing that disturb the laminar flow can be provided. In an aspect, the mixing means is a passive mixer, in another aspect, the mixing means is an active mixer. See, e.g., Nam-Trung Nguyen and Zhigang Wu 2005 J. Micromech. Microeng. 15 R1, doi: 10.1088/0960-1317/15/2/R01 for discussion of mixing approaches. In an aspect, the active mixer can be based on external sources such as pressure, temperature, hydrodynamics (with electrical or magnetic forces), dielectrophoresis, electrokinetics, or acoustics. Examples of passive mixing means can be provided by use of geometric approaches, such as a curved path or channel, see, e.g., U.S. Patent 7,160,025, or an expansion/contraction of a channel cross section or diameter. When the cartridge is utilized with beads, channels and wells are configured and sized for the flow of beads.

Means for Reading the Results of the Assay

[0387] A means for reading the results of the assay can be provided in the system. The means for reading the results of the assay will depend in part on the type of detectable signal generated by the assay. In particular embodiments, the assay generates a detectable fluorescent or color readaout. In these instances, the means for reading the results of the assay will be an optic means, for example a single channel or multi-channel optical means such as a fluorimeter, colorimeter or other spectroscopic sensor.

[0388] A combination of means for reading the results of the assay can be utilized, and may include readings such as turbidity, temperature, magnetic, radio, or electrical properties and or optical properties, including scattering, polarization effects, etc.

[0389] The system may further comprise a user interface for programming the device and/or readout of the results of the assay. The user interface may comprise an LED screen. The system can be further configured for a USB port that can allow for docking of four or more devices.

[0390] In an aspect, the system comprises a means for activating a magnet that is disposed within or on the cartridge.

Wearable Devices

[0391] The systems described herein, may further be incorporated into wearable medical devices that assess biological samples, such as biological fluids or an environmental sample, of a subject or in a subject's environment outside the clinic setting and report the outcome of the assay remotely to a central server accessible by a medical care professional. In some embodiments the device may include the ability to self-sample blood, saliva, sweat, such as the devices disclosed in U.S. Patent Application Publication No. 2015/0342509 entitled "Needle-free Blood Draw to Peeters *et al.*, U.S. Patent Application Publication No. 2015/0065821 entitled "Nanoparticle Phoresies" to Andrew Conrad.

[0392] In some embodiments, the device is configured as a dosimeter or badge that serves as a sensor or indicator such that the wearer is notified of exposure to certain microbes or other agents. For example, the systems described herein may be used to detect a particular pathogen. Likewise, aptamer-based embodiments disclosed above may be used to detect both polypeptide as well as other agents, such as chemical agents, to which a specific aptamer may bind. Such a device may be useful for surveillance of soldiers or other military personnel, as well as clinicians, researchers, hospital staff, and the like, in order to provide information relating to exposure to potentially dangerous microbes as quickly as possible, for example for biological or chemical warfare agent detection. In other embodiments, such a surveillance badge may be used for preventing exposure to dangerous microbes or pathogens in immunocompromised patients, burn patients, patients undergoing chemotherapy, children, or elderly individuals.

Other Device Features

[0393] In certain example embodiments, the device may comprise individual wells, such as microplate wells. The size of the microplate wells may be the size of standard 6, 24, 96, 384,

1536, 3456, or 9600 sized wells. In certain example embodiments, the elements of the systems described herein may be freeze dried and applied to the surface of the well prior to distribution and use.

[0394] The devices disclosed herein may further comprise inlet and outlet ports, or openings, which in turn may be connected to valves, tubes, channels, chambers, and syringes and/or pumps for the introduction and extraction of fluids into and from the device. The devices may be connected to fluid flow actuators that allow directional movement of fluids within the microfluidic device. Example actuators include, but are not limited to, syringe pumps, mechanically actuated recirculating pumps, electroosmotic pumps, bulbs, bellows, diaphragms, or bubbles intended to force movement of fluids. In certain example embodiments, the devices are connected to controllers with programmable valves that work together to move fluids through the device. In certain example embodiments, the devices are connected to the controllers discussed in further detail below. The devices may be connected to flow actuators, controllers, and sample loading devices by tubing that terminates in metal pins for insertion into inlet ports on the device.

[0395] As shown herein, the elements of the system are stable when freeze dried or lyophilized, therefore embodiments that do not require a supporting device are also contemplated, i.e., the system may be applied to any surface or fluid that will support the reactions disclosed herein and allow for detection of a positive detectable signal from that surface or solution. In addition to freeze-drying, the systems may also be stably stored and utilized in a pelletized form. Polymers useful in forming suitable pelletized forms are known in the art.

[0396] The devices disclosed herein may also include elements of point of care (POC) devices known in the art for analyzing samples by other methods. See, for example St John and Price, "Existing and Emerging Technologies for Point-of-Care Testing" (Clin Biochem Rev. 2014 Aug; 35(3): 155–167).

[0397] Radio frequency identification (RFID) tag systems include an RFID tag that transmits data for reception by an RFID reader (also referred to as an interrogator). In a typical RFID system, individual objects (e.g., store merchandise) are equipped with a relatively small tag that contains a transponder. The transponder has a memory chip that is given a unique electronic product code. The RFID reader emits a signal activating the transponder within the tag through the use of a communication protocol. Accordingly, the RFID reader is capable of

reading and writing data to the tag. Additionally, the RFID tag reader processes the data according to the RFID tag system application. Currently, there are passive and active type RFID tags. The passive type RFID tag does not contain an internal power source, but is powered by radio frequency signals received from the RFID reader. Alternatively, the active type RFID tag contains an internal power source that enables the active type RFID tag to possess greater transmission ranges and memory capacity. The use of a passive versus an active tag is dependent upon the particular application.

[0398] Since the electrical conductivity of the surface area can be measured precisely quantitative results are possible on the disposable wireless RFID electro-assays. Furthermore, the test area can be very small allowing for more tests to be done in a given area and therefore resulting in cost savings. In certain embodiments, separate sensors each associated with a different CRISPR effector protein and guide RNA immobilized to a sensor are used to detect multiple target molecules. Not being bound by a theory, activation of different sensors may be distinguished by the wireless device.

[0399] In addition to the conductive methods described herein, other methods may be used that rely on RFID or Bluetooth as the basic low-cost communication and power platform for a disposable RFID assay. For example, optical means may be used to assess the presence and level of a given target molecule. In certain embodiments, an optical sensor detects unmasking of a fluorescent masking agent.

[0400] In certain embodiments, the device of the present invention may include handheld portable devices for diagnostic reading of an assay (see e.g., Vashist et al., Commercial Smartphone-Based Devices and Smart Applications for Personalized Healthcare Monitoring and Management, *Diagnostics* 2014, *4*(3), 104-128; mReader from Mobile Assay; and Holomic Rapid Diagnostic Test Reader).

[0401] As noted herein, certain embodiments allow detection via colorimetric change which has certain attendant benefits when embodiments are utilized in POC situations and or in resource poor environments where access to more complex detection equipment to readout the signal may be limited. However, portable embodiments disclosed herein may also be coupled with hand-held spectrophotometers that enable detection of signals outside the visible range. An example of a hand-held spectrophotometer device that may be used in combination with the present invention is described in Das *et al.* "Ultra-portable, wireless smartphone spectrophotometer for rapid, non-destructive testing of fruit ripeness." Nature Scientific

Reports. 2016, 6:32504, DOI: 10.1038/srep32504. Finally, in certain embodiments utilizing quantum dot-based detection constructs, use of a handheld UV light, or other suitable device, may be successfully used to detect a signal owing to the near complete quantum yield provided by quantum dots.

KITS

[0402] Any of the compounds, compositions, formulations, particles, cells, devices, and combinations thereof, described herein or a combination thereof can be presented as a combination kit. As used herein, the terms "combination kit" or "kit of parts" refers to the compounds, compositions, formulations, particles, cells and any additional components that are used to package, sell, market, deliver, and/or administer the combination of elements or a single element, such as the active ingredient, contained therein. Such additional components include, but are not limited to, packaging, syringes, blister packages, dipsticks, substrates, bottles, and the like. The separate kit components can be contained in a single package or in separate packages within the kit.

[0403] In some embodiments, the combination kit also includes instructions printed on or otherwise contained in a tangible medium of expression. The instructions can provide information regarding the content of the compounds, compositions, formulations, particles, cells, devices, described herein or a combination thereof contained therein, safety information regarding the content of the compounds, compositions, formulations, particles, devices, and cells described herein or a combination thereof contained therein, information regarding the dosages, working amounts, indications for use, and/or recommended treatment regimen(s) for the compound(s) formulations, devices, and combinations thereof contained therein. In some embodiments, the instructions can provide directions for sample collection, sample preparation, and/or use of the compounds, compositions, formulations, particles, devices and cells described herein or a combination thereof. In some embodiments, the instructions can be specific to the target(s) being detected by an effector composition or system of the present invention (e.g., a programmable pattern recognition composition or system described herein).

METHODS OF USE

Target Molecule Modification

[0404] The compositions and systems of the present invention can be used to modify a target cell or molecule, such as a target polypeptide and/or target polynucleotide. In some embodiments, a method of modifying a target molecule and/or cell comprises delivering an

engineered protein of the present invention, a polynucleotide of the present invention, a vector or vector system of the present invention, a formulation thereof, or any combination thereof to the target molecule and/or cell, or a sample containing the same, wherein the target molecule and/or cell is or comprises a target polypeptide and activating an effector domain of the engineered protein by allowing binding of the target polypeptide to the recognition domain thereby activating the effector domain via the effector activation domain, wherein effector domain activity modifies the target molecule and/or cell. In some embodiments, modification comprises nucleotide or nucleic acid modification, such including, but not limited to, cleavage, nicking, methylation, demethylation, sequence mutation or modification, base exchange, base editing, any combination thereof and/or the like. In some embodiments, modification comprises polypeptide or amino acid modification, including but not limited to, cleavage, hydrolyzing, acetylation, deacetylation, glycosylation, deglycosylation, phosphorylation, dephosphorylation, any combination thereof, and/or the like.

[0405] In some embodiments the target molecule (e.g., a target polypeptide and/or target polynucleotide) contains or is otherwise associated with a target molecular recognition pattern, such as a PAMP. In some embodiments, the target molecule (e.g., a target polypeptide or polynucleotide) is contained within or on the surface of a cell. In some embodiments, the target molecule (e.g., a target polypeptide or polynucleotide) is exogenous (i.e., not native) to a cell or organism. In some embodiments, the target molecule (e.g., a target polypeptide or polynucleotide) is endogenous or native to the cell or organism to which is introduced. In some embodiments, the exogenous target molecule (e.g., a target polypeptide or polynucleotide) is or is part of a detection composition (e.g., a detection construct) of the present invention. In some embodiments, such as in those methods where an endogenous or exogenous target molecule (e.g., a target polypeptide or polynucleotide) is to be modified, compositions and systems of the present invention are configured to detect an exogenous target molecule (e.g., a target polypeptide or polynucleotide) and thus activation of the engineered protein of the present invention and target molecule modification can be controlled, at least in part, by controlling delivery of the target polynucleotide. In some embodiments, such as in those methods where an endogenous or exogenous target molecule (e.g., a target polypeptide or polynucleotide) is to be modified, compositions and systems of the present invention are configured to detect an endogenous target molecule (e.g., a target polypeptide or polynucleotide), activation of the system and thus target polypeptide modification, occurs only

in cells that contain the target molecule (e.g., a target polypeptide or polynucleotide), such as target proteins, DNA, and/or RNA. In some embodiments target molecule (e.g., a target polypeptide or polynucleotide) modification is cleavage of the target molecule (e.g., a target polypeptide or polynucleotide).

[0406] In some embodiments, the target molecule (e.g., a target polypeptide or polynucleotide) is not contained in or associated with a cell. In some embodiments, the target molecule (e.g., a target polypeptide or polynucleotide) is or is part of a detection construct, such as one that is configured for use in an in vitro detection assay. Such embodiments are described in greater detail elsewhere herein. In some embodiments, the target molecule (e.g., a target polypeptide or polynucleotide) is not contained in or associated with a cell.

[0407] In certain example embodiments, introducing into the sample comprises *in vitro*, *ex vivo*, or *in vivo* delivery of the programable nuclease-peptidase composition into a cell or cell population.

[0408] In certain example embodiments, modification of the one or more target polypeptides and/or polynucleotides results in activation or deactivation of one or more cell-signaling proteins and/or pathways. In some embodiments the cell-signaling protein is a protein involved in any one or more of the following pathways: Akt signaling pathway, AMPK signaling pathway, apoptosis signaling pathway, estrogen signaling pathway, insulin signaling pathway, JAK-STAT signaling pathway, MAPK signaling pathway, mTOR signaling pathway, NF-kappaB signaling pathway, Notch signaling pathway, p53 signaling pathway, TGF-beta signaling pathway, Toll-like receptor signaling pathway, VEGF signaling pathway, Wnt signaling pathway, hedgehog signaling pathway, a cytokine signaling pathway, a growth factor signaling pathway, a PI3K signaling pathway, a PKC signaling pathway, a MEK signaling pathway, a GSK3 beta signaling pathway, and/or the like. In some embodiments the cell-signaling protein is a protein involved in a cytokine receptor mediated pathway, a survival factor receptor mediated signaling pathway, a G-protein coupled receptor mediated signaling pathway, a growth factor receptor, mediated signaling pathway, an integrin mediated signaling pathway, a Frizzled receptor mediated signaling pathway, a Fas receptor mediated signaling pathway, a Patched/SMO receptor mediated signaling pathway.

[0409] In some embodiments, the cell signaling protein is JAK, STAT3, STAT5, Bcl-xL, cytochrome C, caspase 9, caspase 8, FADD, Bad, Bim, Bcl-2, PI3K, Akt, Akkalpha, IkapppaB, PLC, PKC, NFkappaB, G-protein, adenylate cyclase, PKA, Grb2, SOS, Ras, Raf, MEK,

MEKK, MAPK, MKK, Myc, Mad, Max, CREB, ARF, mdm2, Mt, Bax, p53, ERK, Fos, a JNK, Jun, beta cadherin, TCF, a disheveled protein, GSK3beta, APC, Gli, p16, p15, p21, CycIE, CDK2, CycID, CDK4, Rb, E2F, a heat shock protein, insulin, ghrelin, preproghrelin, obestatin, neuropeptideY, erythropoietin, growth hormone, glucagon, vasopressin, calcitonin, adrenocortical hormone, amylin, angiotensin, atrial natriuretic peptide, cholecystokinin, gastrin, secretin, C-peptide, relaxin, pancreatic polypeptide, follicle-stimulating hormone, leptin, luteinizing hormone, melanocyte stimulating hormone, melanotropin, oxytocin, parathyroid hormone, prolactin, renin, somatostatin, thyroid-stimulating hormone, thyrotropin-releasing hormone, substance P, vasoactive intestinal peptide, IFN-gamma, MHC, TCRs, BCRs, activin, inhibin, bone-morophogeneitc proteins, TGF-beta, Smad transcription factors, RXR, IL-1, TNF, and/or the like.

[0410] In certain example embodiments, the one or more target polynucleotides are a specific transcript or set of transcripts and wherein modification of the one or more target polypeptides triggers cell death upon activating the peptidase in response to binding of the nuclease-peptidase to the specific transcript or set of transcripts. In certain embodiments, the guide molecule is configured to detect one or more mutations in the specific transcript or set of transcripts.

[0411] In some embodiments, the method of modifying a polypeptide can be used for, e.g., treating a disease or eliminating a pathogenic microorganism, by triggering apoptosis in the cell or otherwise disrupting signaling, or other function activity of the cell by modifying a polypeptide within said cell. Other applications of the methods of modifying a polypeptide will be appreciated in view of the description herein and, in particular, the polypeptides modified.

Biologic Activity Modulation

[0412] The engineered proteins of the present invention can have application for biologic activity modulation. In some embodiments, the engineered proteins of the present invention are included in an effector system that generally includes a substrate for an effector domain of the programmable pattern recognition composition that is coupled to an effector of interest. Cleavage of the substrate for the effector of the programmable pattern recognition composition directly or indirectly results in activity of the effector of interest, which in turn initiates a biological activity, stops a biological activity, or modulates/modifies a biological activity.

[0413] In some embodiments, one or more components of an effector of interest is expressed in an organism or a cell or cell population thereof. Activity of the effector of interest

is stimulated, stopped, increased, or decreased when the programmable pattern recognition composition of the present invention is activated by recognizing and/or binding a target molecule (e.g., a target polypeptide and/or target polynucleotide) that is contained in, coupled to, or otherwise associated with a target cell, polypeptide, and/or polynucleotide. In some embodiments, the target polynucleotide is endogenous to the cell in which the effector system is expressed. In some embodiments, the target polynucleotide is exogenous to the cell in which the effector system of interest is expressed.

[0414] In some embodiments, the effector of interest is separately expressed from the programmable pattern recognition composition, a target polynucleotide, a target polypeptide, or any combination thereof. Thus, in some embodiments, effector of interest activity is controlled by controlling the timing of co-expression of the effector of interest, the programmable pattern recognition composition, the target polypeptide, and/or the target polynucleotide.

[0415] The effector system of interest can be used to modify a biological activity in a cell or cells so as to impart a functionality to an organism or cell(s) thereof and/or treat and/or prevent a disease, condition, infection, disorder, or any combination thereof in an organism or cell(s) thereof.

Perturbation Screening

[0416] The programmable pattern recognition compositions and systems of the present invention can be used for functional screening, such as a method of perturbation screening. Described in several exemplary embodiments herein are methods for screening cell perturbations comprising introducing a perturbation to a cell population comprising engineered cells as described in greater detail elsewhere herein, along with any elements of a detection composition not already expressed by the engineered cells, and wherein the programmable pattern recognition composition of the present invention is configured to introduce a perturbation in a polynucleotide and/or polypeptide via activation of an effector domain of the engineered programmable pattern recognition composition of the present invention by recognizing and/or binding, a target molecule (e.g., a target polypeptide or target polynucleotide) that is contained in, coupled to, or otherwise associated with (e.g., on a cell surface containing a target polynucleotide or polypeptide) the polynucleotide and/or polypeptide that is to be perturbed. In some embodiments, the programmable pattern recognition composition or component thereof (e.g., an effector domain contained therein),

comprises a nuclease, nickase, protease, peptidase, methylase, acetylase, deacetylase, demethylase, transferase, phosphorylase, dephosphorylatse glycosylase, deglycosylase, and/or the like that is effective to introduce a perturbation in the target polynucleotide and/or polypeptide. In some embodiments the effector is a STAND protein, optionally a STAND NTPase or component thereof. In some embodiments the nuclease or nickase is a Cas. In some embodiments, the programmable pattern recognition composition includes or is delivered with a guide molecule for a CRISPR-Cas system. In some embodiments, when the Cas nuclease is activated by target recognition by the programmable pattern recognition composition of the present invention, a perturbation is introduced into a target polynucleotide. In some embodiments, the target molecule (e.g., target polypeptide or target polynucleotide) and/or target molecular pattern (e.g, a PAMP) recognized by the pattern recognition composition is associated with a specific cell type, cell stated, or cell species (or subspecies). In some embodiments, the guide molecules are configured to detect one or more target transcripts associated with a specific cell type or cell state. In some embodiments, activation of the programmable pattern recognition composition results in production of a detectable product and/or signal, optionally from a detection construct, thus allowing for detection of the perturbation to modify expression of a target polynucleotide and/or target polypeptide by measuring a change in the detectable product or signal relative to a control.

[0417] As is described in greater detail elsewhere herein, the engineered cells into which one or more perturbations are introduced contain a programmable pattern recognition composition or system, such as a detection composition system, of the present invention. Detection constructs and detection assays and devices are described in greater detail elsewhere herein.

[0418] In general perturbation screening is a method of introducing one or more modifications (e.g., perturbations) into the genome and evaluating any change in gene and/or protein expression, phenotype, characteristic, functionality, and/or the like. Methods and tools for genome-scale screening of perturbations in cells, including single cells, using CRISPR-Cas9 have been described, herein referred to as perturb-seq (see e.g., Dixit et al., "Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens" 2016, Cell 167, 1853–1866; Adamson et al., "A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response" 2016, Cell 167, 1867–1882; and International publication serial number WO/2017/075294). A

similar approach may be used with the compositions and systems of the present invention provided herein.

[0419] The compositions and systems present invention are compatible with a detection reaction utilizing a detection composition of the present invention, such that genes, such as signature and/or target, genes may be perturbed, and the perturbation may be identified and assigned to the proteomic and gene expression readouts of single cells or cell populations. In certain embodiments, genes, such as signature or target genes, may be perturbed in single cells and gene expression analyzed. Not being bound by a theory, networks of genes that are disrupted due to perturbation of a signature gene may be determined. Understanding the network of genes effected by a perturbation may allow for a gene to be linked to a specific pathway that may be targeted to modulate the signature and treat a cancer. Thus, in certain embodiments, perturbation is used to discover novel drug and other targets to allow treatment of specific diseases, conditions, etc. at the population, subpopulation, and/or individual patient level.

[0420] The perturbation methods and tools allow reconstructing of a cellular network or circuit. In one embodiment, the method comprises (1) introducing single-order or combinatorial perturbations to a population of cells, (2) measuring genomic, genetic, proteomic, epigenetic and/or phenotypic differences in single cells and (3) assigning a perturbation(s) to the single cells. Not being bound by a theory, a perturbation may be linked to a phenotypic change, preferably changes in gene or protein expression. In preferred embodiments, measured differences that are relevant to the perturbations are determined by applying a model accounting for co-variates to the measured differences. The model may include the capture rate of measured signals, whether the perturbation actually perturbed the cell (phenotypic impact), the presence of subpopulations of either different cells or cell states, and/or analysis of matched cells without any perturbation. In certain embodiments, the measuring of phenotypic differences and assigning a perturbation to a cell or single cell is determined by performing a detection reaction utilizing a detection composition described herein. In some embodiments, barcodes such as nucleic acid barcodes, can be included in the detection composition and/or detection construct such that single cells, or cell populations, detection compositions, detection constructs, target molecules, target polypeptides of the compositions of the present invention, can be distinguished and/or associated with a particular

perturbation and/or result. In some embodiments, the barcode comprises a Unique Molecular Identifier (UMI).

[0421] Perturbations may be introduced into an engineered cell described herein using any suitable method or technique. In some embodiments, perturbations are introduced at least in part using a CRISPR-Cas system or component thereof. In certain embodiments, a programmable pattern recognition protein composition of the present invention and/or CRISPR system or component thereof is used to create an INDEL at one or more target genes. In other embodiments, epigenetic screening is performed by applying CRISPRa/i/x technology (see, e.g., Konermann et al. "Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex" Nature. 2014 Dec 10. doi: 10.1038/nature14136; Qi, L. S., et al. (2013). "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression". Cell. 152 (5): 1173–83; Gilbert, L. A., et al., (2013). "CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes". Cell. 154 (2): 442–51; Komor et al., 2016, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage, Nature 533, 420–424; Nishida et al., 2016, Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems, Science 353(6305); Yang et al., 2016, Engineering and optimising deaminase fusions for genome editing, Nat Commun. 7:13330; Hess et al., 2016, Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells, Nature Methods 13, 1036–1042; and Ma et al., 2016, Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells, Nature Methods 13, 1029–1035). Numerous genetic variants associated with disease phenotypes are found to be in non-coding region of the genome, and frequently coincide with transcription factor (TF) binding sites and non-coding RNA genes. Not being bound by a theory, CRISPRa/i/x approaches may be used to achieve a more thorough and precise understanding of the implication of epigenetic regulation. In one embodiment, a CRISPR system may be used to activate gene transcription. A nuclease-dead RNA-guided DNA binding domain, dCas9, tethered to transcriptional repressor domains that promote epigenetic silencing (e.g., KRAB) may be used for "CRISPRi" that represses transcription. To use dCas9 as an activator (CRISPRa), a guide RNA is engineered to carry RNA binding motifs (e.g., MS2) that recruit effector domains fused to RNA-motif binding proteins, increasing transcription. A key dendritic cell molecule, p65, may be used as a signal amplifier, but is not required. In certain embodiments, the CRISPR-Cas system used to introduce the perturbation(s) includes a Cpf1.

[0422] The engineered cells into which the perturbation(s) are introduced may comprise a cell in a model non-human organism, a model non-human mammal, such as a mouse, non-human primate, and/or the like, that expresses a composition or system of the present invention or component(s) thereof, a mouse that expresses a composition or system of the present invention or component(s) thereof, a cell *in vivo,* or a cell *ex vivo,* or a cell *in vitro* (see e.g., WO 2014/093622 (PCT/US13/074667); US Patent Publication Nos. 20120017290 and 20110265198 assigned to Sangamo BioSciences, Inc.; US Patent Publication No. 20130236946 assigned to Cellectis; Platt et al., "CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling" Cell (2014), 159(2): 440-455; "Oncogenic models based on delivery and use of the crispr-cas systems, vectors and compositions" WO2014204723A1 "Delivery and use of the crispr-cas systems, vectors and compositions for hepatic targeting and therapy" WO2014204726A1; "Delivery, use and therapeutic applications of the crispr-cas systems and compositions for modeling mutations in leukocytes" WO2016049251; and Chen et al., "Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis" 2015, Cell 160, 1246–1260), which can be adapted for use with the present invention described herein.

[0423] In some embodiments, the cell or cells into which perturbations are introduced are tumor cells, such as tumor cells obtained from a subject in need of treatment. In some embodiments, the subject has or is suspected of having a cancer.

[0424] In one embodiment, one or more perturbations are introduced into one or more protein-coding genes or non-protein-coding DNA. In some embodiments, a programmable pattern recognition protein composition of the present invention and/or a CRISPR system or component thereof may be used to knockout protein-coding genes by frameshifts, point mutations, inserts, or deletions. An extensive toolbox may be used for efficient and specific CRISPR system mediated knockout as described herein, including a double-nicking CRISPR to efficiently modify both alleles of a target gene or multiple target loci and a smaller Cas protein for delivery on smaller vectors (Ran, F.A., et al., In vivo genome editing using *Staphylococcus aureus* Cas9. Nature. 520, 186-191 (2015)), which can be used as a guide and adapted for use with the present invention. A genome-wide sgRNA mouse library (~10 sgRNAs/gene) may also be used in a mouse that expresses a suitable Cas protein (see, e.g., WO2014204727A1).

[0425] In one embodiment, perturbation is by deletion of regulatory elements. Non-coding elements may be targeted by using pairs of guide RNAs to delete regions of a defined size, and by tiling deletions covering sets of regions in pools.

[0426] In certain embodiments, whole genome screens can be used for understanding the phenotypic readout of perturbing potential target genes. In preferred embodiments, perturbations target expressed genes as defined by a gene signature using a focused sgRNA library. Libraries may be focused on expressed genes in specific networks or pathways. In other preferred embodiments, regulatory drivers are perturbed.

[0427] Not being bound by a theory, perturbation studies targeting the genes and gene signatures described herein could (1) generate new insights regarding regulation and interaction of molecules within the system that contribute to suppression of an immune response, such as in the case within the tumor microenvironment, and (2) establish potential therapeutic targets or pathways that could be translated into clinical application.

Target Detection

[0428] The programmable pattern recognition compositions and detection compositions described herein can be used in a method of detecting target cells, polypeptides, polynucleotides and/or combinations thereof, such as those present in a sample that contain, are coupled to, or are otherwise associated with a target molecule (e.g., a target polypeptide, target polynucleotide) and/or target molecular pattern, such as a PAMP. Such methods employ one or more of the detection compositions described herein, systems, cells, described herein, and/or devices described herein. Exemplary aspects of the method, e.g., detection constructs and detectable signal generation, are also described in greater detail elsewhere herein. Generally, a method of detection includes binding, complexing, or associating a programmable pattern recognition composition (such as a detection composition) of the present invention with a target molecule, such as one containing, coupled to, or otherwise associated with a molecular pattern (e.g., a PAMP), whereby the programmable pattern recognition protein composition or component thereof (e.g., an effector domain) is activated so as to modify a polypeptide or polynucleotide of a detection composition or component thereof (e.g., a detection construct) to produce a detectable signal, thereby indicating detection of a target cell, polypeptide, polynucleotide or other molecule. Detection can occur, in vitro, in vivo, in situ, or ex vivo. The system can be configured to detect one or more, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more different target polynucleotides. In some embodiments, the STAND NTPase contains peptidase activity

so as to cleave a target peptide in a detection construct. In some embodiments, the programmable pattern recognition composition includes an effector, such as a molecule with nuclease activity that cleaves a polynucleotide of a detection construct. In some embodiments such an effector can be a Cas with or without collateral cleavage activity.

[0429] Described in certain example embodiments are methods of detecting a target molecule and/or cell, the method comprising combining a detection composition of the present invention (i.e., one comprising an engineered protein of the present invention) or a formulation thereof and a sample or component thereof and activating an effector domain of the engineered protein via binding of a target polypeptide in the sample to the recognition domain thereby mediating effector domain modification of the detection construct and generation of a detectable signal.

[0430] In some embodiments, the method further comprising amplifying and/or enriching the target polynucleotide. In some embodiments, activating the peptidase further results in activation or generation of one or more signal amplification molecules.

[0431] Methods employing Cas13 or Cas12 based detection can be used as a general guide for configuration and design of a method, including sample processing, for target molecule, such as nucleic acids, detection methods employing the programmable nuclease-peptidase compositions of the present invention as they related to target nucleic acid preparation and processing (see e.g., Jong et al. N Engl J Med. 2020. 383(15):1492-1494; Broughton, *et al.* CRISPR–Cas12-based detection of SARS-CoV-2. *Nat Biotechnol* (2020), doi:10.1038/s41587-020-0513-4 (DETECTR detection); Gootenberg *et al.*, Science. 2018 Apr 27; 360(6387):439-444. doi: 10.1126/science.aaq0179 (multiplexing lateral flow platform for point-of-care diagnostics); and Chen, *et al.,* Science. 2018 Apr 27;360(6387):436-439. doi: 10.1126/science.aar6245 (Cas12 detection), Myrhvold *et al.*, Science 27 Apr 2018: 360:6387, pp. 444-448; doi:10.1126/science.aas8836 (field deployable viral diagnostics), Joung *et al.*, Point-of-care testing for COVID-19 using SHERLOCK diagnostics" doi: 10.1101/2020.05.04.20091231; Schmid-Burgk, et al., "LAMP-Seq: Population-Scale COVID-19 Diagnostics Using Combinatorial Barcoding," doi: 10.1101/2020.04.06.025635, Gootenberg, 2018; Gootenberg, *et al*, Science. 2017 Apr 28;356(6336):438-442 (2017); Myhrvold, *et al.*, Science 360, 444-448 (2018)). Nucleic acid detection with SHERLOCK relies on the collateral activity of Type VI and Type V Cas proteins, such as Cas13 and Cas12, which unleashes promiscuous cleavage of reporters upon target detection (Gooteneberg *et al.,*

2018)(Abudayyeh, *et al.*, Science. 353(6299)(2016); East-Seletsky *et al. Nature* 538:270-273 (2016); Smargon *et al. Mol Cell* 65(4):618-630 (2017)), Gootenberg, 2018 ; Myhrvold et al. *Science* 360(6387):444-448 (2018); Gootenberg, 2017; Chen et al. *Science* 360(6387):436-439 (2018); Li et al. *Cell Rep* 25(12):3262-3272 (2018); Li et al. *Nat Protoc* 13(5):899-914 (2018), WO 2017/219027, WO2018/107129, US20180298445, US 2018-0274017, US 2018-0305773, WO 2018/170340, U.S. Application 15/922,837, filed March 15, 2018 entitled "Devices for CRISPR Effector System Based Diagnostics", PCT/US18/50091, filed September 7, 2018 "Multi-Effector CRISPR Based Diagnostic Systems", PCT/US18/66940 filed December 20, 2018 entitled "CRISPR Effector System Based Multiplex Diagnostics", PCT/US18/054472 filed October 4, 2018 entitled "CRISPR Effector System Based Diagnostic", U.S. Provisional 62/740,728 filed October 3, 2018 entitled "CRISPR Effector System Based Diagnostics for Hemorrhagic Fever Detection", U.S. Provisional 62/690,278 filed June 26, 2018 and U.S. Provisional 62/767,059 filed November 14, 2018 both entitled "CRISPR Double Nickase Based Amplification, Compositions, Systems and Methods", U.S. Provisional 62/690,160 filed June 26, 2018 and 62,767,077 filed Novemebr 14, 2018, both entitled "CRISPR/CAS and Transposase Based Amplification Compositions, Systems, And Methods", U.S. Provisional 62/690,257 filed June 26, 2018 and 62/767,052 filed November 14, 2018 both entitled "CRISPR Effector System Based Amplification Methods, Systems, And Diagnostics", US Provisional 62/767,076 filed November 14, 2018 entitled "Multiplexing Highly Evolving Viral Variants With SHERLOCK" and 62/767,070 filed November 14, 2018 entitled "Droplet SHERLOCK." Reference is further made to WO2017/127807, WO2017/184786, WO 2017/184768, WO 2017/189308, WO 2018/035388, WO 2018/170333, WO 2018/191388, WO 2018/213708, WO 2019/005866, PCT/US18/67328 filed December 21, 2018 entitled "Novel CRISPR Enzymes and Systems", PCT/US18/67225 filed December 21, 2018 entitled "Novel CRISPR Enzymes and Systems"and PCT/US18/67307 filed December 21, 2018 entitled "Novel CRISPR Enzymes and Systems", US 62/712,809 filed July 31, 2018 entitled "Novel CRISPR Enzymes and Systems", U.S. 62/744,080 filed October 10, 2018 entitled "Novel Cas12b Enzymes and Systems" and U.S. 62/751,196 filed October 26 2018 entitled "Novel Cas12b Enzymes and Systems", U.S. 715,640 filed August 7, 2-18 entitled "Novel CRISPR Enzymes and Systems", WO 2016/205711, U.S. 9,790,490, WO 2016/205749, WO 2016/205764, WO 2017/070605, WO 2017/106657, and WO 2016/149661, WO2018/035387, WO2018/194963, Cox DBT, et al., RNA editing with CRISPR-Cas13,

Science. 2017 Nov 24;358(6366):1019-1027; Gootenberg JS, et al., Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6., Science. 2018 Apr 27;360(6387):439-444; Gootenberg JS, et al., Nucleic acid detection with CRISPR-Cas13a/C2c2., Science. 2017 Apr 28;356(6336):438-442; Abudayyeh OO, et al., RNA targeting with CRISPR-Cas13, Nature. 2017 Oct 12;550(7675):280-284; Smargon AA, et al., Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. Mol Cell. 2017 Feb 16;65(4):618-630.e7; Abudayyeh OO, et al., C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector, Science. 2016 Aug 5;353(6299):aaf5573; Yang L, et al., Engineering and optimising deaminase fusions for genome editing. Nat Commun. 2016 Nov 2;7:13330, Myhrvold et al., Field deployable viral diagnostics using CRISPR-Cas13, Science 2018 360, 444-448, Shmakov et al. "Diversity and evolution of class 2 CRISPR-Cas systems," Nat Rev Microbiol. 2017 15(3):169-182, each of which is incorporated herein by reference in its entirety. Differences in the mechanism of nucleic acid detection and signal generation by a detection construct from such guiding methods and systems will be readily apparent in view of the description herein.

[0432] The low cost and adaptability of the assay platform described herein lends itself to a number of applications including (i) general bacterial and viral potein RNA/DNA quantitation, (ii) rapid, multiplexed RNA/DNA/protein expression detection, and (iii) sensitive detection of target nucleic acids, proteins, and cells, in both clinical and environmental samples. Additionally, the systems disclosed herein may be adapted for detection of transcripts within biological settings, such as cells. Given the highly specific nature of the effectors described herein, it may possible to track allelic specific expression of transcripts or disease-associated mutations and/or the presence of microorganisms in live cells.

[0433] In certain example embodiments, a library of programmable pattern recognition compositions of the present invention is generated with each programmable pattern recognition compositions being capable of recognizing and/or binding a different target molecule (e.g., a target polypeptide and/or polynucleotide) and/or target molecular pattern (e.g., a PAMP), thus being activated by a different target molecule (e.g., a target polypeptide and/or polynucleotide) and/or target molecular pattern and thus recognizing a different target or groups of targets having the same target molecule (e.g., a target polypeptide and/or polynucleotide) and/or target molecular pattern. In some embodiments, each of the programmable pattern recognition compositions are placed in separate volumes and/or compartments. Each volume and/or

compartment may then receive a different sample or aliquot of the same sample. In some embodiments, two or more programmable pattern recognition compositions, each recognizing and/or binding a different target molecule (e.g., a target polypeptide and/or polynucleotide) and/or target molecular pattern, are placed in a single volume, such as droplet, cell, well, or other discrete individual volume. Each volume may then receive a different sample or aliquot of the same sample.

[0434] In certain example embodiments where a guide molecule is co-delivered or is a component of the programmable pattern recognition composition that is a guide nucleic acid for a CRISPR-Cas system, a single guide RNA specific to a single target is placed in separate volumes. Each volume may then receive a different sample or aliquot of the same sample. In certain example embodiments, multiple guide RNA each to separate target may be placed in a single well such that multiple targets may be screened in a different well. In order to detect multiple guide RNAs in a single volume, in certain example embodiments, multiple effector proteins with different specificities may be used. For example, different orthologs with different sequence specificities may be used. For example, one orthologue may preferentially cut A, while others preferentially cut C, U, or T. Accordingly, guide RNAs that are all, or comprise a substantial portion, of a single nucleotide may be generated, each with a different fluorophore. In this way up to four different targets may be screened in a single individual discrete volume.

[0435] In some embodiments, the programmable pattern recognition compositions and systems and methods herein are capable of detecting down to at least attomolar concentrations of target molecules, such as bacterial or viral polynucleotides or polypeptides. In some embodiments, the programmable pattern recognition compositions and systems and methods herein are capable of detecting down to about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or about 100 copies of bacterial or viral DNA, RNA, or polypeptides per microliter (cp/μL). In some embodiments, the programmable pattern recognition compositions and methods herein are capable of detecting down to about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55,

56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or about 100 copies of bacterial or viral DNA, RNA, or polypeptides per microliter (cp/μL) using a fluorescent or colorimetric readout.

[0436] In some embodiments, the detection reaction can occur as a two-step reaction in which amplification of target(s) and target detection via the effector composition/system of the present invention occur in separate reactions. In some embodiments, the detection reaction (including any target and/or signal amplification) can occur as a single, one-pot reaction. In some embodiments where the detection reaction is a one-pot reaction, target amplification is achieved using LAMP or RPA (see also below).

[0437] In some embodiments, the total time to perform the detection method (from sample preparation to detection) can be greater than 0 hours but less than about 4, 3.5, 3, 2.5, 2, 1.5, 1, or 0.5 hours. In some embodiments, the total time to perform the detection method (from sample preparation to detection) can occur within about 20 to 120 minutes, such as within about 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, to/or 120 minutes. In some embodiments, the total time to perform the detection method (from sample preparation to detection) can occur within about 20 to about 60 minutes, e.g. within about 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, or/to 60 minutes. In some embodiments, the total time to perform the detection method (from sample preparation to detection) can occur within about 20 to about 45 minutes, e.g., within about 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, and/or 45 minutes. In some embodiments, the total time to perform the detection method (from sample preparation to detection) can occur within about 20 to about 30 minutes, e.g., within about 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, and/or 30 minutes.

[0438] In some embodiments, the detection reaction can occur within about 1 to about 60 minutes, e.g. within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, to/or about 60 minutes. In some embodiments,

the detection reaction can occur within about 1 to about 45 minutes, e.g. within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, to/or about 45 minutes. In some embodiments, the reaction can occur within about 1 to about 30 minutes, e.g., within about 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, to/or about 30 minutes. In some embodiments, the detection reaction can occur within about 1 to about 25 minutes, e.g., within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, to/or about 25 minutes. In some embodiments, the detection reaction can occur within about 1 to about 20 minutes, e.g., within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, to/or about 20 minutes. In some embodiments, the detection reaction can occur within about 1 to about 15 minutes, e.g., within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, to/or about 15 minutes. In some embodiments, the detection reaction can occur within about 1 to about 10 minutes, e.g., within about 1, 2, 3, 4, 5, 6, 7, 8, 9, to/or about 10 minutes. In some embodiments, the detection reaction can occur within about 1 to about 5 minutes, e.g., within about 1, 2, 3, 4, to/or about 5 minutes.

Sample and Target Nucleic Acid Processing, Isolation, Amplification, and Enrichment

[0439] In some embodiments, a sample and/or target polynucleotides or polypeptides is/are isolated, amplified, and/or enriched, and/or otherwise processed prior to amplification, enrichment, and/or detection. Such processing can include lysis of one or more cells or particles (e.g., viruses, exosomes, virus like particles, and/or the like) present in the sample to release target nucleic acids. In some embodiments, nucleic acids are isolated or otherwise separated from the one or more cells or particles (e.g., viruses, exosomes, virus like particles, and/or the like) present in the sample or sample lysate. In some embodiments, the method does not require or include extraction of the nucleic acids from the sample prior to amplification and/or target detection. In some embodiments, the sample preparation (e.g., lysis) and amplification occur in the same reaction vessel or location.

[0440] In some embodiments, the sample preparation (e.g., lysis), target amplification, and detection occur in the same reaction vessel or location. In some embodiments, the reaction vessel or location contains the sample preparation, amplification, and/or detection compositions and/or systems. In these embodiments, the sample can be added to the vessel and processing, amplification and detection can occur in the same vessel with no requirement to remove or add reagents to the vessel prior to obtaining a result. In some embodiments, the

reagents, compositions, and systems are included in a vessel in a dehydrated (e.g., freeze dried, lyophilized, etc.) form and can be reconstituted when ready to use.

[0441] In some embodiments, the method includes preparation of the reagents for one or more steps, such as sample preparation, amplification, and/or detection, for storage. Such storage preparation can include, but is not limited to lyophilizing, freeze drying, or otherwise dehydrating them. They can be prepared for storage inside of individual reaction vessels or locations within a device or other vessel. In some of these embodiments, the reagents, compositions, systems or combinations thereof are e.g., lyophilized or freeze dried inside of the reaction vessel or at the specific discreet locations on a substrate or otherwise in a device. They can be stored at a suitable temperature ranging from ambient temperature (e.g., about 25-32 degrees C) to about -20 or -80 degrees Celsius. In some embodiments, they are stored for 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 days, weeks, months or years. In some embodiments, the reagents, compositions, systems or combinations thereof are prepared and stored at about 4 degrees C for about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 days, weeks, months or years or more.

[0442] Due to the sensitivity of said systems, a number of applications that require from the rapid and sensitive detection may benefit from the embodiments disclosed herein and are contemplated to be within the scope of the invention. Further, any of the sample and/or nucleic acid processing methods described in this section can be applied, as relevant, to other methods employing the programmable nuclease-peptidase and detection compositions of the present invention herein. It is not intended to limit these features to just methods specifically designed to detect target polynucleotides.

Sample Preparation

[0443] In some embodiments, the sample preparation can include release of polynucleotides (e.g., DNA and/or RNA) and/or polypeptides from cells and/or microorganisms, such as viruses, bacteria, engineered or other cells, particles (e.g., exosomes) etc., present in the sample. In some embodiments, the sample preparation can include virus, bacteria, inactivation and/or nuclease inactivation. The step of sample preparation can occur prior to any target amplification and/or detection. In some embodiments, sample preparation can include nuclease inactivation and/or viral inactivation by 1, 2, 3, 4 or more thermal (heat or cold) inactivation steps, chemical inactivation steps, biologic inactivation, physiologic inactivation, physical inactivation steps, or any combination thereof. The phrase "physiological inactivation" refers to conditions that deviate from the normal working physiological

conditions (e.g., pH, osmolarity, temperature, salinity, etc.) necessary for causing or maintaining the activation of a component (e.g., an enzyme) present in a sample that result in the inactivation or inhibition of the function or activity of the component. Inactivation can, in some embodiments, result in lysis of the cells, microorganisms, viruses, and/or particles. In some embodiments, the same methods and reagents can be applied to other microbes (e.g., bacteria and eukaryotic cells).

Amplification and Enrichment of Target and/or Signal

Target amplification

[0444] In certain example embodiments, target RNAs and/or DNAs may be amplified prior to activating the effector protein of the composition and/or system of the present invention. Any suitable RNA or DNA amplification technique may be used. In certain example embodiments, the RNA or DNA amplification is an isothermal amplification. In certain example embodiments, the isothermal amplification may be nucleic-acid sequenced-based amplification (NASBA), recombinase polymerase amplification (RPA), loop-mediated isothermal amplification (LAMP), strand displacement amplification (SDA), helicase-dependent amplification (HDA), or nicking enzyme amplification reaction (NEAR). In certain example embodiments, non-isothermal amplification methods may be used which include, but are not limited to, PCR, multiple displacement amplification (MDA), rolling circle amplification (RCA), ligase chain reaction (LCR), or ramification amplification method (RAM). In certain embodiments, the amplification can utilize a transposase-based isothermal amplification method (see e.g. WO 2020/006049, which is incorporated by reference herein as if expressed in its entirety), nickase-based isothermal amplification method (see e.g. WO 2020/006067, which is incorporated by reference herein as if expressed in its entirety), or a helicase-based amplification method (see e.g. WO 2020/006036, which is incorporated by reference herein as if expressed in its entirety). In some embodiments, amplification is via LAMP. In some embodiments, amplification is via RPA.

[0445] In certain example embodiments, the RNA or DNA amplification is nucleic acid sequence-based amplification is NASBA, which is initiated with reverse transcription of target RNA by a sequence-specific reverse primer to create a RNA/DNA duplex. RNase H is then used to degrade the RNA template, allowing a forward primer containing a promoter, such as the T7 promoter, to bind and initiate elongation of the complementary strand, generating a double-stranded DNA product. The RNA polymerase promoter-mediated transcription of the

DNA template then creates copies of the target RNA sequence. Importantly, each of the new target RNAs can be detected by the guide RNAs thus further enhancing the sensitivity of the assay. Binding of the target RNAs by the guide RNAs then leads to activation of the effector protein effector protein of the composition and/or system of the present invention and the methods proceed as outlined above. The NASBA reaction has the additional advantage of being able to proceed under moderate isothermal conditions, for example at approximately 41°C, making it suitable for systems and devices deployed for early and direct detection in the field and far from clinical laboratories.

[0446] In certain other example embodiments, a recombinase polymerase amplification (RPA) reaction may be used to amplify the target nucleic acids. RPA reactions employ recombinases which are capable of pairing sequence-specific primers with homologous sequence in duplex DNA. If target DNA is present, DNA amplification is initiated and no other sample manipulation such as thermal cycling or chemical melting is required. The entire RPA amplification system is stable as a dried formulation and can be transported safely without refrigeration. RPA reactions may also be carried out at isothermal temperatures with an optimum reaction temperature of 37-42° C. The sequence specific primers are designed to amplify a sequence comprising the target nucleic acid sequence to be detected. In certain example embodiments, a RNA polymerase promoter, such as a T7 promoter, is added to one of the primers. This results in an amplified double-stranded DNA product comprising the target sequence and a RNA polymerase promoter. After, or during, the RPA reaction, a RNA polymerase is added that will produce RNA from the double-stranded DNA templates. The amplified target RNA can then in turn be detected by the effector system effector protein of the composition and/or system of the present invention. In this way target DNA can be detected using the embodiments disclosed herein. RPA reactions can also be used to amplify target RNA. The target RNA is first converted to cDNA using a reverse transcriptase, followed by second strand DNA synthesis, at which point the RPA reaction proceeds as outlined above.

[0447] Accordingly, in certain example embodiments the systems disclosed herein may include amplification reagents. Different components or reagents useful for amplification of nucleic acids are described herein. For example, an amplification reagent as described herein may include a buffer, such as a Tris buffer. A Tris buffer may be used at any concentration appropriate for the desired application or use, for example including, but not limited to, a concentration of 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, 10 mM, 11

mM, 12 mM, 13 mM, 14 mM, 15 mM, 25 mM, 50 mM, 75 mM, 1 M, or the like. One of skill in the art will be able to determine an appropriate concentration of a buffer such as Tris for use with the present invention.

[0448] A salt, such as magnesium chloride ($MgCl_2$), potassium chloride (KCl), or sodium chloride (NaCl), may be included in an amplification reaction, such as PCR, in order to improve the amplification of nucleic acid fragments. Although the salt concentration will depend on the particular reaction and application, in some embodiments, nucleic acid fragments of a particular size may produce optimum results at particular salt concentrations. Larger products may require altered salt concentrations, typically lower salt, in order to produce desired results, while amplification of smaller products may produce better results at higher salt concentrations. One of skill in the art will understand that the presence and/or concentration of a salt, along with alteration of salt concentrations, may alter the stringency of a biological or chemical reaction, and therefore any salt may be used that provides the appropriate conditions for a reaction of the present invention and as described herein.

[0449] Other components of a biological or chemical reaction may include a cell lysis component in order to break open or lyse a cell for analysis of the materials therein. A cell lysis component may include, but is not limited to, a detergent, a salt as described above, such as NaCl, KCl, ammonium sulfate [$(NH_4)_2SO_4$], or others. Detergents that may be appropriate for the invention may include Triton X-100, sodium dodecyl sulfate (SDS), CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate), ethyl trimethyl ammonium bromide, nonyl phenoxypolyethoxylethanol (NP-40). Concentrations of detergents may depend on the particular application and may be specific to the reaction in some cases. Amplification reactions may include dNTPs and nucleic acid primers used at any concentration appropriate for the invention, such as including, but not limited to, a concentration of 100 nM, 150 nM, 200 nM, 250 nM, 300 nM, 350 nM, 400 nM, 450 nM, 500 nM, 550 nM, 600 nM, 650 nM, 700 nM, 750 nM, 800 nM, 850 nM, 900 nM, 950 nM, 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, 10 mM, 20 mM, 30 mM, 40 mM, 50 mM, 60 mM, 70 mM, 80 mM, 90 mM, 100 mM, 150 mM, 200 mM, 250 mM, 300 mM, 350 mM, 400 mM, 450 mM, 500 mM, or the like. Likewise, a polymerase useful in accordance with the invention may be any specific or general polymerase known in the art and useful or the invention, including Taq polymerase, Q5 polymerase, or the like.

[0450] In some embodiments, amplification reagents as described herein may be appropriate for use in hot-start amplification. Hot start amplification may be beneficial in some embodiments to reduce or eliminate dimerization of adaptor molecules or oligos, or to otherwise prevent unwanted amplification products or artifacts and obtain optimum amplification of the desired product. Many components described herein for use in amplification may also be used in hot-start amplification. In some embodiments, reagents or components appropriate for use with hot-start amplification may be used in place of one or more of the composition components as appropriate. For example, a polymerase or other reagent may be used that exhibits a desired activity at a particular temperature or other reaction condition. In some embodiments, reagents may be used that are designed or optimized for use in hot-start amplification, for example, a polymerase may be activated after transposition or after reaching a particular temperature. Such polymerases may be antibody-based or apatamer-based. Polymerases as described herein are known in the art. Examples of such reagents may include, but are not limited to, hot-start polymerases, hot-start dNTPs, and photo-caged dNTPs. Such reagents are known and available in the art. One of skill in the art will be able to determine the optimum temperatures as appropriate for individual reagents.

[0451] Amplification reagents can include one or more primers and/or probes optimized for amplification of a target sequence by one or more of the amplification methods previously described. Primer and probe design for the methods described herein will be within the purview of one of ordinary skill in the art in view of the context and disclosure only provided herein.

[0452] Amplification of nucleic acids may be performed using specific thermal cycle machinery or equipment and may be performed in single reactions or in bulk, such that any desired number of reactions may be performed simultaneously. In some embodiments, amplification may be performed using microfluidic or robotic devices, or may be performed using manual alteration in temperatures to achieve the desired amplification. In some embodiments, optimization may be performed to obtain the optimum reactions conditions for the particular application or materials. One of skill in the art will understand and be able to optimize reaction conditions to obtain sufficient amplification.

[0453] In certain embodiments, detection of DNA with the methods or systems of the invention requires transcription of the (amplified) DNA into RNA prior to detection.

[0454] In some embodiments, the amplification reagent or component thereof is shelf-stable. In some embodiments, the amplification reagent or component thereof is shelf-stable at ambient temperature.

Target Polynucleotide Enrichment

[0455] In certain example embodiments, target polypeptides, RNA, and/or DNA may first be enriched prior to detection or amplification of the target polypeptides, RNA, and/or DNA. In certain example embodiments, this enrichment may be achieved by binding of the target nucleic acids by a CRISPR effector system or other suitable affinity based capture strategy capable of specifically capturing target nucleic acids so as to allow separation from non-target nucleic acids. In some embodiments, polypeptides are enriched by using a suitable immunoseparation technique or other pull-down type assay. Such techniques for enriching polypeptides are generally known in the art.

[0456] Current target-specific enrichment protocols require single-stranded nucleic acid prior to hybridization with probes. Among various advantages, the present embodiments can skip this step and enable direct targeting to double-stranded DNA (either partly or completely double-stranded). In addition, the embodiments disclosed herein are enzyme-driven targeting methods that offer faster kinetics and easier workflow allowing for isothermal enrichment. In certain example embodiments, a set of guide RNAs to different target nucleic acids are used in a single assay, allowing for detection of multiple targets an/or multiple variants of a single target.

[0457] In certain example embodiments, a dead CRISPR effector protein may bind the target nucleic acid in solution and then subsequently be isolated from said solution. For example, the dead CRISPR effector protein bound to the target nucleic acid, may be isolated from the solution using an antibody or other molecule, such as an aptamer, that specifically binds the dead CRISPR effector protein.

[0458] In other example embodiments, the dead CRISPR effector protein may bound to a solid substrate. A fixed substrate may refer to any material that is appropriate for or can be modified to be appropriate for the attachment of a polypeptide or a polynucleotide. Possible substrates include, but are not limited to, glass and modified functionalized glass, plastics (including acrylics, polystyrene and copolymers of styrene and other materials, polypropylene, polyethylene, polybutylene, polyurethanes, Teflon™, etc.), polysaccharides, nylon or nitrocellulose, ceramics, resins, silica or silica-based materials including silicon and modified

silicon, carbon, metals, inorganic glasses, plastics, optical fiber bundles, and a variety of other polymers. In some embodiments, the solid support comprises a patterned surface suitable for immobilization of molecules in an ordered pattern. In certain embodiments a patterned surface refers to an arrangement of different regions in or on an exposed layer of a solid support. In some embodiments, the solid support comprises an array of wells or depressions in a surface. The composition and geometry of the solid support can vary with its use. In some embodiments, the solids support is a planar structure such as a slide, chip, microchip and/or array. As such, the surface of the substrate can be in the form of a planar layer. In some embodiments, the solid support comprises one or more surfaces of a flowcell. The term "flowcell" as used herein referes to a chamber comprising a solid surface across which one or more fluid reagent can be flowed. Example flowcells and related fluidic systems and detection platforms that can be readily used in the methods of the present disclosure are described, for example, in Bentley *et al.* Nature 456:53-59 (2008), WO 04/0918497, U.S. 7,057,026; WO 91/06678; WO 07/123744; US 7,329,492; US 7,211,414; US 7,315,019; U.S. 7,405,281, and US 2008/0108082. In some embodiments, the solid support or its surface is non-planar, such as the inner or outer surface of a tube or vessel. In some embodiments, the solid support comprise microspheres or beads. "Microspheres," "bead," "particles," are intended to mean within the context of a solid substrate to mean small discrete particles made of various material including, but not limited to, plastics, ceramics, glass, and polystyrene. In certain embodiments, the microspheres are magnetic microspheres or beads. Alternatively or additionally, the beads may be porous. The bead sizes range from nanometers, e.g., 100 nm, to millimeters, e.g., 1 mm.

[0459] A sample containing, or suspected of containing, the target nucleic acids may then be exposed to the substrate to allow binding of the target nucleic acids to the bound dead CRISPR effector protein. Non-target molecules may then be washed away. In certain example embodiments, the target nucleic acids may then be released from the CRISPR effector protein/guide RNA complex for further detection using the methods disclosed herein. In certain example embodiments, the target nucleic acids may first be amplified as described herein. In certain example embodiments, the CRISPR effector may be labeled with a binding tag. In certain example embodiments the CRISPR effector may be chemically tagged. For example, the CRISPR effector may be chemically biotinylated. In another example embodiment, a fusion may be created by adding additional sequence encoding a fusion to the CRISPR effector. One

example of such a fusion is an AviTag™, which employs a highly targeted enzymatic conjugation of a single biotin on a unique 15 amino acide peptide tag. In certain embodiments, the CRISPR effector may be labeled with a capture tag such as, but not limited to, GST, Myc, hemagglutinin (HA), green fluorescent protein (GFP), flag, His tag, TAP tag, and Fc tag. The binding tag, whether a fusion, chemical tag, or capture tag, may be used to either pull down the CRISPR effector system once it has bound a target nucleic acid or to fix the CRISPR effector system on the solid substrate.

[0460] In certain example embodiments, a guide RNA may be labeled with a binding tag. In certain example embodiments, the entire guide RNA may be labeled using *in vitro* transcription (IVT) incorporating one or more biotinylated nucleotides, such as, biotinylated uracil. In some embodiments, biotin can be chemically or enzymatically added to the guide RNA, such as, the addition of one or more biotin groups to the 3' end of the guide RNA. The binding tag may be used to pull down the guide RNA/target nucleic acid complex after binding has occurred, for example, by exposing the guide RNA/target nucleic acid to a streptavidin coated solid substrate.

[0461] Accordingly, in certain example embodiments, an engineered or non-naturally-occurring CRISPR effector may be used for enrichment purposes. In an embodiment, the modification may comprise mutation of one or more amino acid residues of the effector protein. The one or more mutations may be in one or more catalytically active domains of the effector protein. The effector protein may have reduced or abolished nuclease activity compared with an effector protein lacking said one or more mutations. The effector protein may not direct cleavage of the RNA strand at the target locus of interest. In a preferred embodiment, the one or more mutations may comprise two mutations. In a preferred embodiment the one or more amino acid residues are modified in a C2c2 effector protein, e.g., an engineered or non-naturally-occurring effector protein or C2c2. In particular embodiments, the one or more modified of mutated amino acid residues are one or more of those in C2c2 corresponding to R597, H602, R1278 and H1283 (referenced to Lsh C2c2 amino acids), such as mutations R597A, H602A, R1278A and H1283A, or the corresponding amino acid residues in Lsh C2c2 orthologues.

[0462] In particular embodiments, the one or more modified of mutated amino acid residues are one or more of those in C2c2 corresponding to K2, K39, V40, E479, L514, V518, N524, G534, K535, E580, L597, V602, D630, F676, L709, I713, R717 (HEPN), N718, H722

(HEPN), E773, P823, V828, I879, Y880, F884, Y997, L1001, F1009, L1013, Y1093, L1099, L1111, Y1114, L1203, D1222, Y1244, L1250, L1253, K1261, I1334, L1355, L1359, R1362, Y1366, E1371, R1372, D1373, R1509 (HEPN), H1514 (HEPN), Y1543, D1544, K1546, K1548, V1551, I1558, according to C2c2 consensus numbering. In certain embodiments, the one or more modified of mutated amino acid residues are one or more of those in C2c2 corresponding to R717 and R1509. In certain embodiments, the one or more modified of mutated amino acid residues are one or more of those in C2c2 corresponding to K2, K39, K535, K1261, R1362, R1372, K1546 and K1548. In certain embodiments, said mutations result in a protein having an altered or modified activity. In certain embodiments, said mutations result in a protein having a reduced activity, such as reduced specificity. In certain embodiments, said mutations result in a protein having no catalytic activity (i.e., "dead" C2c2). In an embodiment, said amino acid residues correspond to Lsh C2c2 amino acid residues, or the corresponding amino acid residues of a C2c2 protein from a different species.

[0463] The above enrichment systems may also be used to deplete a sample of certain nucleic acids. For example, guide RNAs may be designed to bind non-target RNAs to remove the non-target RNAs from the sample. In one example embodiment, the guide RNAs may be designed to bind nucleic acids that do carry a particular nucleic acid variation. For example, in a given sample a higher copy number of non-variant nucleic acids may be expected. Accordingly, the embodiments disclosed herein may be used to remove the non-variant nucleic acids from a sample, to increase the efficiency with which the detection effector system effector protein of the composition and/or system of the present invention can detect the target variant sequences in a given sample.

Amplification and/or Enhancement of Detectable Signal

[0464] In certain example embodiments, further modification or reagents may be introduced that further amplify the detectable positive signal. For example, an activated effector domain of of an engineered protein the present invention may be used to generate a secondary target or additional guide sequence, or both. In one example embodiment, the reaction solution would contain a secondary target polypeptide that is spiked in at high concentration. The secondary target polypeptide may be distinct from the primary target polypeptide (i.e., the first target polypeptide e for which the assay is designed to detect) and in certain instances may be common across all reaction volumes. A secondary polypeptide may include a protecting group such that is not active until acted upon by the effector protein.

Cleavage of the protecting group by an activated effector protein (i.e., after activation by formation of complex with the primary target(s) in solution) and formation of a complex with free effector protein in solution and activation from the spiked in secondary target polypeptide.

[0465] In some embodiments a CRISPR system can be used to enrich or amplify the detectable signal. In some embodiments, the programmable pattern recognition compositions and systems of the present invention that is/are activated upon target recognition and/or binding can produce, such as via collateral (e.g., peptidase, nuclease, etc.) activity of one or more components of the programmable pattern recognition composition, species that can activate (or be targets of) a CRISPR system (such as a Cas-12 or Cas-13 detection system) thus amplifying the signal for detection. In some embodiments a CRISPR type-III effector can be used as the signal amplifying system. In some embodiments, the type III effector is Csm6, which is which is activated by cyclic adenylate molecules or linear adenine homopolymers terminated with a 2′,3′-cyclic phosphate. In some embodiments, the first CRISPR system includes a Cas13 (e.g., Cas 13a, 13b, 13c, or 13d) and/or a Cas 12a effector(s) and the amplification system or molecule is or includes Csm6. See also Gootenberg et al. 2018. Science. 360:439-44 and WO 2019/051318, which are incorporated by reference herein as if expressed in their entireties.

Exemplary Applications of the Target Polynucleotide Detection Methods

Microbe and Virus Detection and Applications

[0466] In certain example embodiments, the systems, devices, and methods, disclosed herein are directed to detecting the presence of one or more microbial agents in a sample, such as a biological sample obtained from a subject. In certain example embodiments, the microbe may be a bacterium, a fungus, a yeast, a protozoa, a parasite, or a virus. Accordingly, the methods disclosed herein can be adapted for use in other methods (or in combination) with other methods that require quick identification of microbe species, monitoring the presence of microbial proteins (antigens), antibodies, antibody genes, detection of certain phenotypes (e.g., bacterial resistance), monitoring of disease progression and/or outbreak, and antibiotic screening. Because of the rapid and sensitive diagnostic capabilities of the embodiments disclosed here, detection of microbe species type, down to a single nucleotide difference, and the ability to be deployed as a point of care (POC) device, the embodiments disclosed herein may be used guide therapeutic regimens, such as selection of the appropriate antibiotic or antiviral. The embodiments disclosed herein may also be used to screen environmental samples (air, water, surfaces, food etc.) for the presence of microbial contamination.

[0467] Disclosed is a method to identify microbial species, such as bacterial, viral, fungal, yeast, or parasitic species, or the like. Particular embodiments disclosed herein describe methods and systems that will identify and distinguish microbial species within a single sample, or across multiple samples, allowing for recognition of many different microbes. The present methods allow the detection of pathogens and distinguishing between two or more species of one or more organisms, e.g., bacteria, viruses, yeast, protozoa, and fungi or a combination thereof, in a biological or environmental sample, by detecting the presence of a target nucleic acid sequence in the sample. A positive signal obtained from the sample indicates the presence of the microbe. Multiple microbes can be identified simultaneously using the methods and systems of the invention, by employing the use of more than one effector protein, wherein each effector protein targets a specific microbial target sequence. In this way, a multi-level analysis can be performed for a particular subject in which any number of microbes can be detected at once. In some embodiments, simultaneous detection of multiple microbes may be performed using a set of probes that can identify one or more microbial species.

[0468] Multiplex analysis of samples enables large-scale detection of samples, reducing the time and cost of analyses. However, multiplex analyses are often limited by the availability of a biological sample. In accordance with the invention, however, alternatives to multiplex analysis may be performed such that multiple effector proteins can be added to a single sample and each detection construct may be combined with a separate quencher dye. In this case, positive signals may be obtained from each quencher dye separately for multiple detection in a single sample.

[0469] Disclosed herein are methods for distinguishing between two or more species of one or more organisms in a sample. The methods are also amenable to detecting one or more species of one or more organisms in a sample.

Microbe Detection

[0470] In some embodiments, a method for detecting microbes in samples is provided comprising distributing a sample or set of samples into one or more individual discrete volumes, the individual discrete volumes comprising a programmable pattern recognition composition of the present invention; incubating the sample or set of samples under conditions sufficient to allow recognition and/or binding of the programmable pattern recognition composition to a target molecule (e.g., a target polypeptide and/or polynucleotide) and/or target molecular pattern on, in or secreted by one or more microbe targets; activating one or more

effector domains of the programmable pattern recognition composition and/or component thereof via recognition and/or binding of the programmable pattern recognition composition to a target molecule (e.g., a target polypeptide and/or polynucleotide) and/or target molecular pattern of the one or more target microbes or molecules, wherein activating the programmable pattern recognition composition results in modification of a detection construct, such as a polypeptide or polynucleotide (e.g., RNA)-based detection construct, such that a detectable positive signal is generated; and detecting the detectable positive signal, wherein detection of the detectable positive signal indicates a presence of one or more target molecules or cells in the sample containing a target molecule (e.g., a target polypeptide and/or polynucleotide) and/or target molecular pattern or other recognized molecular pattern. The one or more target molecules may be mRNA, gDNA (coding or non-coding), trRNA, RNA, or peptides or polypeptides. The guide RNAs may be designed to detect target sequences. Where the systems include or involve guide RNAs, cartain embodiments disclosed herein may also utilize certain steps to improve hybridization between guide RNA and target RNA sequences. Methods for enhancing ribonucleic acid hybridization are disclosed in WO 2015/085194, entitled "Enhanced Methods of Ribonucleic Acid Hybridization" which is incorporated herein by reference. The microbe-specific target may be RNA or DNA or a protein. If DNA, the method may further comprise the use of DNA primers that introduce an RNA polymerase promoter as described herein. If the target is a protein then aptamers can be utilized and the method includes one or more specific to protein detection described herein.

Detection of Single Nucleotide Variants

[0471] In some embodiments, one or more identified target sequences may be detected using engineered proteins of the present invention and/or guide RNAs that are specific for and bind to the target sequence as described herein. The systems and methods of the present invention can distinguish even between single nucleotide polymorphisms present among different microbial species and therefore, use of multiple guide RNAs in accordance with the invention may further expand on or improve the number of target sequences that may be used to distinguish between species. For example, in some embodiments, the one or more guide RNAs may distinguish between microbes at the species, genus, family, order, class, phylum, kingdom, or phenotype, or a combination thereof. This application can also apply to non-microbial cells, such as human cells in detection of disease or genotyping.

Detection Based on rRNA Sequences

[0472] In certain example embodiments, the devices, systems, and methods disclosed herein may be used to distinguish multiple microbial species in a sample. In certain example embodiments, identification may be based on ribosomal RNA sequences, including the 16S, 23S, and 5S subunits. Methods for identifying relevant rRNA sequences are disclosed in U.S. Patent Application Publication No. 2017/0029872. In certain example embodiments, a set of guide RNA may designed to distinguish each species by a variable region that is unique to each species or strain. Guide RNAs may also be designed to target RNA genes that distinguish microbes at the genus, family, order, class, phylum, kingdom levels, or a combination thereof. In certain example embodiments where amplification is used, a set of amplification primers may be designed to flanking constant regions of the ribosomal RNA sequence and a guide RNA designed to distinguish each species by a variable internal region. In certain example embodiments, the primers and guide RNAs may be designed to conserved and variable regions in the 16S subunit respectfully. Other genes or genomic regions that uniquely variable across species or a subset of species such as the RecA gene family, RNA polymerase β subunit, may be used as well. Other suitable phylogenetic markers, and methods for identifying the same, are discussed for example in Wu et al. arXiv:1307.8690 [q-bio.GN].

[0473] In certain example embodiments, a method or diagnostic is designed to screen microbes across multiple phylogenetic and/or phenotypic levels at the same time. For example, the method or diagnostic may comprise the use of multiple detection compositions or systems of the present invention with different guide RNAs. A first set of guide RNAs may distinguish, for example, between mycobacteria, gram positive, and gram-negative bacteria. These general classes can be even further subdivided. For example, guide RNAs could be designed and used in the method or diagnostic that distinguish enteric and non-enteric within gram negative bacteria. A second set of guide RNA can be designed to distinguish microbes at the genus or species level. Thus, a matrix may be produced identifying all mycobacteria, gram positive, gram negative (further divided into enteric and non-enteric) with each genus of species of bacteria identified in a given sample that fall within one of those classes. In some embodiments, identification of microbes is based on other target molecules, such as polypeptides, or other microbe specific structural features. In some embodiments, identification of microbes is based on target molecular patterns, such as PAMPS. The foregoing is for example purposes only.

Other means for classifying other microbe types are also contemplated and would follow the general structure described above.

Screening for Drug Resistance

[0474] In certain example embodiments, the devices, systems and methods disclosed herein may be used to screen for microbial genes and/or proteins of interest, for example antibiotic and/or antiviral resistance genes/proteins. Guide RNAs may be designed to distinguish between known genes of interest. Samples, including clinical samples, may then be screened using the embodiments disclosed herein for detection of such genes. The ability to screen for drug resistance at POC would have tremendous benefit in selecting an appropriate treatment regime. In certain example embodiments, the antibiotic resistance genes are carbapenemases including KPC, NDM1, CTX-M15, OXA-48. Other antibiotic resistance genes are known and may be found for example in the Comprehensive Antibiotic Resistance Database (Jia *et al.* "CARD 2017: expansion and model-centric curation of the Comprehensive Antibiotic Resistance Database." Nucleic Acids Research, 45, D566-573).

[0475] Ribavirin is an effective antiviral that hits a number of RNA viruses. Several clinically important viruses have evolved ribavirin resistance including Foot and Mouth Disease Virus doi:10.1128/JVI.03594-13; polio virus (Pfeifer and Kirkegaard. PNAS, 100(12):7289-7294, 2003); and hepatitis C virus (Pfeiffer and Kirkegaard, J. Virol. 79(4):2346-2355, 2005). A number of other persistent RNA viruses, such as hepatitis and HIV, have evolved resistance to existing antiviral drugs: hepatitis B virus (lamivudine, tenofovir, entecavir) doi:10/1002/hep22900; hepatitis C virus (telaprevir, BILN2061, ITMN-191, SCh6, boceprevir, AG-021541, ACH-806) doi:10.1002/hep.22549; and HIV (many drug resistance mutations) hivb.standford.edu. The embodiments disclosed herein may be used to detect such variants among others.

[0476] Aside from drug resistance, there are a number of clinically relevant mutations that could be detected with the embodiments disclosed herein , such as persistent versus acute infection in LCMV (doi:10.1073/pnas.1019304108), and increased infectivity of Ebola (Diehl *et al.* Cell. 2016, 167(4):1088-1098.

[0477] As described herein elsewhere, closely related microbial species (e.g., having only a single nucleotide difference in a given target sequence) may be distinguished by introduction of a synthetic mismatch in the gRNA.

Set Cover Approaches

[0478] In particular embodiments, a set of guide RNAs is employed and designed that can identify, for example, all microbial species within a defined set of microbes. In certain example embodiments, the methods for generating guide RNAs as described herein may be compared to methods disclosed in WO 2017/040316, incorporated herein by reference. As described in WO 2017040316, a set cover solution may identify the minimal number of target sequences probes or guide RNAs needed to cover an entire target sequence or set of target sequences, e.g. a set of genomic sequences. Set cover approaches have been used previously to identify primers and/or microarray probes, typically in the 20 to 50 base pair range. *See*, *e.g.*, Pearson *et al.*, cs.virginia.edu/~robins/papers/primers_dam11_final.pdf., Jabado *et al.* Nucleic Acids Res. 2006 34(22):6605-11, Jabado *et al.* Nucleic Acids Res. 2008, 36(1):e3 doi10.1093/nar/gkm1106, Duitama *et al.* Nucleic Acids Res. 2009, 37(8):2483-2492, Phillippy *et al.* BMC Bioinformatics. 2009, 10:293 doi:10.1186/1471-2105-10-293. However, such approaches generally involved treating each primer/probe as k-mers and searching for exact matches or allowing for inexact matches using suffix arrays. In addition, the methods generally take a binary approach to detecting hybridization by selecting primers or probes such that each input sequence only needs to be bound by one primer or probe and the position of this binding along the sequence is irrelevant. Alternative methods may divide a target genome into pre-defined windows and effectively treat each window as a separate input sequence under the binary approach – *i.e.*, they determine whether a given probe or guide RNA binds within each window and require that all of the windows be bound by the same of some probe or guide RNA. Effectively, these approaches treat each element of the "universe" in the set cover problem as being either an entire input sequence or a pre-defined window of an input sequence, and each element is considered "covered" if the start of a probe or guide RNA binds within the element. These approaches limit the fluidity to which different probe or guide RNA designs are allowed to cover a given target sequence.

[0479] In contrast, the embodiments disclosed herein are directed to detecting longer probe or guide RNA lengths, for example, in the range of 70 bp to 200 bp that are suitable for hybrid selection sequencing. In addition, the methods disclosed WO 2017/040316 herein may be applied to take a pan-target sequence approach capable of defining a probe or guide RNA sets that can identify and facilitate the detection sequencing of all species and/or strains sequences in a large and/or variable target sequence set. For example, the methods disclosed herein may

be used to identify all variants of a given virus, or multiple different viruses in a single assay. Further, the method disclosed herein treat each element of the "universe" in the set cover problem as being a nucleotide of a target sequence, and each element is considered "covered" as long as a probe or guide RNA binds to some segment of a target genome that includes the element. These type of set cover methods may be used instead of the binary approach of previous methods, the methods disclosed in herein better model how a probe or guide RNA may hybridize to a target sequence. Rather than only asking if a given guide RNA sequence does or does not bind to a given window, such approaches may be used to detect a hybridization pattern – *i.e.*, where a given probe or guide RNA binds to a target sequence or target sequences – and then determines from those hybridization patterns the minimum number of probes or guide RNAs needed to cover the set of target sequences to a degree sufficient to enable both enrichment from a sample and sequencing of any and all target sequences. These hybridization patterns may be determined by defining certain parameters that minimize a loss function, thereby enabling identification of minimal probe or guide RNA sets in a way that allows parameters to vary for each species, *e.g.*, to reflect the diversity of each species, as well as in a computationally efficient manner that cannot be achieved using a straightforward application of a set cover solution, such as those previously applied in the probe or guide RNA design context.

[0480] The ability to detect multiple transcript abundances may allow for the generation of unique microbial signatures indicative of a particular phenotype. Various machine learning techniques may be used to derive the gene signatures. Accordingly, the guide RNAs of the detection compositions/systems of the present invention may be used to identify and/or quantitate relative levels of biomarkers defined by the gene signature in order to detect certain phenotypes. In certain example embodiments, the gene signature indicates susceptibility to an antibiotic, resistance to an antibiotic, or a combination thereof.

[0481] In one aspect of the invention, a method comprises detecting one or more pathogens. In this manner, differentiation between infection of a subject by individual microbes may be obtained. In some embodiments, such differentiation may enable detection or diagnosis by a clinician of specific diseases, for example, different variants of a disease. Preferably the pathogen sequence is a genome of the pathogen or a fragment thereof. The method further may comprise determining the evolution of the pathogen. Determining the evolution of the pathogen may comprise identification of pathogen mutations, e.g., nucleotide

deletion, nucleotide insertion, nucleotide substitution. Amongst the latter, there are non-synonymous, synonymous, and noncoding substitutions. Mutations are more frequently non-synonymous during an outbreak. The method may further comprise determining the substitution rate between two pathogen sequences analyzed as described above. Whether the mutations are deleterious or even adaptive would require functional analysis, however, the rate of non-synonymous mutations suggests that continued progression of this epidemic could afford an opportunity for pathogen adaptation, underscoring the need for rapid containment. Thus, the method may further comprise assessing the risk of viral adaptation, wherein the number non-synonymous mutations is determined. (Gire, *et al.*, *Science* 345, 1369, 2014).

Monitoring Microbe Outbreaks

[0482] In some embodiments, a detection composition of the present invention or methods of use thereof as described herein may be used to determine the evolution of a pathogen outbreak. The method may comprise detecting one or more target sequences from a plurality of samples from one or more subjects, wherein the target sequence is a sequence from a microbe causing the outbreaks. Such a method may further comprise determining a pattern of pathogen transmission, or a mechanism involved in a disease outbreak caused by a pathogen.

[0483] The pattern of pathogen transmission may comprise continued new transmissions from the natural reservoir of the pathogen or subject-to-subject transmissions (e.g., human-to-human transmission) following a single transmission from the natural reservoir or a mixture of both. In one embodiment, the pathogen transmission may be bacterial or viral transmission, in such case, the target sequence is preferably a microbial genome or fragments thereof. In one embodiment, the pattern of the pathogen transmission is the early pattern of the pathogen transmission, i.e., at the beginning of the pathogen outbreak. Determining the pattern of the pathogen transmission at the beginning of the outbreak increases likelihood of stopping the outbreak at the earliest possible time thereby reducing the possibility of local and international dissemination.

[0484] Determining the pattern of the pathogen transmission may comprise detecting a pathogen sequence according to the methods described herein. Determining the pattern of the pathogen transmission may further comprise detecting shared intra-host variations of the pathogen sequence between the subjects and determining whether the shared intra-host variations show temporal patterns. Patterns in observed intrahost and interhost variation provide important insight about transmission and epidemiology (Gire, *et al.*, 2014).

[0485] Detection of shared intra-host variations between the subjects that show temporal patterns is an indication of transmission links between subject (in particular between humans) because it can be explained by subject infection from multiple sources (superinfection), sample contamination recurring mutations (with or without balancing selection to reinforce mutations), or co-transmission of slightly divergent viruses that arose by mutation earlier in the transmission chain (Park, *et al.*, *Cell* 161(7):1516–1526, 2015). Detection of shared intra-host variations between subjects may comprise detection of intra-host variants located at common single nucleotide polymorphism (SNP) positions. Positive detection of intra-host variants located at common (SNP) positions is indicative of superinfection and contamination as primary explanations for the intra-host variants. Superinfection and contamination can be parted on the basis of SNP frequency appearing as inter-host variants (Park, et al., 2015). Otherwise, superinfection and contamination can be ruled out. In this latter case, detection of shared intra-host variations between subjects may further comprise assessing the frequencies of synonymous and nonsynonymous variants and comparing the frequency of synonymous and nonsynonymous variants to one another. A nonsynonymous mutation is a mutation that alters the amino acid of the protein, likely resulting in a biological change in the microbe that is subject to natural selection. Synonymous substitution does not alter an amino acid sequence. Equal frequency of synonymous and nonsynonymous variants is indicative of the intra-host variants evolving neutrally. If frequencies of synonymous and nonsynonymous variants are divergent, the intra-host variants are likely to be maintained by balancing selection. If frequencies of synonymous and nonsynonymous variants are low, this is indicative of recurrent mutation. If frequencies of synonymous and nonsynonymous variants are high, this is indicative of co-transmission (Park, et al., 2015).

[0486] Like Ebola virus, Lassa virus (LASV) can cause hemorrhagic fever with high case fatality rates. Andersen et al. generated a genomic catalog of almost 200 LASV sequences from clinical and rodent reservoir samples (Andersen, et al., Cell Volume 162, Issue 4, p 738–750, 13 August 2015). Andersen et al. show that whereas the 2013–2015 EVD epidemic is fueled by human-to-human transmissions, LASV infections mainly result from reservoir-to-human infections. Andersen et al. elucidated the spread of LASV across West Africa and show that this migration was accompanied by changes in LASV genome abundance, fatality rates, codon adaptation, and translational efficiency. The method may further comprise phylogenetically comparing a first pathogen sequence to a second pathogen sequence, and determining whether

there is a phylogenetic link between the first and second pathogen sequences. The second pathogen sequence may be an earlier reference sequence. If there is a phylogenetic link, the method may further comprise rooting the phylogeny of the first pathogen sequence to the second pathogen sequence. Thus, it is possible to construct the lineage of the first pathogen sequence. (Park, et al., 2015).

[0487] The method may further comprise determining whether the mutations are deleterious or adaptive. Deleterious mutations are indicative of transmission-impaired viruses and dead-end infections, thus normally only present in an individual subject. Mutations unique to one individual subject are those that occur on the external branches of the phylogenetic tree, whereas internal branch mutations are those present in multiple samples (i.e., in multiple subjects). Higher rate of nonsynonymous substitution is a characteristic of external branches of the phylogenetic tree (Park, et al., 2015).

[0488] In internal branches of the phylogenetic tree, selection has had more opportunity to filter out deleterious mutants. Internal branches, by definition, have produced multiple descendent lineages and are thus less likely to include mutations with fitness costs. Thus, lower rate of nonsynonymous substitution is indicative of internal branches (Park, et al., 2015).

[0489] Synonymous mutations, which likely have less impact on fitness, occurred at more comparable frequencies on internal and external branches (Park, et al., 2015).

[0490] By analyzing the sequenced target sequence, such as viral genomes, it is possible to discover the mechanisms responsible for the severity of the epidemic episode such as during the 2014 Ebola outbreak. For example, Gire et al. made a phylogenetic comparison of the genomes of the 2014 outbreak to all 20 genomes from earlier outbreaks suggests that the 2014 West African virus likely spread from central Africa within the past decade. Rooting the phylogeny using divergence from other ebolavirus genomes was problematic (6, 13). However, rooting the tree on the oldest outbreak revealed a strong correlation between sample date and root-to-tip distance, with a substitution rate of $8 \times 10{-}4$ per site per year (13). This suggests that the lineages of the three most recent outbreaks all diverged from a common ancestor at roughly the same time, around 2004, which supports the hypothesis that each outbreak represents an independent zoonotic event from the same genetically diverse viral population in its natural reservoir. They also found out that the 2014 EBOV outbreak might be caused by a single transmission from the natural reservoir, followed by human-to-human transmission during the outbreak. Their results also suggested that the epidemic episode in Sierra Leon might

stem from the introduction of two genetically distinct viruses from Guinea around the same time (Gire, et al., 2014).

[0491] It has been also possible to determine how the Lassa virus spread out from its origin point, in particular thanks to human-to-human transmission and even retrace the history of this spread 400 years back (Andersen, et al., *Cell* 162(4):738–50, 2015).

[0492] In relation to the work needed during the 2013–2015 EBOV outbreak and the difficulties encountered by the medical staff at the site of the outbreak, and more generally, the method of the invention makes it possible to carry out sequencing using fewer selected probes such that sequencing can be accelerated, thus shortening the time needed from sample taking to results procurement. Further, kits and systems can be designed to be usable on the field so that diagnostics of a patient can be readily performed without need to send or ship samples to another part of the country or the world.

[0493] In any method described above, sequencing the target sequence or fragment thereof may used any of the sequencing processes described above. Further, sequencing the target sequence or fragment thereof may be a near-real-time sequencing. Sequencing the target sequence or fragment thereof may be carried out according to previously described methods (Experimental Procedures: Matranga et al., 2014; and Gire, et al., 2014). Sequencing the target sequence or fragment thereof may comprise parallel sequencing of a plurality of target sequences. Sequencing the target sequence or fragment thereof may comprise Illumina sequencing.

[0494] Analyzing the target sequence or fragment thereof that hybridizes to one or more of the selected probes may be an identifying analysis, wherein hybridization of a selected probe to the target sequence or a fragment thereof indicates the presence of the target sequence within the sample.

[0495] Currently, primary diagnostics are based on the symptoms a patient has. However, various diseases may share identical symptoms so that diagnostics rely much on statistics. For example, malaria triggers flu-like symptoms: headache, fever, shivering, joint pain, vomiting, hemolytic anemia, jaundice, hemoglobin in the urine, retinal damage, and convulsions. These symptoms are also common for septicemia, gastroenteritis, and viral diseases. Amongst the latter, Ebola hemorrhagic fever has the following symptoms fever, sore throat, muscular pain, headaches, vomiting, diarrhea, rash, decreased function of the liver and kidneys, internal and external hemorrhage.

[0496] When a patient is presented to a medical unit, for example in tropical Africa, basic diagnostics will conclude to malaria because statistically, malaria is the most probable disease within that region of Africa. The patient is consequently treated for malaria although the patient might not actually have contracted the disease and the patient ends up not being correctly treated. This lack of correct treatment can be life-threatening especially when the disease the patient contracted presents a rapid evolution. It might be too late before the medical staff realizes that the treatment given to the patient is ineffective and comes to the correct diagnostics and administers the adequate treatment to the patient.

[0497] The method of the invention provides a solution to this situation. Indeed, because the number of guide RNAs can be dramatically reduced, this makes it possible to provide on a single chip selected probes divided into groups, each group being specific to one disease, such that a plurality of diseases, e.g., viral infection, can be diagnosed at the same time. Thanks to the invention, more than 3 diseases can be diagnosed on a single chip, preferably more than 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 diseases at the same time, preferably the diseases that most commonly occur within the population of a given geographical area. Since each group of selected probes is specific to one of the diagnosed diseases, a more accurate diagnostics can be performed, thus diminishing the risk of administering the wrong treatment to the patient.

[0498] In other cases, a disease such as a viral infection may occur without any symptoms, or had caused symptoms but they faded out before the patient is presented to the medical staff. In such cases, either the patient does not seek any medical assistance or the diagnostics is complicated due to the absence of symptoms on the day of the presentation.

[0499] The present invention may also be used in concert with other methods of diagnosing disease, identifying pathogens and optimizing treatment based upon detection of nucleic acids, such as mRNA in crude, non-purified samples.

[0500] The method of the invention also provides a powerful tool to address this situation. Indeed, since a plurality of groups of selected guide RNAs, each group being specific to one of the most common diseases that occur within the population of the given area, are comprised within a single diagnostic, the medical staff only need to contact a biological sample taken from the patient with the chip. Reading the chip reveals the diseases the patient has contracted.

[0501] In some cases, the patient is presented to the medical staff for diagnostics of particular symptoms. The method of the invention makes it possible not only to identify which

disease causes these symptoms but at the same time determine whether the patient suffers from another disease he was not aware of.

[0502] This information might be of utmost importance when searching for the mechanisms of an outbreak. Indeed, groups of patients with identical viruses also show temporal patterns suggesting a subject-to-subject transmission links.

[0503] In some embodiments, a programmable pattern recognition protein composition or method of use thereof as described herein may be used to predict disease outcome in patients suffering from viral diseases. In specific embodiments, such viral diseases may include, but are not necessarily limited to, Lassa fever. Specific factors related to Lassa fever disease outcome may include but are not necessarily limited to, age, extent of kidney injury, and/or CNS injury.

Screening Microbial Genetic Perturbations

[0504] In certain example embodiments, the programmable pattern recognition compositions and systems of the present invention disclosed herein may be used to screen microbial genetic perturbations. Such methods may be useful, for example to map out microbial pathways and functional networks. Microbial cells may be genetically modified and then screened under different experimental conditions. As described above, the embodiments disclosed herein can screen for multiple target molecules in a single sample, or a single target in a single individual discrete volume in a multiplex fashion. Genetically modified microbes may be modified to include a nucleic acid barcode sequence that identifies the particular genetic modification carried by a particular microbial cell or population of microbial cells. A barcode is s short sequence of nucleotides (for example, DNA, RNA, or combinations thereof) that is used as an identifier. A nucleic acid barcode may have a length of 4-100 nucleotides and be either single or double-stranded. Methods for identifying cells with barcodes are known in the art. Accordingly, guide RNAs of the effector compositions and systems of the present invention described herein may be used to detect the barcode. Detection of the positive detectable signal indicates the presence of a particular genetic modification in the sample. The methods disclosed herein may be combined with other methods for detecting complimentary genotype or phenotypic readouts indicating the effect of the genetic modification under the experimental conditions tested. Genetic modifications to be screened may include, but are not limited to, a gene knock-in, a gene knock-out, inversions, translocations, transpositions, or one or more nucleotide insertions, deletions, substitutions, mutations, or addition of nucleic acids encoding an epitope with a functional consequence such as altering protein stability or

detection. In a similar fashion, the methods described herein may be used in synthetic biology application to screen the functionality of specific arrangements of gene regulatory elements and gene expression modules.

[0505] In certain example embodiments, the methods may be used to screen hypomorphs. Generation of hypomorphs and their use in identifying key bacterial functional genes and identification of new antibiotic therapeutics as disclosed in PCT/US2016/060730 entitled "Multiplex High-Resolution Detection of Micro-organism Strains, Related Kits, Diagnostic Methods and Screening Assays" filed November 4, 2016, which is incorporated herein by reference.

[0506] The different experimental conditions may comprise exposure of the microbial cells to different chemical agents, combinations of chemical agents, different concentrations of chemical agents or combinations of chemical agents, different durations of exposure to chemical agents or combinations of chemical agents, different physical parameters, or both. In certain example embodiments the chemical agent is an antibiotic or antiviral. Different physical parameters to be screened may include different temperatures, atmospheric pressures, different atmospheric and non-atmospheric gas concentrations, different pH levels, different culture media compositions, or a combination thereof.

Screening Environmental Samples

[0507] The methods disclosed herein may also be used to screen environmental samples for contaminants by detecting the presence of target nucleic acids. For example, in some embodiments, the invention provides a method of detecting microbes, comprising: exposing a detection composition (e.g., a programmable pattern recognition composition configured to detect one or more target cells or molecules) of the present invention as described herein to a sample; activating the programmable pattern recognition composition and/or system and/or an effector component thereof, by binding a PAMP or other recognized pattern associated with a target cell or molecule so as to modify a detection construct to produce a detectable signal. In some embodiments the programmable pattern recognition composition and/or system and/or an effector component thereof includes an RNA effector protein that is activated via binding of one or more guide RNAs to one or more microbe-specific target RNAs or one or more trigger RNAs such that a detectable positive signal is produced. The positive signal can be detected and is indicative of the presence of one or more microbes in the sample. In some embodiments, the detection composition or system of the present invention or component thereof may be on

a substrate as described herein, and the substrate may be exposed to the sample. In other embodiments, the same detection composition or system of the present invention, and/or a different detection composition or system of the present invention may be applied to multiple discrete locations on the substrate. In further embodiments, the different detection composition or system of the present invention may detect a different microbe at each location. As described in further detail above, a substrate may be a flexible materials substrate, for example, including, but not limited to, a paper substrate, a fabric substrate, or a flexible polymer-based substrate.

[0508] In accordance with the invention, the substrate may be exposed to the sample passively, by temporarily immersing the substrate in a fluid to be sampled, by applying a fluid to be tested to the substrate, or by contacting a surface to be tested with the substrate. Any means of introducing the sample to the substrate may be used as appropriate.

[0509] As described herein, a sample for use with the invention may be a biological or environmental sample, such as a food sample (fresh fruits or vegetables, meats), a beverage sample, a paper surface, a fabric surface, a metal surface, a wood surface, a plastic surface, a soil sample, a freshwater sample, a wastewater sample, a saline water sample, exposure to atmospheric air or other gas sample, or a combination thereof. For example, household/commercial/industrial surfaces made of any materials including, but not limited to, metal, wood, plastic, rubber, or the like, may be swabbed and tested for contaminants. Soil samples may be tested for the presence of pathogenic bacteria or parasites, or other microbes, both for environmental purposes and/or for human, animal, or plant disease testing. Water samples such as freshwater samples, wastewater samples, or saline water samples can be evaluated for cleanliness and safety, and/or potability, to detect the presence of, for example, *Cryptosporidium parvum*, *Giardia lamblia*, or other microbial contamination. In further embodiments, a biological sample may be obtained from a source including, but not limited to, a tissue sample, saliva, blood, plasma, sera, stool, urine, sputum, mucous, lymph, synovial fluid, cerebrospinal fluid, ascites, pleural effusion, seroma, pus, or swab of skin or a mucosal membrane surface. In some particular embodiments, an environmental sample or biological samples may be crude samples and/or the one or more target molecules may not be purified or amplified from the sample prior to application of the method. Identification of microbes may be useful and/or needed for any number of applications, and thus any type of sample from any source deemed appropriate by one of skill in the art may be used in accordance with the invention.

[0510] In some embodiments, checking for food contamination by bacteria, such as *E. coli*, in restaurants or other food providers; food surfaces; Testing water for pathogens like *Salmonella*, *Campylobacter*, or *E. coli*; also checking food quality for manufacturers and regulators to determine the purity of meat sources; identifying air contamination with pathogens such as legionella; Checking whether beer is contaminated or spoiled by pathogens like Pediococcus and Lactobacillus; contamination of pasteurized or un-pasteurized cheese by bacteria or fungi during manufacture.

[0511] A microbe in accordance with the invention may be a pathogenic microbe or a microbe that results in food or consumable product spoilage. A pathogenic microbe may be pathogenic or otherwise undesirable to humans, animals, or plants. For human or animal purposes, a microbe may cause a disease or result in illness. Animal or veterinary applications of the present invention may identify animals infected with a microbe. For example, the methods and systems of the invention may identify companion animals with pathogens including, but not limited to, kennel cough, rabies virus, and heartworms. In other embodiments, the methods and systems of the invention may be used for parentage testing for breeding purposes. A plant microbe may result in harm or disease to a plant, reduction in yield, or alter traits such as color, taste, consistency, odor, For food or consumable contamination purposes, a microbe may adversely affect the taste, odor, color, consistency or other commercial properties of the food or consumable product. In certain example embodiments, the microbe is a bacterial species. The bacteria may be a psychrotroph, a coliform, a lactic acid bacteria, or a spore-forming bacteria. In certain example embodiments, the bacteria may be any bacterial species that causes disease or illness, or otherwise results in an unwanted product or trait. Bacteria in accordance with the invention may be pathogenic to humans, animals, or plants.

Example Microbes

[0512] The embodiment disclosed herein may be used to detect a number of different microbes. The term microbe as used herein includes bacteria, fungus, protozoa, parasites and viruses.

Bacteria

[0513] The following provides an example list of the types of microbes that might be detected using the embodiments disclosed herein. In certain example embodiments, the microbe is a bacterium. Examples of bacteria that can be detected in accordance with the

disclosed methods include without limitation any one or more of (or any combination of) *Acinetobacter baumanii*, *Actinobacillus sp.*, *Actinomycetes*, *Actinomyces sp.* (such as *Actinomyces israelii* and *Actinomyces naeslundii*), *Aeromonas sp.* (such as *Aeromonas hydrophila*, *Aeromonas veronii biovar sobria* (*Aeromonas sobria*), and *Aeromonas caviae*), *Anaplasma phagocytophilum*, *Anaplasma marginale Alcaligenes xylosoxidans*, *Acinetobacter baumanii*, *Actinobacillus actinomycetemcomitans*, *Bacillus sp.* (such as *Bacillus anthracis*, *Bacillus cereus*, *Bacillus subtilis*, *Bacillus thuringiensis*, and *Bacillus stearothermophilus*), *Bacteroides sp.* (such as *Bacteroides fragilis*), *Bartonella sp.* (such as *Bartonella bacilliformis* and *Bartonella henselae*, *Bifidobacterium sp.*, *Bordetella sp.* (such as *Bordetella pertussis*, *Bordetella parapertussis*, and *Bordetella bronchiseptica*), *Borrelia sp.* (such as *Borrelia recurrentis*, and *Borrelia burgdorferi*), *Brucella sp.* (such as *Brucella abortus*, *Brucella canis*, *Brucella melintensis* and *Brucella suis*), *Burkholderia sp.* (such as *Burkholderia pseudomallei* and *Burkholderia cepacia*), *Campylobacter sp.* (such as *Campylobacter jejuni*, *Campylobacter coli*, *Campylobacter lari* and *Campylobacter fetus*), *Capnocytophaga sp.*, *Cardiobacterium hominis*, *Chlamydia trachomatis*, *Chlamydophila pneumoniae*, *Chlamydophila psittaci*, *Citrobacter sp. Coxiella burnetii*, *Corynebacterium sp.* (such as, *Corynebacterium diphtheriae*, *Corynebacterium jeikeum* and *Corynebacterium*), *Clostridium sp.* (such as *Clostridium perfringens*, *Clostridium difficile*, *Clostridium botulinum* and *Clostridium tetani*), *Eikenella corrodens*, *Enterobacter sp.* (such as *Enterobacter aerogenes*, *Enterobacter agglomerans*, *Enterobacter cloacae* and *Escherichia coli*, *including opportunistic Escherichia coli*, such as *enterotoxigenic E. coli*, *enteroinvasive E. coli*, *enteropathogenic E. coli*, *enterohemorrhagic E. coli*, *enteroaggregative E. coli* and *uropathogenic E. coli*) *Enterococcus sp.* (such as *Enterococcus faecalis* and *Enterococcus faecium*) *Ehrlichia sp.* (such as *Ehrlichia chafeensia* and *Ehrlichia canis*), *Epidermophyton floccosum, Erysipelothrix rhusiopathiae*, *Eubacterium sp.*, *Francisella tularensis*, *Fusobacterium nucleatum*, *Gardnerella vaginalis*, *Gemella morbillorum*, *Haemophilus sp.* (such as *Haemophilus influenzae*, *Haemophilus ducreyi*, *Haemophilus aegyptius*, *Haemophilus parainfluenzae*, *Haemophilus haemolyticus* and *Haemophilus parahaemolyticus*, *Helicobacter sp.* (such as *Helicobacter pylori*, *Helicobacter cinaedi* and *Helicobacter fennelliae*), *Kingella kingii*, *Klebsiella sp.* (such as *Klebsiella pneumoniae*, *Klebsiella granulomatis* and *Klebsiella oxytoca*), *Lactobacillus sp.*, *Listeria monocytogenes*, *Leptospira interrogans*, *Legionella pneumophila*, *Leptospira interrogans*, *Peptostreptococcus sp.*, *Mannheimia hemolytica, Microsporum canis, Moraxella catarrhalis*,

Morganella sp., *Mobiluncus sp.*, *Micrococcus sp.*, *Mycobacterium sp.* (such as *Mycobacterium leprae*, *Mycobacterium tuberculosis*, *Mycobacterium paratuberculosis*, *Mycobacterium intracellulare*, *Mycobacterium avium*, *Mycobacterium bovis*, and *Mycobacterium marinum*), *Mycoplasm sp.* (such as *Mycoplasma pneumoniae*, *Mycoplasma hominis*, and *Mycoplasma genitalium*), *Nocardia sp.* (such as *Nocardia asteroides*, *Nocardia cyriacigeorgica* and *Nocardia brasiliensis*), *Neisseria sp.* (such as *Neisseria gonorrhoeae* and *Neisseria meningitidis*), *Pasteurella multocida*, *Pityrosporum orbiculare* (Malassezia furfur), *Plesiomonas shigelloides. Prevotella sp.*, *Porphyromonas sp.*, *Prevotella melaninogenica*, *Proteus sp.* (such as *Proteus vulgaris* and *Proteus mirabilis*), *Providencia sp.* (such as *Providencia alcalifaciens*, *Providencia rettgeri* and *Providencia stuartii*), *Pseudomonas aeruginosa*, *Propionibacterium acnes*, *Rhodococcus equi*, *Rickettsia sp.* (such as *Rickettsia rickettsii*, *Rickettsia akari* and *Rickettsia prowazekii*, *Orientia tsutsugamushi* (*formerly: Rickettsia tsutsugamushi*) and *Rickettsia typhi*), *Rhodococcus sp.*, *Serratia marcescens*, *Stenotrophomonas maltophilia*, *Salmonella sp.* (such as *Salmonella enterica*, *Salmonella typhi*, *Salmonella paratyphi*, *Salmonella enteritidis*, *Salmonella cholerasuis* and *Salmonella typhimurium*), *Serratia sp.* (such as *Serratia marcesans* and *Serratia liquifaciens*), *Shigella sp.* (such as *Shigella dysenteriae*, *Shigella flexneri*, *Shigella boydii* and *Shigella sonnei*), *Staphylococcus sp.* (such as *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Staphylococcus hemolyticus*, *Staphylococcus saprophyticus*), *Streptococcus sp.* (such as *Streptococcus pneumoniae* (for example *chloramphenicol-resistant serotype 4 Streptococcus pneumoniae*, *spectinomycin-resistant serotype 6B Streptococcus pneumoniae*, *streptomycin-resistant serotype 9V Streptococcus pneumoniae*, *erythromycin-resistant serotype 14 Streptococcus pneumoniae*, *optochin-resistant serotype 14 Streptococcus pneumoniae*, *rifampicin-resistant serotype 18C Streptococcus pneumoniae*, *tetracycline-resistant serotype 19F Streptococcus pneumoniae*, *penicillin-resistant serotype 19F Streptococcus pneumoniae*, and *trimethoprim-resistant serotype 23F Streptococcus pneumoniae*, *chloramphenicol-resistant serotype 4 Streptococcus pneumoniae*, *spectinomycin-resistant serotype 6B Streptococcus pneumoniae*, *streptomycin-resistant serotype 9V Streptococcus pneumoniae*, *optochin-resistant serotype 14 Streptococcus pneumoniae*, *rifampicin-resistant serotype 18C Streptococcus pneumoniae*, *penicillin-resistant serotype 19F Streptococcus pneumoniae*, *or trimethoprim-resistant serotype 23F Streptococcus pneumoniae*), *Streptococcus agalactiae*, *Streptococcus mutans*, *Streptococcus pyogenes*, *Group A streptococci*, *Streptococcus*

pyogenes, *Group B streptococci*, *Streptococcus agalactiae*, *Group C streptococci*, *Streptococcus anginosus*, *Streptococcus equismilis*, *Group D streptococci*, *Streptococcus bovis*, *Group F streptococci*, and *Streptococcus anginosus Group G streptococci*), *Spirillum minus*, *Streptobacillus moniliformi*, *Treponema sp.* (such as *Treponema carateum*, *Treponema petenue*, *Treponema pallidum* and *Treponema endemicum*, *Trichophyton rubrum, T. mentagrophytes, Tropheryma whippelii*, *Ureaplasma urealyticum*, *Veillonella sp.*, *Vibrio sp.* (such as *Vibrio cholerae*, *Vibrio parahemolyticus*, *Vibrio vulnificus*, *Vibrio parahaemolyticus*, *Vibrio vulnificus*, *Vibrio alginolyticus*, *Vibrio mimicus*, *Vibrio hollisae*, *Vibrio fluvialis*, *Vibrio metchnikovii*, *Vibrio damsela* and *Vibrio furnisii*), *Yersinia sp.* (such as *Yersinia enterocolitica*, *Yersinia pestis*, and *Yersinia pseudotuberculosis*) and *Xanthomonas maltophilia* among others.

[0514] Near-real-time microbial diagnostics are needed for food, clinical, industrial, and other environmental settings (see e.g., Lu TK, Bowers J, and Koeris MS., Trends Biotechnol. 2013 Jun;31(6):325-7). In certain embodiments, the assay described herein is configured for detection of foodborne pathogens using guide RNAs specific to a pathogen (e.g., *Campylobacter jejuni, Clostridium perfringens, Salmonella spp., Escherichia coli, Bacillus cereus, Listeria monocytogenes, Shigella spp., Staphylococcus aureus, Staphylococcal enteritis, Streptococcus, Vibrio cholerae, Vibrio parahaemolyticus, Vibrio vulnificus, Yersinia enterocolitica and Yersinia pseudotuberculosis, Brucella spp., Corynebacterium ulcerans, Coxiella burnetii,* or *Plesiomonas shigelloides*).

Fungi

[0515] In certain example embodiments, the microbe is a fungus or a fungal species. Examples of fungi that can be detected in accordance with the disclosed methods include without limitation any one or more of (or any combination of), *Aspergillus*, *Blastomyces*, *Candidiasis*, *Coccidiodomycosis*, *Cryptococcus neoformans*, *Cryptococcus gatti*, *sp. Histoplasma sp.* (such as *Histoplasma capsulatum*), *Pneumocystis sp.* (such as *Pneumocystis jirovecii*), *Stachybotrys* (such as *Stachybotrys chartarum*), *Mucroymcosis*, Sporothrix, fungal eye infections ringworm, *Exserohilum*, *Cladosporium*.

[0516] In certain example embodiments, the fungus is a yeast. Examples of yeast that can be detected in accordance with disclosed methods include without limitation one or more of (or any combination of), *Aspergillus* species (such as *Aspergillus fumigatus*, *Aspergillus flavus* and *Aspergillus clavatus*), *Cryptococcus sp.* (such as *Cryptococcus neoformans, Cryptococcus gattii, Cryptococcus laurentii* and *Cryptococcus albidus*), a *Geotrichum* species, a

Saccharomyces species, a *Hansenula* species, a *Candida* species (such as *Candida albicans*), a *Kluyveromyces* species, a *Debaryomyces* species, a *Pichia* species, or combination thereof. In certain example embodiments, the fungi is a mold. Example molds include, but are not limited to, a *Penicillium* species, a *Cladosporium* species, a *Byssochlamys* species, or a combination thereof.

Protozoa

[0517] In certain example embodiments, the microbe is a protozoan. Examples of protozoa that can be detected in accordance with the disclosed methods and devices include without limitation any one or more of (or any combination of), Euglenozoa, Heterolobosea, Diplomonadida, Amoebozoa, Blastocystic, and Apicomplexa. Example Euglenoza include, but are not limited to, *Trypanosoma cruzi* (Chagas disease), *T. brucei gambiense, T. brucei rhodesiense, Leishmania braziliensis, L. infantum, L. mexicana, L. major, L. tropica,* and *L. donovani.* Example Heterolobosea include, but are not limited to, *Naegleria fowleri.* Example Diplomonadid include, but are not limited to, *Giardia intestinalis* (*G. lamblia, G. duodenalis*). Example Amoebozoa include, but are not limited to, *Acanthamoeba castellanii, Balamuthia madrillaris, Entamoeba histolytica.* Example Blastocystis include, but are not limited to, *Blastocystic hominis.* Example Apicomplexa include, but are not limited to, *Babesia microti, Cryptosporidium parvum, Cyclospora cayetanensis, Plasmodium falciparum, P. vivax, P. ovale, P. malariae,* and *Toxoplasma gondii.Babesia microti, Cryptosporidium parvum, Cyclospora cayetanensis, Plasmodium falciparum, P. vivax, P. ovale, P. malariae,* and *Toxoplasma gondii*.

Parasites

[0518] In certain example embodiments, the microbe is a parasite. Examples of parasites that can be detected in accordance with disclosed methods include without limitation one or more of (or any combination of), an *Onchocerca* species and a *Plasmodium* species.

Viruses

[0519] In certain example embodiments, the systems, devices, and methods, disclosed herein are directed to detecting viruses in a sample. The embodiments disclosed herein may be used to detect viral infection (*e.g.*, of a subject or plant), or determination of a viral strain, including viral strains that differ by a single nucleotide polymorphism. The virus may be a DNA virus, a RNA virus, or a retrovirus. Non-limiting example of viruses useful with the present invention include, but are not limited to Ebola, measles, SARS, Chikungunya, hepatitis,

Marburg, yellow fever, MERS, Dengue, Lassa, influenza, rhabdovirus or HIV. A hepatitis virus may include hepatitis A, hepatitis B, or hepatitis C. An influenza virus may include, for example, influenza A or influenza B. An HIV may include HIV 1 or HIV 2. In certain example embodiments, the viral sequence may be a human respiratory syncytial virus, Sudan ebola virus, Bundibugyo virus, Tai Forest ebola virus, Reston ebola virus, Achimota, Aedes flavivirus, Aguacate virus, Akabane virus, Alethinophid reptarenavirus, Allpahuayo mammarenavirus, Amapari mmarenavirus, Andes virus, Apoi virus, Aravan virus, Aroa virus, Arumwot virus, Atlantic salmon paramyoxivirus, Australian bat lyssavirus, Avian bornavirus, Avian metapneumovirus, Avian paramyoxviruses, penguin or Falkland Islandsvirus, BK polyomavirus, Bagaza virus, Banna virus, Bat hepevirus, Bat sapovirus, Bear Canon mammarenavirus, Beilong virus, Betacoronoavirus, Betapapillomavirus 1-6, Bhanja virus, Bokeloh bat lyssavirus, Borna disease virus, Bourbon virus, Bovine hepacivirus, Bovine parainfluenza virus 3, Bovine respiratory syncytial virus, Brazoran virus, Bunyamwere virus, Caliciviridae virus. California encephalitis virus, Candiru virus, Canine distemper virus, Canaine pneumovirus, Cedar virus, Cell fusing agent virus, Cetacean morbillivirus, Chandipura virus, Chaoyang virus, Chapare mammarenavirus, Chikungunya virus, Colobus monkey papillomavirus, Colorado tick fever virus, Cowpox virus, Crimean-Congo hemorrhagic fever virus, Culex flavivirus, Cupixi mammarenavirus, Dengue virus, Dobrava-Belgrade virus, Donggang virus, Dugbe virus, Duvenhage virus, Eastern equine encephalitis virus, Entebbe bat virus, Enterovirus A-D, European bat lyssavirus 1-2, Eyach virus, Feline morbillivirus, Fer-de-Lance paramyxovirus, Fitzroy River virus, Flaviviridae virus, Flexal mammarenavirus, GB virus C, Gairo virus, Gemycircularvirus, Goose paramyoxiviurs SF02, Great Island virus, Guanarito mammarenavirus, Hantaan virus, Hantavirus Z10, Heartland virus, Hendra virus, Hepatitis A/B/C/E, Hepatitis delta virus, Human bocavirus, Human coronavirus, Human endogenous retrovirus K, Human enteric coronavirus, Human gential-associated circular DNA virus-1, Human herpesvirus 1-8, Human immunodeficiency virus 1/2, Huan mastadenovirus A-G, Human papillomavirus, Human parainfluenza virus 1-4, Human paraechovirus, Human picobirnavirus, Human smacovirus, Ikoma lyssavirus, Ilheus virus, Influenza A-C, Ippy mammarenavirus, Irkut virus, J-virus, JC polyomavirus, Japanses encephalitis virus, Junin mammarenavirus, KI polyomavirus, Kadipiro virus, Kamiti River virus, Kedougou virus, Khujand virus, Kokobera virus, Kyasanur forest disease virus, Lagos bat virus, Langat virus, Lassa mammarenavirus, Latino mammarenavirus, Leopards Hill virus,

Liao ning virus, Ljungan virus, Lloviu virus, Louping ill virus, Lujo mammarenavirus, Luna mammarenavirus, Lunk virus, Lymphocytic choriomeningitis mammarenavirus, Lyssavirus Ozernoe, MSSI2\.225 virus, Machupo mammarenavirus, Mamastrovirus 1, Manzanilla virus, Mapuera virus, Marburg virus, Mayaro virus, Measles virus, Menangle virus, Mercadeo virus, Merkel cell polyomavirus, Middle East respiratory syndrome coronavirus, Mobala mammarenavirus, Modoc virus, Moijang virus, Mokolo virus, Monkeypox virus, Montana myotis leukoenchalitis virus, Mopeia lassa virus reassortant 29, Mopeia mammarenavirus, Morogoro virus, Mossman virus, Mumps virus, Murine pneumonia virus, Murray Valley encephalitis virus, Nariva virus, Newcastle disease virus, Nipah virus, Norwalk virus, Norway rat hepacivirus, Ntaya virus, O'nyong-nyong virus, Oliveros mammarenavirus, Omsk hemorrhagic fever virus, Oropouche virus, Parainfluenza virus 5, Parana mammarenavirus, Parramatta River virus, Peste-des-petits-ruminants virus, Pichande mammarenavirus, Picornaviridae virus, Pirital mammarenavirus, Piscihepevirus A, Procine parainfluenza virus 1, porcine rubulavirus, Powassan virus, Primate T-lymphotropic virus 1-2, Primate erythroparvovirus 1, Punta Toro virus, Puumala virus, Quang Binh virus, Rabies virus, Razdan virus, Reptile bornavirus 1, Rhinovirus A-B, Rift Valley fever virus, Rinderpest virus, Rio Bravo virus, Rodent Torque Teno virus, Rodent hepacivirus, Ross River virus, Rotavirus A-I, Royal Farm virus, Rubella virus, Sabia mammarenavirus, Salem virus, Sandfly fever Naples virus, Sandfly fever Sicilian virus, Sapporo virus, Sathuperi virus, Seal anellovirus, Semliki Forest virus, Sendai virus, Seoul virus, Sepik virus, Severe acute respiratory syndrome-related coronavirus, Severe fever with thrombocytopenia syndrome virus, Shamonda virus, Shimoni bat virus, Shuni virus, Simbu virus, Simian torque teno virus, Simian virus 40-41, Sin Nombre virus, Sindbis virus, Small anellovirus, Sosuga virus, Spanish goat encephalitis virus, Spondweni virus, St. Louis encephalitis virus, Sunshine virus, TTV-like mini virus, Tacaribe mammarenavirus, Taila virus, Tamana bat virus, Tamiami mammarenavirus, Tembusu virus, Thogoto virus, Thottapalayam virus, Tick-borne encephalitis virus, Tioman virus, Togaviridae virus, Torque teno canis virus, Torque teno douroucouli virus, Torque teno felis virus, Torque teno midi virus, Torque teno sus virus, Torque teno tamarin virus, Torque teno virus, Torque teno zalophus virus, Tuhoko virus, Tula virus, Tupaia paramyxovirus, Usutu virus, Uukuniemi virus, Vaccinia virus, Variola virus, Venezuelan equine encephalitis virus, Vesicular stomatitis Indiana virus, WU Polyomavirus, Wesselsbron virus, West Caucasian bat virus, West Nile virus, Western equine encephalitis virus, Whitewater Arroyo mammarenavirus, Yellow fever

virus, Yokose virus, Yug Bogdanovac virus, Zaire ebolavirus, Zika virus, or Zygosaccharomyces bailii virus Z viral sequence. Examples of RNA viruses that may be detected include one or more of (or any combination of) Coronaviridae virus, a Picornaviridae virus, a Caliciviridae virus, a Flaviviridae virus, a Togaviridae virus, a Bornaviridae, a Filoviridae, a Paramyxoviridae, a Pneumoviridae, a Rhabdoviridae, an Arenaviridae, a Bunyaviridae, an Orthomyxoviridae, or a Deltavirus. In certain example embodiments, the virus is Coronavirus, SARS, Poliovirus, Rhinovirus, Hepatitis A, Norwalk virus, Yellow fever virus, West Nile virus, Hepatitis C virus, Dengue fever virus, Zika virus, Rubella virus, Ross River virus, Sindbis virus, Chikungunya virus, Borna disease virus, Ebola virus, Marburg virus, Measles virus, Mumps virus, Nipah virus, Hendra virus, Newcastle disease virus, Human respiratory syncytial virus, Rabies virus, Lassa virus, Hantavirus, Crimean-Congo hemorrhagic fever virus, Influenza, or Hepatitis D virus.

[0520] In certain example embodiments, the virus may be a plant virus selected from the group comprising Tobacco mosaic virus (TMV), Tomato spotted wilt virus (TSWV), Cucumber mosaic virus (CMV), Potato virus Y (PVY), the RT virus Cauliflower mosaic virus (CaMV), Plum pox virus (PPV), Brome mosaic virus (BMV), Potato virus X (PVX), Citrus tristeza virus (CTV), Barley yellow dwarf virus (BYDV), Potato leafroll virus (PLRV), Tomato bushy stunt virus (TBSV), rice tungro spherical virus (RTSV), rice yellow mottle virus (RYMV), rice hoja blanca virus (RHBV), maize rayado fino virus (MRFV), maize dwarf mosaic virus (MDMV), sugarcane mosaic virus (SCMV), Sweet potato feathery mottle virus (SPFMV), sweet potato sunken vein closterovirus (SPSVV), Grapevine fanleaf virus (GFLV), Grapevine virus A (GVA), Grapevine virus B (GVB), Grapevine fleck virus (GFkV), Grapevine leafroll-associated virus-1, -2, and -3, (GLRaV-1, -2, and -3), Arabis mosaic virus (ArMV), or Rupestris stem pitting-associated virus (RSPaV). In a preferred embodiment, the target RNA molecule is part of said pathogen or transcribed from a DNA molecule of said pathogen. For example, the target sequence may be comprised in the genome of an RNA virus. It is further preferred that CRISPR effector protein hydrolyzes said target RNA molecule of said pathogen in said plant if said pathogen infects or has infected said plant. It is thus preferred that the CRISPR system is capable of cleaving the target RNA molecule from the plant pathogen both when the CRISPR system (or parts needed for its completion) is applied therapeutically, *i.e,*. after infection has occurred or prophylactically, *i.e.*, before infection has occurred.

[0521] In certain example embodiments, the virus may be a retrovirus. Example retroviruses that may be detected using the embodiments disclosed herein include one or more of or any combination of viruses of the Genus Alpharetrovirus, Betaretrovirus, Gammaretrovirus, Deltaretrovirus, Epsilonretrovirus, Lentivirus, Spumavirus, or the Family Metaviridae, Pseudoviridae, and Retroviridae (including HIV), Hepadnaviridae (including Hepatitis B virus), and Caulimoviridae (including Cauliflower mosaic virus).

[0522] In certain example embodiments, the virus is a DNA virus. Example DNA viruses that may be detected using the embodiments disclosed herein include one or more of (or any combination of) viruses from the Family Myoviridae, Podoviridae, Siphoviridae, Alloherpesviridae, Herpesviridae (including human herpes virus, and Varicella Zozter virus), Malocoherpesviridae, Lipothrixviridae, Rudiviridae, Adenoviridae, Ampullaviridae, Ascoviridae, Asfarviridae (including African swine fever virus), Baculoviridae, Cicaudaviridae, Clavaviridae, Corticoviridae, Fuselloviridae, Globuloviridae, Guttaviridae, Hytrosaviridae, Iridoviridae, Maseilleviridae, Mimiviridae, Nudiviridae, Nimaviridae, Pandoraviridae, Papillomaviridae, Phycodnaviridae, Plasmaviridae, Polydnaviruses, Polyomaviridae (including Simian virus 40, JC virus, BK virus), Poxviridae (including Cowpox and smallpox), Sphaerolipoviridae, Tectiviridae, Turriviridae, Dinodnavirus, Salterprovirus, Rhizidovirus, among oIn some embodiments, a method of diagnosing a species-specific bacterial infection in a subject suspected of having a bacterial infection is described as obtaining a sample comprising bacterial ribosomal ribonucleic acid from the subject; contacting the sample with one or more of the probes described, and detecting hybridization between the bacterial ribosomal ribonucleic acid sequence present in the sample and the probe, wherein the detection of hybridization indicates that the subject is infected with *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Staphylococcus aureus, Acinetobacter baumannii, Candida albicans, Enterobacter cloacae, Enterococcus faecalis, Enterococcus faecium, Proteus mirabilis, Staphylococcus agalactiae,* or *Staphylococcus maltophilia* or a combination thereof.

SARS-CoV-2

[0523] The present disclosure relates to and/or involves detection of SARS-CoV-2.

[0524] As used herein, the term “variant” refers to any virus having one or more mutations as compared to a known virus. A strain is a genetic variant or subtype of a virus. The terms ‘strain’, ‘variant’, and ‘isolate’ may be used interchangeably. In certain embodiments, a variant

has developed a "specific group of mutations" that causes the variant to behave differently than that of the strain it originated from. While there are many thousands of variants of SARS-CoV-2, (Koyama, Takahiko Koyama; Platt, Daniela; Parida, Laxmi (June 2020). "Variant analysis of SARS-CoV-2 genomes". Bulletin of the World Health Organization. 98: 495–504) there are also much larger groupings called clades. Several different clade nomenclatures for SARS-CoV-2 have been proposed. As of December 2020, GISAID, referring to SARS-CoV-2 as hCoV-19 identified seven clades (O, S, L, V, G, GH, and GR) (Alm E, Broberg EK, Connor T, et al. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020 [published correction appears in Euro Surveill. 2020 Aug;25(33):]. Euro Surveill. 2020;25(32):2001410). Also as of December 2020, Nextstrain identified five (19A, 19B, 20A, 20B, and 20C) (Cited in Alm et al. 2020). Guan et al. identified five global clades (G614, S84, V251, I378 and D392) (Guan Q, Sadykov M, Mfarrej S, et al. A genetic barcode of SARS-CoV-2 for monitoring global distribution of different clades during the COVID-19 pandemic. Int J Infect Dis. 2020;100:216-223). Rambaut et al. proposed the term "lineage" in a 2020 article in Nature Microbiology; as of December 2020, there have been five major lineages (A, B, B.1, B.1.1, and B.1.777) identified (Rambaut, A.; Holmes, E.C.; O'Toole, Á.; et al. "A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology". 5: 1403–1407).

[0525] Genetic variants of SARS-CoV-2 have been emerging and circulating around the world throughout the COVID-19 pandemic (see, e.g., The US Centers for Disease Control and Prevention; www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html). Exemplary, non-limiting variants applicable to the present disclosure include variants of SARS-CoV-2, particularly those having substitutions of therapeutic concern. **Table 9** below shows exemplary, non-limiting genetic substitutions in SARS-CoV-2 variants.

Table 9.	
Spike Protein Substitution	**Common Pango Lineages with Spike Protein Substitutions**
L452R	A.2.5, B.1, B.1.429, B.1.427, B.1.617.1, B.1.526.1, B.1.617.2, C.36.3
E484K	B.1.1.318, B.1.1.7, B.1.351, B.1.525, B.1.526, B.1.621, B.1.623, P.1, P.1.1, P.1.2, R.1
K417N, E484K, N501Y	B.1.351, B.1.351.3
K417T, E484K, N501Y	P.1, P.1.1, P.1.2
A67V, del69-70, T95I, del142-144, Y145D, del211, L212I, ins214EPE, G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F	B.1.1.529 and BA lineages

Phylogenetic Assignment of Named Global Outbreak (PANGO) Lineages is software tool developed by members of the Rambaut Lab. The associated web application was developed by the Centre for Genomic Pathogen Surveillance in South Cambridgeshire and is intended to implement the dynamic nomenclature of SARS-CoV-2 lineages, known as the PANGO nomenclature. It is available at cov-lineages.org.

[0526] In some embodiments, the SARS-CoV-2 variant is and/or includes: B.1.1.7, also known as Alpha (WHO) or UK variant, having the following spike protein substitutions: 69del, 70del, 144del, (E484K*), (S494P*), N501Y, A570D, D614G, P681H, T716I, S982A, and D1118H (K1191N*); B.1.351, also known as Beta (WHO) or South Africa variant, having the following spike protein substitutions: D80A, D215G, 241del, 242del, 243del, K417N, E484K, N501Y, D614G, and A701V; B.1.427, also known as Epsilon (WHO) or US California variant, having the following spike protein substitutions: L452R, and D614G; B.1.429, also known as Epsilon (WHO) or US California variant, having the following spike protein substitutions: S13I, W152C, L452R, and D614G; B.1.617.2, also known as Delta (WHO) or India variant, having the following spike protein substitutions: T19R, (G142D), 156del, 157del, R158G, L452R, T478K, D614G, P681R, and D950N; P.1, also known as Gamma (WHO) or Japan/Brazil variant, having the following spike protein substitutions: L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, and T1027I; and B.1.1.529 also known as Omicron (WHO), having the following spike protein substitutions: A67V, del69-70, T95I, del142-144, Y145D, del211, L212I, ins214EPE, G339D, S371L, S373P, S375F, K417N,

N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K, D614G, H655Y, N679K, P681H, N764K, D796Y, N856K, Q954H, N969K, L981F, or any combination thereof.

[0527] In some embodiments, the SARS-CoV-2 variant is classified and/or otherwise identified as a Variant of Concern (VOC) by the World Health Organization and/or the U.S. Centers for Disease Control. A VOC is a variant for which there is evidence of an increase in transmissibility, more severe disease (e.g., increased hospitalizations or deaths), significant reduction in neutralization by antibodies generated during previous infection or vaccination, reduced effectiveness of treatments or vaccines, or diagnostic detection failures.

[0528] In some embodiments, the SARS-Cov-2 variant is classified and/or otherwise identified as a Variant of High Consequence (VHC) by the World Health Organization and/or the U.S. Centers for Disease Control. A variant of high consequence has clear evidence that prevention measures or medical countermeasures (MCMs) have significantly reduced effectiveness relative to previously circulating variants.

[0529] In some embodiments, the SARS-Cov-2 variant is classified and/or otherwise identified as a Variant of Interest (VOI) by the World Health Organization and/or the U.S. Centers for Disease Control. A VOI is a variant with specific genetic markers that have been associated with changes to receptor binding, reduced neutralization by antibodies generated against previous infection or vaccination, reduced efficacy of treatments, potential diagnostic impact, or predicted increase in transmissibility or disease severity.

[0530] In some embodiments, the SARS-Cov-2 variant is classified and/or is otherwise identified as a Variant of Note (VON). As used herein, VON refers to both "variants of concern" and "variants of note" as the two phrases are used and defined by Pangolin (cov-lineages.org) and provided in their available "VOC reports" available at cov-lineages.org.

[0531] In some embodiments the SARS-Cov-2 variant is a VOC. In some embodiments, the SARS-CoV-2 variant is or includes an Alpha variant (e.g., Pango lineage B.1.1.7), a Beta variant (e.g., Pango lineage B.1.351, B.1.351.1, B.1.351.2, and/or B.1.351.3), a Delta variant (e.g., Pango lineage B.1.617.2, AY.1, AY.2, AY.3 and/or AY.3.1); a Gamma variant (e.g., Pango lineage P.1, P.1.1, P.1.2, P.1.4, P.1.6, and/or P.1.7), a Omicon variant (B.1.1.529) or any combination thereof.

[0532] In some embodiments the SARS-Cov-2 variant is a VOI. In some embodiments, the SARS-CoV-2 variant is or includes an Eta variant (e.g., Pango lineage B.1.525 (Spike

protein substitutions A67V, 69del, 70del, 144del, E484K, D614G, Q677H, F888L)); an Iota variant (e.g., Pango lineage B.1.526 (Spike protein substitutions L5F, (D80G*), T95I, (Y144-*), (F157S*), D253G, (L452R*), (S477N*), E484K, D614G, A701V, (T859N*), (D950H*), (Q957R*))); a Kappa variant (e.g., Pango lineage B.1.617.1 (Spike protein substitutions (T95I), G142D, E154K, L452R, E484Q, D614G, P681R, Q1071H)); Pango lineage variant B.1.617.2 (Spike protein substitutions T19R, G142D, L452R, E484Q, D614G, P681R, D950N)), Lambda (e.g., Pango lineage C.37); or any combination thereof.

[0533] In some embodiments SARS-Cov-2 variant is a VON. In some embodiments, the SARS-Cov-2 variant is or includes Pango lineage variant P.1 (alias, B.1.1.28.1.) as described in Rambaut et al. 2020. Nat. Microbiol. 5:1403-1407) (spike protein substitutions: T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, H655Y, TI027I)); an Alpha variant (e.g., Pango lineage B.1.1.7); a Beta variant (e.g., Pango lineage B.1.351, B.1.351.1, B.1.351.2, and/or B.1.351.3); Pango lineage variant B.1.617.2 (Spike protein substitutions T19R, G142D, L452R, E484Q, D614G, P681R, D950N)); an Eta variant (e.g., Pango lineage B.1.525); Pango lineage variant A.23.1 (as described in Bugembe et al. medRxiv. 2021. doi: https://doi.org/10.1101/2021.02.08.21251393) (spike protein substitutions: F157L, V367F, Q613H, P681R); or any combination thereof.

Drug Resistant Viruses

[0534] In certain embodiments, the virus is a drug resistant virus. By means of example, and without limitation, the virus may be a ribavirin resistant virus. Ribavirin is a very effective antiviral that hits a number of RNA viruses. Below are a few important viruses that have evolved ribavirin resistance. Foot and Mouth Disease Virus: doi:10.1128/JVI.03594-13. Polio virus: www.pnas.org/content/100/12/7289.full.pdf. Hepatitis C Virus: jvi.asm.org/content/79/4/2346.full. A number of other persistent RNA viruses, such as hepatitis and HIV, have evolved resistance to existing antiviral drugs. Hepatitis B Virus (lamivudine, tenofovir, entecavir): doi:10.1002/hep.22900. Hepatitis C Virus (Telaprevir, BILN2061, ITMN-191, SCH6, Boceprevir, AG-021541, ACH-806): doi:10.1002/hep.22549. HIV has many drug resistant mutations, see hivdb.stanford.edu/ for more information. Aside from drug resistance, there are a number of clinically relevant mutations that could be targeted with the CRISPR systems according to the invention as described herein. For instance, persistent versus acute infection in LCMV: doi:10.1073/pnas.1019304108; or increased infectivity of Ebola: http://doi.org/10.1016/j.cell.2016.10.014 and http://doi.org/10.1016/j.cell.2016.10.013.

Malaria Detection and Monitoring

[0535] Malaria is a mosquito-borne pathology caused by *Plasmodium* parasites. The parasites are spread to people through the bites of infected female *Anopheles* mosquitoes. Five *Plasmodium* species cause malaria in humans: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae*, and *Plasmodium knowlesi.* Among them, according to the World Health Organization (WHO), *Plasmodium falciparum* and *Plasmodium vivax* are responsible for the greatest threat. *P. falciparum* is the most prevalent malaria parasite on the African continent and is responsible for most malaria-related deaths globally. *P. vivax* is the dominant malaria parasite in most countries outside of sub-Saharan Africa.

[0536] Treatment against Plasmodium sp. include aryl-amino alcohols such as quinine or quinine derivatives such as chloroquine, amodiaquine, mefloquine, piperaquine, lumefantrine, primaquine; lipophilic hydroxynaphthoquinone analog, such as atovaquone; antifolate drugs, such as the sulfa drugs sulfadoxine, dapsone and pyrimethamine; proguanil; the combination of atovaquone/proguanil; atemisins drugs; and combinations thereof. In some embodiments. The method includes screening for resistance against one or more of these compounds.

[0537] Target sequences for the assays described herein include those that are diagnostic for the presence of a mosquito-borne pathogen include a sequence that diagnostic for the presence of Plasmodium, notably *Plasmodia* species affecting humans such as *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae*, and *Plasmodium knowlesi, including sequences from the genomes thereof.*

[0538] Target sequences for the assays described herien include those that are diagnostic for monitoring drug resistance to treatment against Plasmodium, including but not limited to, *Plasmodia* species affecting humans such as *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae*, and *Plasmodium knowlesi.*

[0539] Further target sequences include sequences include target molecules/nucleic acid molecules coding for proteins involved in essential biological process for the Plasmodium parasite and notably transporter proteins, such as protein from drug/metabolite transporter family, the ATP-binding cassette (ABC) protein involved in substrate translocation, such as the ABC transporter C subfamily or the Na+/H+ exchanger, membrane glutathione S-transferase; proteins involved in the folate pathway, such as the dihydropteroate synthase, the dihydrofolate reductase activity or the dihydrofolate reductase-thymidylate synthase; and proteins involved in the translocation of protons across the inner mitochondrial membrane and

notably the cytochrome b complex. Additional target may also include the gene(s) coding for the heme polymerase.

[0540] Further target sequences include target molecules/nucleic acid molecules coding for proteins involved in essential biological process may be selected from the P. falciparum chloroquine resistance transporter gene (pfcrt), the P. falciparum multidrug resistance transporter 1 (pfmdr1), the P. falciparum multidrug resistance-associated protein gene (Pfmrp), the P. falciparum Na+/H+ exchanger gene (pfnhe), the gene coding for the P. falciparum exported protein 1, the P. falciparum Ca2+ transporting ATPase 6 (pfatp6); the P. falciparum dihydropteroate synthase (pfdhps), dihydrofolate reductase activity (pfdhpr) and dihydrofolate reductase-thymidylate synthase (pfdhfr) genes, the cytochrome b gene, gtp cyclohydrolase and the Kelch13 (K13) gene as well as their functional heterologous genes in other Plasmodium species.

[0541] A number of mutations, notably single point mutations, have been identified in the proteins which are the targets of the current malaria treatments and associated with specific resistance phenotypes. Accordingly, the invention allows for the detection of various resistance phenotypes of mosquito-borne parasites, such as plasmodium by detection of those targets that are associated with the specific resistance phenotypes.

[0542] In some embodiments, the method detects one or more mutation(s) and/or one or more single nucleotide polymorphisms in target nucleic acids/molecules. In some embodiments, any one of the mutations below, or their combination thereof, can be used as drug resistance marker and can be detected using the methods, assays, devices, compositions, and/or devices described herein.

[0543] Single point mutations in P. falciparum K13 that can be detected by an assay described herein include the following single point mutations in positions 252, 441, 446, 449, 458, 493, 539, 543, 553, 561, 568, 574, 578, 580, 675, 476, 469, 481, 522, 537, 538, 579, 584 and 719 and notably mutations E252Q, P441L, F446I, G449A, N458Y, Y493H, R539T, I543T, P553L, R561H, V568G, P574L, A578S, C580Y, A675V, M476I; C469Y; A481V; S522C; N537I; N537D; G538V; M579I; D584V; and H719N. These mutations are generally associated with artemisins drugs resistance phenotypes (Artemisinin and artemisinin-based combination therapy resistance, April 2016 WHO/HTM/GMP/2016.5).

[0544] Mutations in the P. falciparum dihydrofolate reductase (DHFR) (PfDHFR-TS, PFD0830w) that can be detected by the assays described herein include mutations in positions

108, 51, 59 and 164, notably 108 D, 164L, 51I and 59R which modulate resistance to pyrimethamine. Other polymorphisms that can be detected by the methods described herein include 437G, 581G, 540E, 436A and 613S, which are associated with resistance to sulfadoxine. Additional mutations that can be detected by the assays described herein include Ser108Asn, Asn51Ile, Cys59Arg, Ile164Leu, Cys50Arg, Ile164Leu, Asn188Lys, Ser189Arg and Val213Ala, Ser108Thr and Ala16Val. Mutations Ser108Asn, Asn51Ile, Cys59Arg, Ile164Leu, Cys50Arg, Ile164Leu are notably associated with pyrimethamine based therapy and/or chloroguanine-dapsone combination therapy resistances and can be detected by the assays described herein. Cycloguanil resistance appears to be associated with the double mutations Ser108Thr and Ala16Val, which can be detected by the assays described herein. Amplification of dhfr may also be of high relevance for therapy resistance notably pyrimethamine resistance and can be detected ny the assays described herein.

[0545] Mutations in the P. falciparum dihydropteroate synthase (DHPS) (PfDHPS, PF08_0095) can be detected by the assays described herein, and include, without limitation, mutations in positions 436, 437, 581 and 613 Ser436Ala/Phe, Ala437Gly, Lys540Glu, Ala581Gly and Ala613Thr/Ser. Polymorphism in position 581 and / or 613 have also been associated with resistance to sulfadoxine–pyrimethamine base therapies and can be detected by an assay described herein.

[0546] Mutations in the P. falciparum chloroquine-resistance transporter (PfCRT) can be detected by the assays described herein. In some embodiments, the polymorphism in position 76, notably the mutation Lys76Thr, is associated with resistance to chloroquine and can be detected by an assay described herein. Further polymorphisms include Cys72Ser, Met74Ile, Asn75Glu, Ala220Ser, Gln271Glu, Asn326Ser, Ile356Thr and Arg371Ile which may be associated with chloroquine resistance can be detected by an assay described herein. PfCRT is also phosphorylated at the residues S33, S411 and T416, which may regulate the transport activity or specificity of the protein, which can be detected by an assay described herein.

[0547] Mutations in the P. falciparum multidrug-resistance transporter 1 (PfMDR1) (PFE1150w) can be detected by an assay described herein. For example, polymorphisms in positions 86, 184, 1034, 1042, notably Asn86Tyr, Tyr184-Phe, Ser1034Cys, Asn1042Asp and Asp1246Tyr have been identified and reported to influence have been reported to influence susceptibilities to lumefantrine, artemisinin, quinine, mefloquine, halofantrine and chloroquine and can be detected by an assay described herein. Additionally, amplification of PfMDR1 is

associated with reduced susceptibility to lumefantrine, artemisinin, quinine, mefloquine, and halofantrine and can be detected by an assay described herein. Deamplification of PfMDR1 leads to an increase in chloroquine resistance and can be detected by an assay described herein. Amplification of pfmdr1 may also be detected. The phosphorylation status of PfMDR1is also of high relevance and can be detected by an assay described herein.

[0548] Mutations in the P. falciparum multidrug-resistance associated protein (PfMRP) (gene reference PFA0590w) can be detected by an assay described herein. For example polymorphisms in positions 191 and/or 437, such as Y191H and A437S have been identified and associated with chloroquine resistance phenotypes and can be detected by an assay described herein.

[0549] Mutations in the P. falciparum NA+/H+ enchanger (PfNHE) (ref PF13_0019) can be detected by an assay described herein. For example, increased repetition of the DNNND in microsatellite ms4670 may be a marker for quinine resistance and can be detected by an assay described herein.

[0550] Mutations altering the ubiquinol binding site of the cytochrome b protein encoded by the cytochrome be gene (cytb, mal_mito_3) are associated with atovaquone resistance and can be detected by an assay described herein. Mutations in positions 26, 268, 276, 133 and 280 and notably Tyr26Asn, Tyr268Ser, M1331 and G280D may be associated with atovaquone resistance and can be detected by an assay described herein.

[0551] In P Vivax, mutations in PvMDR1, the homolog of Pf MDR1 have been associated with chloroquine resistance, notably polymorphism in position 976 such as the mutation Y976F and can be detected by an assay described herein.

[0552] The above mutations are defined in terms of protein sequences. However, the skilled person is able to determine the corresponding mutations, including SNPs, to be identified as a nucleic acid target sequence.

[0553] Other identified drug-resistance markers are known in the art, for example as described in "Susceptibility of Plasmodium falciparum to antimalarial drugs (1996–2004)"; WHO; Artemisinin and artemisinin-based combination therapy resistance (April 2016 WHO/HTM/GMP/2016.5); "Drug-resistant malaria: molecular mechanisms and implications for public health" FEBS Lett. 2011 Jun 6;585(11):1551-62. doi:10.1016/j.febslet.2011.04.042. Epub 2011 Apr 23. Review. PubMed PMID: 21530510; the contents of which are herewith incorporated by reference and can be detected by an assay described herein.

[0554] As to polypeptides that may be detected in accordance with the present invention, gene products of all genes mentioned herein may be used as targets. Correspondingly, it is contemplated that such polypeptides could be used for species identification, typing and/or detection of drug resistance.

[0555] In certain example embodiments, the systems, devices, and methods, disclosed herein are directed to detecting the presence of one or more mosquito-borne parasite in a sample, such as a biological sample obtained from a subject. In certain example embodiments, the parasite may be selected from the species Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale, Plasmodium malariae or Plasmodium knowlesi. Accordingly, the methods disclosed herein can be adapted for use in other methods (or in combination) with other methods that require quick identification of parasite species, monitoring the presence of parasites and parasite forms (for example corresponding to various stages of infection and parasite life-cycle, such as exo-erythrocytic cycle, erythrocytic cyle, sporpogonic cycle; parasite forms include merozoites, sporozoites, schizonts, gametocytes); detection of certain phenotypes (e.g., pathogen drug resistance), monitoring of disease progression and/or outbreak, and treatment (drug) screening. Further, in the case of malaria, a long time may elapse following the infective bite, namely a long incubation period, during which the patient does not show symptoms. Similarly, prophylactic treatments can delay the appearance of symptoms, and long asymptomatic periods can also be observed before a relapse. Such delays can easily cause misdiagnosis or delayed diagnosis, and thus impair the effectiveness of treatment.

[0556] Because of the rapid and sensitive diagnostic capabilities of the embodiments disclosed here, detection of parasite type, down to a single nucleotide difference, and the ability to be deployed as a POC device, the embodiments disclosed herein may be used guide therapeutic regimens, such as selection of the appropriate course of treatment. The embodiments disclosed herein may also be used to screen environmental samples (mosquito population, etc.) for the presence and the typing of the parasite. The embodiments may also be modified to detect mosquito-borne parasites and other mosquito-borne pathogens simultaneously. In some instances, malaria and other mosquito-borne pathogens may present initially with similar symptoms. Thus, the ability to quickly distinguish the type of infection can guide important treatment decisions. Other mosquito-born pathogens that may be detected in conjunction with malaria include dengue, West Nile virus, chikungunya, yellow fever,

filariasis, Japanese encephalitis, Saint Louis encephalitis, western equine encephalitis, eastern equine encephalitis, Venezuelan equine encephalitits, La Crosse encephalitis, and zika.

[0557] In certain example embodiments, the devices, systems, and methods disclosed herein may be used to distinguish multiple mosquito-borne parasite species in a sample. In certain example embodiments, identification may be based on ribosomal RNA sequences, including the 18S, 16S, 23S, and 5S subunits. In certain example embodiments, identification may be based on sequences of genes that are present in multiple copies in the genome, such as mitochondrial genes like CYTB. In certain example embodiments, identification may be based on sequences of genes that are highly expressed and/or highly conserved such as GAPDH, Histone H2B, enolase, or LDH. Methods for identifying relevant rRNA sequences are disclosed in U.S. Patent Application Publication No. 2017/0029872. In certain example embodiments, a set of guide RNA may be designed to distinguish each species by a variable region that is unique to each species or strain. Guide RNAs may also be designed to target RNA genes that distinguish microbes at the genus, family, order, class, phylum, kingdom levels, or a combination thereof. In certain example embodiments where amplification is used, a set of amplification primers may be designed to flanking constant regions of the ribosomal RNA sequence and a guide RNA designed to distinguish each species by a variable internal region. In certain example embodiments, the primers and guide RNAs may be designed to conserved and variable regions in the 16S subunit respectfully. Other genes or genomic regions that uniquely variable across species or a subset of species such as the RecA gene family, RNA polymerase β subunit, may be used as well. Other suitable phylogenetic markers, and methods for identifying the same, are discussed for example in Wu et al. arXiv:1307.8690 [q-bio.GN].

[0558] In certain example embodiments, species identification can be performed based on genes that are present in multiple copies in the genome, such as mitochondrial genes like CYTB. In certain example embodiments, species identification can be performed based on highly expressed and/or highly conserved genes such as GAPDH, Histone H2B, enolase, or LDH.

[0559] In certain example embodiments, a method or diagnostic is designed to screen mosquito-borne parasites across multiple phylogenetic and/or phenotypic levels at the same time. For example, the method or diagnostic may comprise the use of multiple programmable pattern recognition compositions, CRISPR systems, with different guide RNAs when used. A first set of guide RNAs may distinguish, for example, between Plasmodium falciparum or

Plasmodium vivax. These general classes can be even further subdivided. For example, guide RNAs could be designed and used in the method or diagnostic that distinguish drug-resistant strains, in general or with respect to a specific drug or combination of drugs. A second set of guide RNA can be designed to distinguish microbes at the species level. Thus, a matrix may be produced identifying all mosquito-borne parasites species or subspecies, further divided according to drug resistance. The foregoing is for example purposes only. Other means for classifying other types of mosquito-borne parasites are also contemplated and would follow the general structure described above.

[0560] In certain example embodiments, the devices, systems and methods disclosed herein may be used to screen for mosquito-borne parasite genes of interest, for example drug resistance genes. Guide RNAs may be designed to distinguish between known genes of interest. Samples, including clinical samples, may then be screened using the embodiments disclosed herein for detection of one or more such genes. The ability to screen for drug resistance at POC would have tremendous benefit in selecting an appropriate treatment regime. In certain example embodiments, the drug resistance genes are genes encoding proteins such as transporter proteins, such as protein from drug/metabolite transporter family, the ATP-binding cassette (ABC) protein involved in substrate translocation, such as the ABC transporter C subfamily or the Na+/H+ exchanger; proteins involved in the folate pathway, such as the dihydropteroate synthase, the dihydrofolate reductase activity or the dihydrofolate reductase-thymidylate synthase; and proteins involved in the translocation of protons across the inner mitochondrial membrane and notably the cytochrome b complex. Additional targets may also include the gene(s) coding for the heme polymerase. In certain example embodiments, the drug resistance genes are selected from the P. falciparum chloroquine resistance transporter gene (pfcrt), the P. falciparum multidrug resistance transporter 1 (pfmdr1), the P. falciparum multidrug resistance-associated protein gene (Pfmrp), the P. falciparum Na+/H+ exchanger gene (pfnhe), the P. falciparum Ca2+ transporting ATPase 6 (pfatp6), the P. falciparum dihydropteroate synthase (pfdhps), dihydrofolate reductase activity (pfdhpr) and dihydrofolate reductase-thymidylate synthase (pfdhfr) genes, the cytochrome b gene, gtp cyclohydrolase and the Kelch13 (K13) gene as well as their functional heterologous genes in other Plasmodium species. Other identified drug-resistance markers are known in the art, for example as described in “Susceptibility of Plasmodium falciparum to antimalarial drugs (1996–2004)”; WHO; Artemisinin and artemisinin-based combination therapy resistance (April 2016

WHO/HTM/GMP/2016.5); "Drug-resistant malaria: molecular mechanisms and implications for public health" FEBS Lett. 2011 Jun 6;585(11):1551-62. doi:10.1016/j.febslet.2011.04.042. Epub 2011 Apr 23. Review. PubMed PMID: 21530510; the contents of which are herewith incorporated by reference.

[0561] In some embodiments, a programmable pattern recognition compositions, detection system or methods of use thereof as described herein may be used to determine the evolution of a mosquito-borne parasite outbreak. The method may comprise detecting one or more target molecules or cells from a plurality of samples from one or more subjects, wherein the target molecule or cells are from a mosquito-borne parasite spreading or causing the outbreaks. Such a method may further comprise determining a pattern of mosquito-borne parasite transmission, or a mechanism involved in a disease outbreak caused by a mosquito-borne parasite. The samples may be derived from one or more humans, and/or be derived from one or more mosquitoes.

[0562] The pattern of pathogen transmission may comprise continued new transmissions from the natural reservoir of the mosquito-borne parasite or other transmissions (e.g., across mosquitoes) following a single transmission from the natural reservoir or a mixture of both. In one embodiment, the target molecule is preferably a nucleic acid sequence within the mosquito-borne parasite genome or fragments thereof. In one embodiment, the pattern of the mosquito-borne parasite transmission is the early pattern of the mosquito-borne parasite transmission, i.e., at the beginning of the mosquito-borne parasite outbreak. Determining the pattern of the mosquito-borne parasite transmission at the beginning of the outbreak increases likelihood of stopping the outbreak at the earliest possible time thereby reducing the possibility of local and international dissemination.

[0563] Determining the pattern of the mosquito-borne parasite transmission may comprise detecting a mosquito-borne parasite target molecule or cell according to the methods described herein. Determining the pattern of the pathogen transmission may further comprise detecting shared intra-host variations of the mosquito-borne parasite sequence between the subjects and determining whether the shared intra-host variations show temporal patterns. Patterns in observed intrahost and interhost variation provide important insight about transmission and epidemiology (Gire, et al., 2014).

[0564] In addition to other sample types disclosed herein, the sample may be derived from one or more mosquitoes, for example the sample may comprise mosquito saliva.

Biomarker Detection and Applications

[0565] In certain example embodiments, the programmable pattern recognition compositions systems, devices, and methods disclosed herein may be used for biomarker detection. For example, the systems, devices and method disclosed herein may be used for SNP detection and/or genotyping, transcript detection, and/or protein detection. The programmable pattern recognition compositions systems, devices and methods disclosed herein may be also used for the detection of any disease state or disorder characterized by aberrant gene expression. Aberrant gene expression includes aberration in the gene expressed, location of expression and level of expression. Multiple transcripts or protein markers related to cardiovascular, immune disorders, and cancer among other diseases may be detected. In certain example embodiments, the embodiments disclosed herein may be used for cell free DNA detection of diseases that involve lysis, such as liver fibrosis and restrictive/obstructive lung disease. In certain example embodiments, the pattern recognition compositions are utilized for faster and more portable detection for pre-natal testing of cell-free DNA. The embodiments disclosed herein may be used for screening panels of different SNPs or other biomarkers associated with, among others, cardiovascular health, lipid/metabolic signatures, ethnicity identification, paternity matching, human ID (e.g., matching suspect to a criminal database of SNP signatures). The embodiments disclosed herein may also be used for cell free DNA detection of mutations related to and released from cancer tumors. The embodiments disclosed herein may also be used for detection of meat quality, for example, by providing rapid detection of different animal sources in a given meat product. Embodiments disclosed herein may also be used for the detection of GMOs or gene editing related to DNA. As described herein elsewhere, closely related genotypes/alleles or biomarkers (e.g., having only a single nucleotide difference in a given target sequence) may be distinguished by introduction of a synthetic mismatch in a gRNA employed or included with the pattern recognition compositions and/or systems described herein.

[0566] In an aspect, the invention relates to a method for detecting target nucleic acids and/or polypeptides, and/or cells in samples, comprising distributing a sample or set of samples into one or more individual discrete volumes, the individual discrete volumes comprising a detection composition according to the invention as described herein; incubating the sample or set of samples under conditions sufficient to allow recognition and/or binding of the pattern recognition compositions to one or more PAMPs so as to activate the pattern recognition

compositions; activating an effector protein or domain of the detection composition via binding of the one or more guide RNAs to the one or more target molecules, wherein activating the detection composition effector protein results in modification of the detection construct such that a detectable signal is generated; and detecting the detectable signal, wherein detection of the detectable e signal indicates a presence of one or more target molecules in the sample.

Detecting Circulating Tumor Cells

[0567] In one embodiment, circulating cells (e.g., circulating tumor cells (CTC)) can be assayed with the present invention. Isolation of circulating tumor cells (CTC) for use in any of the methods described herein may be performed. Exemplary technologies that achieve specific and sensitive detection and capture of circulating cells that may be used in the present invention have been described (Mostert B, et al., Circulating tumor cells (CTCs): detection methods and their clinical relevance in breast cancer. Cancer Treat Rev. 2009;35:463-474; and Talasaz AH, et al., Isolating highly enriched populations of circulating epithelial cells and other rare cells from blood using a magnetic sweeper device. Proc Natl Acad Sci U S A. 2009; 106:3970-3975). As few as one CTC may be found in the background of 105-106 peripheral blood mononuclear cells (Ross A A, et al., Detection and viability of tumor cells in peripheral blood stem cell collections from breast cancer patients using immunocytochemical and clonogenic assay techniques. Blood. 1993,82:2605-2610). The CellSearch® platform uses immunomagnetic beads coated with antibodies to Epithelial Cell Adhesion Molecule (EpCAM) to enrich for EPCAM-expressing epithelial cells, followed by immunostaining to confirm the presence of cytokeratin staining and absence of the leukocyte marker CD45 to confirm that captured cells are epithelial tumor cells (Momburg F, et al., Immunohistochemical study of the expression of a Mr 34,000 human epithelium-specific surface glycoprotein in normal and malignant tissues. Cancer Res. 1987;47:2883-2891; and Allard WJ, et al., Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. Clin Cancer Res. 2004; 10:6897-6904). The number of cells captured have been prospectively demonstrated to have prognostic significance for breast, colorectal and prostate cancer patients with advanced disease (Cohen SJ, et al., J Clin Oncol. 2008;26:3213-3221; Cristofanilli M, et al. N Engl J Med. 2004;351:781-791; Cristofanilli M, et al., J Clin Oncol. 2005;23: 1420-1430; and de Bono JS, et al. Clin Cancer Res. 2008; 14:6302-6309).

[0568] The present invention also provides for isolating CTCs with CTC-Chip Technology. CTC-Chip is a microfluidic based CTC capture device where blood flows through a chamber containing thousands of microposts coated with anti-EpCAM antibodies to which the CTCs bind (Nagrath S, et al. Isolation of rare circulating tumour cells in cancer patients by microchip technology. Nature. 2007;450: 1235-1239). CTC-Chip provides a significant increase in CTC counts and purity in comparison to the CellSearch® system (Maheswaran S, et al. Detection of mutations in EGFR in circulating lung-cancer cells, N Engl J Med. 2008;359:366-377), both platforms may be used for downstream molecular analysis.

Cell-Free Chromatin

[0569] In certain embodiments, cell free chromatin fragments are isolated and analyzed according to the present invention. Nucleosomes can be detected in the serum of healthy individuals (Stroun et al., Annals of the New York Academy of Sciences 906: 161-168 (2000)) as well as individuals afflicted with a disease state. Moreover, the serum concentration of nucleosomes is considerably higher in patients suffering from benign and malignant diseases, such as cancer and autoimmune disease (Holdenrieder et al (2001) Int J Cancer 95, 1 14-120, Trejo-Becerril et al (2003) Int J Cancer 104, 663-668; Kuroi et al 1999 Breast Cancer 6, 361-364; Kuroi et al (2001) Int j Oncology 19, 143-148; Amoura et al (1997) Arth Rheum 40, 2217-2225; Williams et al (2001) J Rheumatol 28, 81-94). Not being bound by a theory, the high concentration of nucleosomes in tumor bearing patients derives from apoptosis, which occurs spontaneously in proliferating tumors. Nucleosomes circulating in the blood contain uniquely modified histones. For example, U.S. Patent Publication No. 2005/0069931 (Mar. 31, 2005) relates to the use of antibodies directed against specific histone N-terminus modifications as diagnostic indicators of disease, employing such histone-specific antibodies to isolate nucleosomes from a blood or serum sample of a patient to facilitate purification and analysis of the accompanying DNA for diagnostic/screening purposes. Accordingly, the present invention may use chromatin bound DNA to detect and monitor, for example, tumor mutations. The identification of the DNA associated with modified histones can serve as diagnostic markers of disease and congenital defects.

[0570] Thus, in another embodiment, isolated chromatin fragments are derived from circulating chromatin, preferably circulating mono and oligonucleosomes. Isolated chromatin fragments may be derived from a biological sample. The biological sample may be from a

subject or a patient in need thereof. The biological sample may be sera, plasma, lymph, blood, blood fractions, urine, synovial fluid, spinal fluid, saliva, circulating tumor cells or mucous.

Cell-free DNA (cfDNA)

[0571] In certain embodiments, the present invention may be used to detect cell free DNA (cfDNA). Cell free DNA in plasma or serum may be used as a non-invasive diagnostic tool. For example, cell free fetal DNA has been studied and optimized for testing on-compatible RhD factors, sex determination for X-linked genetic disorders, testing for single gene disorders, identification of preeclampsia. For example, sequencing the fetal cell fraction of cfDNA in maternal plasma is a reliable approach for detecting copy number changes associated with fetal chromosome aneuploidy. For another example, cfDNA isolated from cancer patients has been used to detect mutations in key genes relevant for treatment decisions.

[0572] In certain example embodiments, the present disclosure provides detecting cfDNA directly from a patient sample. In certain other example embodiment, the present disclosure provides enriching cfDNA using the enrichment embodiments disclosed above and prior to detecting the target cfDNA.

Exosomes

[0573] In one embodiment, exosomes can be assayed with the present invention. Exosomes are small extracellular vesicles that have been shown to contain RNA. Isolation of exosomes by ultracentrifugation, filtration, chemical precipitation, size exclusion chromatography, and microfluidics are known in the art. In one embodiment exosomes are purified using an exosome biomarker. Isolation and purification of exosomes from biological samples may be performed by any known methods (see e.g., WO2016172598A1).

SNP Detection and Genotyping

[0574] In certain embodiments, the present invention may be used to detect the presence of single nucleotide polymorphisms (SNP) in a biological sample. The SNPs may be related to maternity testing (e.g., sex determination, fetal defects). They may be related to a criminal investigation. In one embodiment, a suspect in a criminal investigation may be identified by the present invention. Not being bound by a theory nucleic acid based forensic evidence may require the most sensitive assay available to detect a suspect or victim's genetic material because the samples tested may be limiting.

[0575] In other embodiments, SNPs associated with a disease are encompassed by the present invention. SNPs associated with diseases are well known in the art and one skilled in

the art can apply the methods of the present invention to design suitable guide RNAs (see e.g., www.ncbi.nlm.nih.gov/clinvar?term=human%5Borgn%5D).

[0576] In an aspect, the invention relates to a method for genotyping, such as SNP genotyping, comprising: distributing a sample or set of samples into one or more individual discrete volumes, the individual discrete volumes comprising a detection composition or system according to the invention as described herein; incubating the sample or set of samples under conditions sufficient to allow binding of the one or more guide RNAs to one or more target molecules; activating the detection composition effector protein via binding of the one or more guide RNAs to the one or more target molecules, wherein activating the detection composition effector protein results in modification of the detection construct such that a detectable signal is generated; and detecting the detectable signal, wherein detection of the detectable signal indicates a presence of one or more target molecules characteristic for a particular genotype in the sample.

[0577] In certain embodiments, the detectable signal is compared to (e.g., by comparison of signal intensity) one or more standard signal, preferably a synthetic standard signal). In certain embodiments, the standard is or corresponds to a particular genotype. In certain embodiments, the standard comprises a particular SNP or other (single) nucleotide variation. In certain embodiments, the standard is a (PCR-amplified) genotype standard. In certain embodiments, the standard is or comprises DNA. In certain embodiments, the standard is or comprises RNA. In certain embodiments, the standard is or comprised RNA which is transcribed from DNA. In certain embodiments, the standard is or comprises DNA which is reverse transcribed from RNA. In certain embodiments, the detectable signal is compared to one or more standard, each of which corresponds to a known genotype, such as a SNP or other (single) nucleotide variation. In certain embodiments, the detectable signal is compared to one or more standard signal and the comparison comprises statistical analysis, such as by parametric or non-parametric statistical analysis, such as by one- or two-way ANOVA, etc. In certain embodiments, the detectable signal is compared to one or more standard signal and when the detectable signal does not (statistically) significantly deviate from the standard, the genotype is determined as the genotype corresponding to said standard.

[0578] In other embodiments, the present invention allows rapid genotyping for emergency pharmacogenomics. In one embodiment, a single point of care assay may be used to genotype a patient brought into the emergency room. The patient may be suspected of having a blood

clot and an emergency physician needs to decide a dosage of blood thinner to administer. In exemplary embodiments, the present invention may provide guidance for administration of blood thinners during myocardial infarction or stroke treatment based on genotyping of markers such as VKORC1, CYP2C9, and CYP2C19. In one embodiment, the blood thinner is the anticoagulant warfarin (Holford, NH (December 1986). "Clinical Pharmacokinetics and Pharmacodynamics of Warfarin Understanding the Dose-Effect Relationship". Clinical Pharmacokinetics. Springer International Publishing. 11 (6): 483–504). Genes associated with blood clotting are known in the art (see e.g., US20060166239A1; Litin SC, Gastineau DA (1995) "Current concepts in anticoagulant therapy". Mayo Clin. Proc. 70 (3): 266–72; and Rusdiana et al., Responsiveness to low-dose warfarin associated with genetic variants of VKORC1, CYP2C9, CYP2C19, and CYP4F2 in an Indonesian population. Eur J Clin Pharmacol. 2013 Mar;69(3):395-405). Specifically, in the VKORC1 1639 (or 3673) single-nucleotide polymorphism, the common ("wild-type") G allele is replaced by the A allele. People with an A allele (or the "A haplotype") produce less VKORC1 than do those with the G allele (or the "non-A haplotype"). The prevalence of these variants also varies by race, with 37% of Caucasians and 14% of Africans carrying the A allele. The end result is a decreased number of clotting factors and therefore, a decreased ability to clot.

[0579] In certain example embodiments, the availability of genetic material for detecting a SNP in a patient allows for detecting SNPs without amplification of a DNA or RNA sample. In the case of genotyping, the biological sample tested is easily obtained. In certain example embodiments, the incubation time of the present invention may be shortened. The assay may be performed in a period of time required for an enzymatic reaction to occur. One skilled in the art can perform biochemical reactions in 5 minutes (e.g., 5 minute ligation). The present invention may use an automated DNA extraction device to obtain DNA from blood. The DNA can then be added to a reaction that generates a target molecule for the effector protein. Immediately upon generating the target molecule the masking agent can be cut and a signal detected. In exemplary embodiments, the present invention allows a POC rapid diagnostic for determining a genotype before administering a drug (e.g., blood thinner). In the case where an amplification step is used, all of the reactions occur in the same reaction in a one step process. In preferred embodiments, the POC assay may be performed in less than an hour, preferably 10 minutes, 20 minutes, 30 minutes, 40 minutes, or 50 minutes.

[0580] In certain embodiments, the systems, devices, and methods disclosed herein may be used for detecting the presence or expression level of long non-coding RNAs (lncRNAs). Expression of certain lncRNAs are associated with disease state and/or drug resistance. In particular, certain lncRNAs (e.g., TCONS_00011252, NR_034078, TCONS_00010506, TCONS_00026344, TCONS_00015940, TCONS_00028298, TCONS_00026380, TCONS_0009861, TCONS_00026521, TCONS_00016127, NR_125939, NR_033834, TCONS_00021026, TCONS_00006579, NR_109890, and NR_026873) are associated with resistance to cancer treatment, such as resistance to one or more BRAF inhibitors (e.g., Vemurafenib, Dabrafenib, Sorafenib, GDC-0879, PLX-4720, and LGX818) for treating melanoma (e.g., nodular melanoma, lentigo maligna, lentigo maligna melanoma, acral lentiginous melanoma, superficial spreading melanoma, mucosal melanoma, polypoid melanoma, desmoplastic melanoma, amelanotic melanoma, and soft-tissue melanoma). The detection of lncRNAs using the various embodiments described herein can facilitate disease diagnosis and/or selection of treatment options.

[0581] In one embodiment, the present invention can guide DNA- or RNA-targeted therapies (e.g., CRISPR, TALE, Zinc finger proteins, RNAi), particularly in settings where rapid administration of therapy is important to treatment outcomes.

LOH Detection

[0582] Cancer cells undergo a loss of genetic material (DNA) when compared to normal cells. This deletion of genetic material which almost all, if not all, cancers undergo is referred to as “loss of heterozygosity” (LOH). Loss of heterozygosity (LOH) is a gross chromosomal event that results in loss of the entire gene and the surrounding chromosomal region. The loss of heterozygosity is a common occurrence in cancer, where it can indicate the absence of a functional tumor suppressor gene in the lost region. However, a loss may be silent because there still is one functional gene left on the other chromosome of the chromosome pair. The remaining copy of the tumor suppressor gene can be inactivated by a point mutation, leading to loss of a tumor suppressor gene. The loss of genetic material from cancer cells can result in the selective loss of one of two or more alleles of a gene vital for cell viability or cell growth at a particular locus on the chromosome.

[0583] An “LOH marker” is DNA from a microsatellite locus, a deletion, alteration, or amplification in which, when compared to normal cells, is associated with cancer or other

diseases. An LOH marker often is associated with loss of a tumor suppressor gene or another, usually tumor related, gene.

[0584] The term "microsatellites" refers to short repetitive sequences of DNA that are widely distributed in the human genome. A microsatellite is a tract of tandemly repeated (i.e., adjacent) DNA motifs that range in length from two to five nucleotides, and are typically repeated 5-50 times. For example, the sequence TATATATATA (SEQ ID NO: 17) is a dinucleotide microsatellite, and GTCGTCGTCGTCGTC (SEQ ID NO: 18) is a trinucleotide microsatellite (with A being Adenine, G Guanine, C Cytosine, and T Thymine). Somatic alterations in the repeat length of such microsatellites have been shown to represent a characteristic feature of tumors. Guide RNAs may be designed to detect such microsatellites. Furthermore, the present invention may be used to detect alterations in repeat length, as well as amplifications and deletions based upon quantitation of the detectable signal. Certain microsatellites are located in regulatory flanking or intronic regions of genes, or directly in codons of genes. Microsatellite mutations in such cases can lead to phenotypic changes and diseases, notably in triplet expansion diseases such as fragile X syndrome and Huntington's disease.

[0585] Frequent loss of heterozygosity (LOH) on specific chromosomal regions has been reported in many kinds of malignancies. Allelic losses on specific chromosomal regions are the most common genetic alterations observed in a variety of malignancies, thus microsatellite analysis has been applied to detect DNA of cancer cells in specimens from body fluids, such as sputum for lung cancer and urine for bladder cancer. (Rouleau, et al. Nature 363, 515-521 (1993); and Latif, et al. Science 260, 1317-1320 (1993)). Moreover, it has been established that markedly increased concentrations of soluble DNA are present in plasma of individuals with cancer and some other diseases, indicating that cell free serum or plasma can be used for detecting cancer DNA with microsatellite abnormalities. (Kamp, et al. Science 264, 436-440 (1994); and Steck, et al. Nat Genet. 15(4), 356-362 (1997)). Two groups have reported microsatellite alterations in plasma or serum of a limited number of patients with small cell lung cancer or head and neck cancer. (Hahn, et al. Science 271, 350-353 (1996); and Miozzo, et al. Cancer Res. 56, 2285-2288 (1996)). Detection of loss of heterozygosity in tumors and serum of melanoma patients has also been previously shown (see, e.g., United States patent number US6465177B1).

[0586] Thus, it is advantageous to detect of LOH markers in a subject suffering from or at risk of cancer. The present invention may be used to detect LOH in tumor cells. In one embodiment, circulating tumor cells may be used as a biological sample. In preferred embodiments, cell free DNA obtained from serum or plasma is used to noninvasively detect and/or monitor LOH. In other embodiments, the biological sample may be any sample described herein (e.g., a urine sample for bladder cancer). Not being bound by a theory, the present invention may be used to detect LOH markers with improved sensitivity as compared to any prior method, thus providing early detection of mutational events. In one embodiment, LOH is detected in biological fluids, wherein the presence of LOH is associated with the occurrence of cancer. The method and systems described herein represents a significant advance over prior techniques, such as PCR or tissue biopsy by providing a non-invasive, rapid, and accurate method for detecting LOH of specific alleles associated with cancer. Thus, the present invention provides a methods and systems which can be used to screen high-risk populations and to monitor high risk patients undergoing chemoprevention, chemotherapy, immunotherapy or other treatments.

[0587] Because the method of the present invention requires only DNA extraction from bodily fluid such as blood, it can be performed at any time and repeatedly on a single patient. Blood can be taken and monitored for LOH before or after surgery; before, during, and after treatment, such as chemotherapy, radiation therapy, gene therapy or immunotherapy; or during follow-up examination after treatment for disease progression, stability, or recurrence. Not being bound by a theory, the method of the present invention also may be used to detect subclinical disease presence or recurrence with an LOH marker specific for that patient since LOH markers are specific to an individual patient's tumor. The method also can detect if multiple metastases may be present using tumor specific LOH markers.

Detection of Epigenetic Modifications

[0588] Histone variants, DNA modifications, and histone modifications indicative of cancer or cancer progression may be used in the present invention. For example, U.S. patent publication 20140206014 describes that cancer samples had elevated nucleosome H2AZ, macroH2A1.1, 5-methylcytosine, P-H2AX(Ser139) levels as compared to healthy subjects. The presence of cancer cells in an individual may generate a higher level of cell free nucleosomes in the blood as a result of the increased apoptosis of the cancer cells. In one embodiment, an antibody directed against marks associated with apoptosis, such as H2B Ser

14(P), may be used to identify single nucleosomes that have been released from apoptotic neoplastic cells. Thus, DNA arising from tumor cells may be advantageously analyzed according to the present invention with high sensitivity and accuracy.

Pre-natal Screening

[0589] In certain embodiments, the method and systems of the present invention may be used in prenatal screening. In certain embodiments, cell-free DNA is used in a method of prenatal screening. In certain embodiments, DNA associated with single nucleosomes or oligonucleosomes may be detected with the present invention. In preferred embodiments, detection of DNA associated with single nucleosomes or oligonucleosomes is used for prenatal screening. In certain embodiments, cell-free chromatin fragments are used in a method of prenatal screening.

[0590] Prenatal diagnosis or prenatal screening refers to testing for diseases or conditions in a fetus or embryo before it is born. The aim is to detect birth defects such as neural tube defects, Down syndrome, chromosome abnormalities, genetic disorders and other conditions, such as spina bifida, cleft palate, Tay Sachs disease, sickle cell anemia, thalassemia, cystic fibrosis, Muscular dystrophy, and fragile X syndrome. Screening can also be used for prenatal sex discernment. Common testing procedures include amniocentesis, ultrasonography including nuchal translucency ultrasound, serum marker testing, or genetic screening. In some cases, the tests are administered to determine if the fetus will be aborted, though physicians and patients also find it useful to diagnose high-risk pregnancies early so that delivery can be scheduled in a tertian,' care hospital where the baby can receive appropriate care.

[0591] It has been realized that there are fetal cells which are present in the mother's blood, and that these cells present a potential source of fetal chromosomes for prenatal DNA-based diagnostics. Additionally, fetal DNA ranges from about 2-10% of the total DNA in maternal blood. Currently available prenatal genetic tests usually involve invasive procedures. For example, chorionic villus sampling (CVS) performed on a pregnant woman around 10-12 weeks into the pregnancy and amniocentesis performed at around 14-16 weeks all contain invasive procedures to obtain the sample for testing chromosomal abnormalities in a fetus. Fetal cells obtained via these sampling procedures are usually tested for chromosomal abnormalities using cytogenetic or fluorescent in situ hybridization (FISH) analyses. Cell-free fetal DNA has been shown to exist in plasma and serum of pregnant women as early as the sixth week of gestation, with concentrations rising during pregnancy and peaking prior to

parturition. Because these cells appear very early in the pregnancy, they could form the basis of an accurate, noninvasive, first trimester test. Not being bound by a theory, the present invention provides unprecedented sensitivity in detecting low amounts of fetal DNA. Not being bound by a theory, abundant amounts of maternal DNA is generally concomitantly recovered along with the fetal DNA of interest, thus decreasing sensitivity in fetal DNA quantification and mutation detection. The present invention overcomes such problems by the unexpectedly high sensitivity of the assay.

[0592] The H3 class of histones consists of four different protein types: the main types, H3.1 and H3.2; the replacement type, H3.3; and the testis specific variant, H3t. Although H3.1 and H3.2 are closely related, only differing at Ser96, H3.1 differs from H3.3 in at least 5 amino acid positions. Further, H3.1 is highly enriched in fetal liver, in comparison to its presence in adult tissues including liver, kidney and heart. In adult human tissue, the H3.3 variant is more abundant than the H3.1 variant, whereas the converse is true for fetal liver. The present invention may use these differences to detect fetal nucleosomes and fetal nucleic acid in a maternal biological sample that comprises both fetal and maternal cells and/or fetal nucleic acid.

[0593] In one embodiment, fetal nucleosomes may be obtained from blood. In other embodiments, fetal nucleosomes are obtained from a cervical mucus sample. In certain embodiments, a cervical mucus sample is obtained by swabbing or lavage from a pregnant woman early in the second trimester or late in the first trimester of pregnancy. The sample may be placed in an incubator to release DNA trapped in mucus. The incubator may be set at 37° C. The sample may be rocked for approximately 15 to 30 minutes. Mucus may be further dissolved with a mucinase for the purpose of releasing DNA. The sample may also be subjected to conditions, such as chemical treatment and the like, as well known in the art, to induce apoptosis to release fetal nucleosomes. Thus, a cervical mucus sample may be treated with an agent that induces apoptosis, whereby fetal nucleosomes are released. Regarding enrichment of circulating fetal DNA, reference is made to U.S. patent publication Nos. 20070243549 and 20100240054. The present invention is especially advantageous when applying the methods and systems to prenatal screening where only a small fraction of nucleosomes or DNA may be fetal in origin.

[0594] Prenatal screening according to the present invention may be for a disease including, but not limited to Trisomy 13, Trisomy 16, Trisomy 18, Klinefelter syndrome (47,

XXY), (47, XYY) and (47, XXX), Turner syndrome, Down syndrome (Trisomy 21), Cystic Fibrosis, Huntington's Disease, Beta Thalassaemia, Myotonic Dystrophy, Sickle Cell Anemia, Porphyria, Fragile-X-Syndrome, Robertsonian translocation, Angelman syndrome, DiGeorge syndrome and Wolf-Hirschhorn Syndrome.

[0595] Several further aspects of the invention relate to diagnosing, prognosing and/or treating defects associated with a wide range of genetic diseases which are further described on the website of the National Institutes of Health under the topic subsection Genetic Disorders (website at health.nih.gov/topic/Genetic Disorders).

Cancer and Cancer Drug Resistance Detection

[0596] In certain embodiments, the present invention may be used to detect genes and mutations associated with cancer. In certain embodiments, mutations associated with resistance are detected. The amplification of resistant tumor cells or appearance of resistant mutations in clonal populations of tumor cells may arise during treatment (see, e.g., Burger JA, et al., Clonal evolution in patients with chronic lymphocytic leukaemia developing resistance to BTK inhibition. Nat Commun. 2016 May 20;7:11589; Landau DA, et al., Mutations driving CLL and their evolution in progression and relapse. Nature. 2015 Oct 22;526(7574):525-30; Landau DA, et al., Clonal evolution in hematological malignancies and therapeutic implications. Leukemia. 2014 Jan;28(1):34-43; and Landau DA, et al., Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell. 2013 Feb 14;152(4):714-26). Accordingly, detecting such mutations requires highly sensitive assays and monitoring requires repeated biopsy. Repeated biopsies are inconvenient, invasive and costly. Resistant mutations can be difficult to detect in a blood sample or other noninvasively collected biological sample (e.g., blood, saliva, urine) using the prior methods known in the art. Resistant mutations may refer to mutations associated with resistance to a chemotherapy, targeted therapy, or immunotherapy.

[0597] In certain embodiments, mutations occur in individual cancers that may be used to detect cancer progression. In one embodiment, mutations related to T cell cytolytic activity against tumors have been characterized and may be detected by the present invention (see e.g., Rooney et al., Molecular and genetic properties of tumors associated with local immune cytolytic activity, Cell. 2015 January 15; 160(1-2): 48–61). Personalized therapies may be developed for a patient based on detection of these mutations (see e.g., WO2016100975A1). In certain embodiments, cancer specific mutations associated with cytolytic activity may be a

mutation in a gene selected from the group consisting of CASP8, B2M, PIK3CA, SMC1A, ARID5B, TET2, ALPK2, COL5A1, TP53, DNER, NCOR1, MORC4, CIC, IRF6, MYOCD, ANKLE1, CNKSR1, NF1, SOS1, ARID2, CUL4B, DDX3X, FUBP1, TCP11L2, HLA-A, B or C, CSNK2A1, MET, ASXL1, PD-L1, PD-L2, IDO1, IDO2, ALOX12B and ALOX15B, or copy number gain, excluding whole-chromosome events, impacting any of the following chromosomal bands: 6q16.1−q21, 6q22.31−q24.1, 6q25.1−q26, 7p11.2−q11.1, 8p23.1, 8p11.23−p11.21 (containing IDO1, IDO2), 9p24.2−p23 (containing PDL1, PDL2), 10p15.3, 10p15.1−p13, 11p14.1, 12p13.32−p13.2, 17p13.1 (containing ALOX12B, ALOX15B), and 22q11.1−q11.21.

[0598] In certain embodiments, the present invention is used to detect a cancer mutation (e.g., resistance mutation) during the course of a treatment and after treatment is completed. The sensitivity of the present invention may allow for noninvasive detection of clonal mutations arising during treatment and can be used to detect a recurrence in the disease.

[0599] In certain example embodiments, detection of microRNAs (miRNA) and/or miRNA signatures of differentially expressed miRNA, may used to detect or monitor progression of a cancer and/or detect drug resistance to a cancer therapy. As an example, Nadal et al. (Nature Scientific Reports, (2015) doi:10.1038/srep12464) describe mRNA signatures that may be used to detect non-small cell lung cancer (NSCLC).

[0600] In certain example embodiments, the presence of resistance mutations in clonal subpopulations of cells may be used in determining a treatment regimen. In other embodiments, personalized therapies for treating a patient may be administered based on common tumor mutations. In certain embodiments, common mutations arise in response to treatment and lead to drug resistance. In certain embodiments, the present invention may be used in monitoring patients for cells acquiring a mutation or amplification of cells harboring such drug resistant mutations.

[0601] Treatment with various chemotherapeutic agents, particularly with targeted therapies such as tyrosine kinase inhibitors, frequently leads to new mutations in the target molecules that resist the activity of the therapeutic. Multiple strategies to overcome this resistance are being evaluated, including development of second generation therapies that are not affected by these mutations and treatment with multiple agents including those that act downstream of the resistance mutation. In an exemplary embodiment, a common mutation to ibrutinib, a molecule targeting Bruton's Tyrosine Kinase (BTK) and used for CLL and certain

lymphomas, is a Cysteine to Serine change at position 481 (BTK/C481S). Erlotinib, which targets the tyrosine kinase domain of the Epidermal Growth Factor Receptor (EGFR), is commonly used in the treatment of lung cancer and resistant tumors invariably develop following therapy. A common mutation found in resistant clones is a threonine to methionine mutation at position 790.

[0602] Non-silent mutations shared between populations of cancer patients and common resistant mutations that may be detected with the present invention are known in the art (see e.g., WO/2016/187508). In certain embodiments, drug resistance mutations may be induced by treatment with ibrutinib, erlotinib, imatinib, gefitinib, crizotinib, trastuzumab, vemurafenib, RAF/MEK, check point blockade therapy, or antiestrogen therapy. In certain embodiments, the cancer specific mutations are present in one or more genes encoding a protein selected from the group consisting of Programmed Death-Ligand 1 (PD-L1), androgen receptor (AR), Bruton's Tyrosine Kinase (BTK), Epidermal Growth Factor Receptor (EGFR), BCR-Abl, c-kit, PIK3CA, HER2, EML4-ALK, KRAS, ALK, ROS1, AKT1, BRAF, MEK1, MEK2, NRAS, RAC1, and ESR1.

[0603] Immune checkpoints are inhibitory pathways that slow down or stop immune reactions and prevent excessive tissue damage from uncontrolled activity of immune cells. In certain embodiments, the immune checkpoint targeted is the programmed death-1 (PD-1 or CD279) gene (PDCD1). In other embodiments, the immune checkpoint targeted is cytotoxic T-lymphocyte-associated antigen (CTLA-4). In additional embodiments, the immune checkpoint targeted is another member of the CD28 and CTLA4 Ig superfamily such as BTLA, LAG3, ICOS, PDL1 or KIR. In further additional embodiments, the immune checkpoint targeted is a member of the TNFR superfamily such as CD40, OX40, CD137, GITR, CD27 or TIM-3.

[0604] Recently, gene expression in tumors and their microenvironments have been characterized at the single cell level (see e.g., Tirosh, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single cell RNA-seq. Science 352, 189-196, doi:10.1126/science.aad0501 (2016)); Tirosh et al., Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature. 2016 Nov 10;539(7628):309-313. doi: 10.1038/nature20123. Epub 2016 Nov 2; and International patent publication serial number WO 2017004153 A1). In certain embodiments, gene signatures may be detected using the present invention. In one embodiment complement genes are monitored or detected in a

tumor microenvironment. In one embodiment MITF and AXL programs are monitored or detected. In one embodiment, a tumor specific stem cell or progenitor cell signature is detected. Such signatures indicate the state of an immune response and state of a tumor. In certain embodiments, the state of a tumor in terms of proliferation, resistance to treatment and abundance of immune cells may be detected.

[0605] Thus, in certain embodiments, the invention provides low-cost, rapid, multiplexed cancer detection panels for circulating DNA, such as tumor DNA, particularly for monitoring disease recurrence or the development of common resistance mutations.

Immunotherapy Applications

[0606] The embodiments disclosed herein can also be useful in further immunotherapy contexts. For instance, in some embodiments methods of diagnosing, prognosing and/or staging an immune response in a subject comprise detecting a first level of expression, activity and/or function of one or more biomarker and comparing the detected level to a control level wherein a difference in the detected level and the control level indicates that the presence of an immune response in the subject.

[0607] In certain embodiments, the present invention may be used to determine dysfunction or activation of tumor infiltrating lymphocytes (TIL). TILs may be isolated from a tumor using known methods. The TILs may be analyzed to determine whether they should be used in adoptive cell transfer therapies. Additionally, chimeric antigen receptor T cells (CAR T cells) may be analyzed for a signature of dysfunction or activation before administering them to a subject. Exemplary signatures for dysfunctional and activated T cell have been described (see e.g., Singer M, et al., A Distinct Gene Module for Dysfunction Uncoupled from Activation in Tumor-Infiltrating T Cells. Cell. 2016 Sep 8;166(6):1500-1511.e9. doi: 10.1016/j.cell.2016.08.052).

[0608] In some embodiments, C2c2 is used to evaluate that state of immune cells, such as T cells (e.g., CD8+ and/or CD4+ T cells). In particular, T cell activation and/or dysfunction can be determined, e.g., based on genes or gene signatures associated with one or more of the T cell states. In this way, c2c2 can be used to determine the presence of one or more subpopulations of T cells.

[0609] In some embodiments, C2c2 can be used in a diagnostic assay or may be used as a method of determining whether a patient is suitable for administering an immunotherapy or another type of therapy. For example, detection of gene or biomarker signatures may be

performed via c2c2 to determine whether a patient is responding to a given treatment or, if the patient is not responding, if this may be due to T cell dysfunction. Such detection is informative regarding the types of therapy the patient is best suited to receive. For example, whether the patient should receive immunotherapy.

[0610] In some embodiments, the systems and assays disclosed herein may allow clinicians to identify whether a patient's response to a therapy (e.g., an adoptive cell transfer (ACT) therapy) is due to cell dysfunction, and if it is, levels of up-regulation and down-regulation across the biomarker signature will allow problems to be addressed. For example, if a patient receiving ACT is non-responsive, the cells administered as part of the ACT may be assayed by an assay disclosed herein to determine the relative level of expression of a biomarker signature known to be associated with cell activation and/or dysfunction states. If a particular inhibitory receptor or molecule is up-regulated in the ACT cells, the patient may be treated with an inhibitor of that receptor or molecule. If a particular stimulatory receptor or molecule is down-regulated in the ACT cells, the patient may be treated with an agonist of that receptor or molecule.

[0611] In certain example embodiments, the systems, methods, and devices described herein may be used to screen gene signatures that identify a particular cell type, cell phenotype, or cell state. Likewise, through the use of such methods as compressed sensing, the embodiments disclosed herein may be used to detect transcriptomes. Gene expression data are highly structured, such that the expression level of some genes is predictive of the expression level of others. Knowledge that gene expression data are highly structured allows for the assumption that the number of degrees of freedom in the system are small, which allows for assuming that the basis for computation of the relative gene abundances is sparse. It is possible to make several biologically motivated assumptions that allow Applicants to recover the nonlinear interaction terms while under-sampling without having any specific knowledge of which genes are likely to interact. In particular, if Applicants assume that genetic interactions are low rank, sparse, or a combination of these, then the true number of degrees of freedom is small relative to the complete combinatorial expansion, which enables Applicants to infer the full nonlinear landscape with a relatively small number of perturbations. Working around these assumptions, analytical theories of matrix completion and compressed sensing may be used to design under-sampled combinatorial perturbation experiments. In addition, a kernel-learning framework may be used to employ under-sampling by building predictive functions of

combinatorial perturbations without directly learning any individual interaction coefficient Compresses sensing provides a way to identify the minimal number of target transcripts to be detected in order obtain a comprehensive gene-expression profile. Methods for compressed sensing are disclosed in PCT/US2016/059230 "Systems and Methods for Determining Relative Abundances of Biomolecules" filed October 27, 2016, which is incorporated herein by reference. Having used methods like compressed sensing to identify a minimal transcript target set, a set of corresponding guide RNAs may then be designed to detect said transcripts. Accordingly, in certain example embodiments, a method for obtaining a gene-expression profile of cell comprises detecting, using the embodiments disclosed, herein a minimal transcript set that provides a gene-expression profile of a cell or population of cells.

Detecting Nucleic Acid Tagged Molecules

[0612] In some embodiments, the detection compositions of the present invention described herein may be used to detect nucleic acid identifiers. Nucleic acid identifiers are non-coding nucleic acids that may be used to identify a particular article. Example nucleic acid identifiers, such as DNA watermarks, are described in Heider and Barnekow. "DNA watermarks: A proof of concept" BMC Molecular Biology 9:40 (2008). The nucleic acid identifiers may also be a nucleic acid barcode. A nucleic-acid based barcode is a short sequence of nucleotides (for example, DNA, RNA, or combinations thereof) that is used as an identifier for an associated molecule, such as a target molecule and/or target nucleic acid. A nucleic acid barcode can have a length of at least, for example, 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, 35, 40, 45, 50, 60, 70, 80, 90, or 100 nucleotides, and can be in single- or double-stranded form. One or more nucleic acid barcodes can be attached, or "tagged," to a target molecule and/or target nucleic acid. This attachment can be direct (for example, covalent or non-covalent binding of the barcode to the target molecule) or indirect (for example, via an additional molecule, for example, a specific binding agent, such as an antibody (or other protein) or a barcode receiving adaptor (or other nucleic acid molecule). Target molecule and/or target nucleic acids can be labeled with multiple nucleic acid barcodes in combinatorial fashion, such as a nucleic acid barcode concatemer. Typically, a nucleic acid barcode is used to identify target molecules and/or target nucleic acids as being from a particular compartment (for example a discrete volume), having a particular physical property (for example, affinity, length, sequence, etc.), or having been subject to certain treatment conditions. Target molecule and/or target nucleic acid can be associated with

multiple nucleic acid barcodes to provide information about all of these features (and more). Methods of generating nucleic acid-barcodes are disclosed, for example, in International Patent Application Publication No. WO/2014/047561.

Cell Labeling

[0613] The programmable pattern recognition compositions and/or detection compositions of the present invention can be used, for example, to label a cell. As previously described in relation to e.g., methods of detecting target polynucleotides, when a detection composition of the present invention is activated by binding a target polynucleotide a detectable signal or product is produced. In some embodiments, the detectable signal or product is such that it allows a cell to which the system is delivered to and activated in to be "labeled" via the detectable signal or product. For example, if the detectable signal is an optical signal (e.g., fluorescence) produced from a protein, then the cell is effectively labeled with fluorescence that can be tracked, imaged, and used for e.g., fluorescence-based sorting or separation techniques. Other signals and products that can be used as labels are described in greater detail elsewhere herein and will be appreciated in view of the description provided herein. In this way cells containing a target polynucleotide can be effectively labeled. Labeling via a method described herein can occur in vivo, ex vivo, in vitro, or in situ. Such methods can be applied to various cell detection, imaging, diagnostic, prognostic, screening, functionality, cell isolation and separation, and other assays and techniques where cell labeling is traditionally employed. Such labeling approaches can be helpful for cell type and cell state evaluation, particularly at the single cell level.

[0614] Described in certain example embodiments herein are methods of labeling cells comprising introducing a pattern recognition composition of the present invention or a detection composition comprising an engineered protein of the present invention as described in greater detail elsewhere herein into a population of cells, wherein the pattern recognition compositions and/or a guide molecule employed is configured to detect one or more target nucleic acids (e.g., DNA or RNA) or polypeptides associated with a particular cell type or cell state; and activating the pattern recognition composition via recognition and/or binding of the protein composition to the one or more target transcripts such that a detection construct is modified by the activated pattern recognition composition such that a detectable product and/or signal is generated, thereby labeling cells within the cell population expressing the one or more target transcripts.

[0615] In some embodiments, a substrate the pattern recognition composition or component thereof (e.g., an effector domain) of the present invention is tethered or anchored to a structure within the cell. Exemplary cell structures to which the substrate can be anchored is the cell or nuclear membrane, mitochondria membrane, endoplasmic reticulum, lysosome, Golgi apparatus, microtubules or other cytoskeleton components, and/or the like. In some embodiments the substrate is coupled to a signal producing molecule or product producing molecule that is inactive until released from the peptidase substrate or is otherwise modified by activity of the pattern recognition composition or effector thereof on the substrate upon recognition of a target molecule (e.g., a target polypeptide and/or polynucleotide) and/or target molecular pattern , which can optionally be associated or contained within or on the surface of a target cell or other molecule (e.g., exosomes or the like).

Therapeutic Delivery and/or Effector Function

[0616] Similar to embodiments of cell labeling, the programmable pattern recognition compositions of the present invention can be configured for in vivo effector function and/or delivery of a cargo, such as a therapeutic molecule. A substrate for the pattern recognition composition or component thereof (e.g., an effector domain) can be tethered or otherwise anchored to a cellular structure. Exemplary cell structures to which the substrate can be anchored is the cell or nuclear membrane, mitochondria membrane, endoplasmic reticulum, lysosome, Golgi apparatus, microtubules or other cytoskeleton components, and/or the like. The substrate can also be coupled to (either directly or via a linker) an effector molecule (e.g., a Cre recombinase, CRISPR-Cas system, or other effector molecule) and/or to a cargo (e.g., a therapeutic molecule). In some embodiments, the effector molecule or other molecule (e.g., a therapeutic molecule), is inactive while coupled to the substrate. When a target molecule (e.g., a target polynucleotide or target polypeptide) and/or target molecular pattern is present in cell that also contains the programmable pattern recognition composition, one or more effector domains of the programmable pattern recognition composition are activated upon recognition and/or binding of the target molecule (e.g., a target polynucleotide or target polypeptide) and/or target molecular pattern that can be associated with a target cell, polypeptide and/or RNA and acts to cleave the tethered substrate. Cleaving of the substrate releases and/or otherwise activates the effector or cargo (e.g., therapeutic molecule) that was linked to the peptidase substrate. Target molecules (e.g., a target polynucleotide or target polypeptide) and/or molecules having a target molecular pattern can be endogenous to the cell that expresses

the programmable pattern recognition composition and/or tethered effector and/or cargo (or therapeutic) complex. In other embodiments, target molecules (e.g., a target polynucleotide or target polypeptide) and/or molecules having a target molecular pattern are exogenous to the cell. Exogenous Target molecules (e.g., a target polynucleotide or target polypeptide) and/or molecules having a target molecular pattern can provide an additional measure of temporal and/or spatial control of effector function and/or therapeutic delivery. Exemplary effectors that can be included in these embodiments are described in greater detail elsewhere herein and will be appreciated by those of ordinary skill in the art in view of the description herein.

[0617] In some embodiments, a method of in vivo effector activation or delivery includes introducing a programmable pattern recognition composition or system of the present invention into a cell comprising a substrate of the programmable pattern recognition composition or component thereof (e.g., a STAND NTPase or other effector domain), wherein the substrate of the programmable pattern recognition composition is optionally tethered to a cellular structure and wherein the substrate the programmable pattern recognition composition is coupled to an effector. In some embodiments, the effector is capable of producing a detectable signal when activated. In some embodiments, the cargo is a therapeutic molecule or prodrug, genetic modifying molecule, or any combination thereof. In some embodiments, the effector is inactive when coupled to an uncleaved substrate. In some embodiments, the effector is inactive when coupled to a cleaved substrate portion (and thus is active when coupled to an uncleaved substrate). In some embodiments, the method further comprises cleaving the substrate in response to target molecule recognition and/or binding and activation of an effector domain of the programmable pattern recognition composition of the present invention. In some embodiments, the target molecule is endogenous to the cell or is exogenous to the cell. In some embodiments, the substrate is tethered to a cell membrane or a nuclear membrane.

[0618] In some embodiments, the method of cargo delivery comprises delivering to a cell (a) an engineered protein of the present invention; (b) a cargo; (c) a detection composition), or any combination thereof, wherein the engineered protein comprises the cargo or wherein the cargo comprises the target polypeptide, wherein the cell optionally comprises the target polypeptide, and

wherein activation of the effector domain by binding of the target polypeptide to the recognition domain results in delivery of the cargo and optionally activation of the detection construct thereby monitoring cargo delivery.

[0619] In some embodiments, the cargo is a prodrug or proenzyme. In certain example embodiments, the target molecule (e.g., a target polypeptide or polynucleotide) modification is cleavage of the target molecule (e.g., a target polypeptide or polynucleotide). In certain example embodiments, the one or more target molecule (e.g., a target polypeptide or polynucleotide) are proenzymes, proproteins, and/or prodrugs, and the modification results in conversion of the proenzyme into an active enzyme, active protein, or active prodrug, respectively. Thus, in some embodiments, the engineered proteins of the present invention can be used to deliver and/or control delivery of a drug or enzyme from its prodrug form.

Targeted Cell Death

[0620] In some embodiments, the programmable pattern recognition compositions are used for targeted cell death. In such embodiments, the programmable pattern recognition compositions are configured to recognize and/or bind a target molecule (e.g., a target polypeptide or target polynucleotide) and/or target molecular pattern of target cells to be killed. The programmable pattern recognition composition can contain an effector domain (e.g., a protease or nuclease) that, when activated upon target molecule (e.g., target polynucleotide or target polypeptide) and/or target molecular pattern recognition and/or binding, induces cell death. In some embodiments, the method includes delivering an engineered protein of the present invention, a polynucleotide of encoding an engineered protein of the present invention, a vector or vector system of the present invention, a formulation thereof, or any combination thereof to the target molecule and/or cell, wherein the target molecule and/or cell is or comprises a target molecule (e.g., a target polypeptide, target polynucleotide) and/or target molecular pattern (e.g., a PAMP); and activating an effector domain of the engineered protein by allowing binding of the target molecule (e.g., target polypeptide or target polynucleotide) and/or target molecular pattern to the recognition domain thereby activating the effector domain via the effector activation domain, wherein effector domain activity induces cell death (e.g., via apoptosis or other cell death mechanism) of the target cell.. Without being bound by theory, an activated nuclease or protease effector domain can induce apoptosis in the target cells, leading to cell death. In some embodiments, the target cells are prokaryotic cells. In some embodiments, the target cells are bacterial cells. Thus, in some embodiments, the engineered proteins of the present invention are antibiotic compositions. In some embodiments the target cells are eukaryotic cells. In some embodiments, the target cells are tumor or cancer cells. Thus, in some embodiments, the engineered proteins of the present invention are anti-cancer or anti-

tumor compositions. In some embodiments, the target cells are diseased, infected or aberrantly functioning cells.

Generating Phage-Resistant Cells

[0621] In some embodiments, the programmable pattern recognition compositions are used to generate or engineer phage-resistant cells. In such embodiments, the programmable pattern recognition compositions are configured to recognize a target molecule (e.g., a target polypeptide or target polynucleotide) and/or a target molecular pattern (e.g., a PAMP) of one or more phages. Further, the programmable pattern recognition compositions are configured to inactivate and/or destroy or eliminate phages comprising the target molecule (e.g., a target polypeptide or target polynucleotide) and/or a target molecular pattern (e.g., a PAMP) that is recognized by the programmable pattern recognition composition. Such a configuration can include a peptidase effector and/or a nuclease effector domain. In some embodiments, the programmable pattern recognition compositions are introduced into and/or expressed by cells, such as bacterial cells. When invaded by a phage, the programmable pattern recognition compositions recognizes a target molecule (e.g., a target polypeptide or target polynucleotide) and/or a target molecular pattern (e.g., a PAMP) thaton the invading phage and is activated by such recognition. Activation results in effector domain function of the programmable pattern recognition composition, and inhibition and/or destruction of the invading phage. Such phage resistant cells can have various industrial applications, such as in industrial bioreactors.

Microbiome Engineering

[0622] In certain example embodiments, the programmable pattern recognition compositions are used to engineer a microbiome, and more particularly the structure of a microbiome. As used herien, the term “microbiome structure” refers to the profile (e.g., relative or absolute numbers) of different microbes present in a microbiome. Engineering a microbiome can include modulating the microbiome structure to a desired structure. The programmable pattern recognition compositions can be configured to recognize target molecule (e.g., a target polypeptide or target polynucleotide) and/or a target molecular pattern (e.g., a PAMP) on a specific microbe type (e.g., species, subspecies, clade and/or the like) to be increased in number, decreased in number, modified, and/or eliminated, so as to modify and engineer the microbiome structure. In some embodiments, the modification results in a phage-resistant bacteria, which can result in an increase in the number of the phage-resistant bacteria within the microbiome as it has an evolutionary advantage of being phage-resistant. In some

embodiments, the modification results in apoptosis of the target cell(s) thus removing it from the microbiome. In some embodiments, the modification in the target cells provides a competitive advantage so as to make the target cell type the dominate cell type within the microbiome population. In some embodiments, the targeted microbe(s) are increased in relative or absolute amount within the microbiome after modification by the programmable pattern recognition composition of the present invention. In some embodiments, the targeted microbe(s) are decreased in relative or absolute amount within the microbiome after modification by the programmable pattern recognition composition of the present invention.

[0623] In some embodiments a method of microbiome engineering includes delivering a programmable pattern recognition protein or composition to one or more microorganisms of microbiome to which the target molecule (e.g., a target polypeptide or target polynucleotide) and/or a target molecular pattern (e.g., a PAMP) is specific to and wherein an effector of the engineered protein modifies, stimulates, or kills the one or more microorganisms to which the protein is delivered, thus resulting in a change (e.g., increase or decrease) in the absolute or relative number of the microorganism within the microbiome, thereby modifying (or engineering) the microbiome structure.

[0624] Further embodiments are illustrated in the following Examples which are given for illustrative purposes only and are not intended to limit the scope of the invention.

EXAMPLES

Example 1 – STAND NTPases

[0625] All organisms have evolved specialized immune proteins, including pattern recognition receptors consisting of nucleotide-binding oligomerization domain-like receptors (NLRs) of the STAND superfamily ubiquitous in eukaryotes. NLRs recognize conserved pathogen-associated molecular patterns, leading to activation of an effector domain and an inflammatory or apoptotic response. The roles of NLRs in eukaryotic immunity are well established, but it is unknown whether prokaryotes use similar defense mechanisms. Here Applicant shows that antiviral STAND (Avs) homologs in bacteria and archaea are pattern recognition receptors that detect conserved viral proteins and activate diverse N-terminal effectors, including DNA endonucleases. See e.g., **Fig. 26**. This work reveals remarkable similarity between the defense strategies of prokaryotes and eukaryotes and extends the paradigm of pattern recognition of pathogen-specific proteins across all domains of life.

[0626] Applicants recently identified a group of STAND NTPases, dubbed Avs (antiviral STAND) (4), that are often encoded next to restriction modification and other defense systems (**Fig. 1A-1B**) and protect bacteria from dsDNA phages (**Fig. 2**). Here, Applicants investigate the mechanism of Avs proteins, which consist of diverse N-terminal effector domains, a conserved central STAND NTPase core, and C-terminal tetratricopeptide repeat (TPR) domains. Applicants demonstrate that Avs systems specifically bind to and are activated by the conserved large terminase subunit and portal proteins of diverse tailed phages; that some Avs systems function as non-specific DNA endonucleases; and that some phages encode Avs inhibitors. Avs systems are activated by two conserved phage proteins.

Avs systems are activated by two conserved phage proteins

[0627] Although the domain architectures of Avs proteins resemble those of eukaryotic NLRs (**Fig. 3A-3B**), it is unclear whether they function via similar molecular mechanisms. Applicant identified four distinct families of Avs proteins (Avs1-4), each of which contains highly divergent tetratrico-peptide repeat (TPR)–like sensor domains (**Fig. 4A-4B**) and selected two representatives for further characterization: SeAvs3 from Salmonella enterica NCTC13175 and EcAvs4 from Escherichia coli NCTC11132, both of which provide robust protection against the T7-like coliphage PhiV-1 (**Fig. 2**). Applicants first asked how phage infection leads to Avs activation and whether a specific phage-encoded trigger exists for these defense systems. Applicants cloned fragments comprising the whole PhiV-1 phage genome into expression plasmids and transformed the resulting phage library into E. coli containing either Avs proteins or empty controls (**Fig. 3C and Table 10**). Applicant hypothesized that co-expression of putative Avs triggers might lead to cell death and depletion of the respective phage genes from the pool, and performed deep sequencing to detect enrichment or depletion of phage genes. Four phage genes were generally toxic to all cells; however, two genes were depleted only in the presence of Avs systems, namely the large terminase subunit (gp19) when co-expressed with SeAvs3, and the portal protein (gp8) when expressed with EcAvs4 (**Fig. 3D** and **Fig. 5A-5B**). By Southern blot, Applicant observed that Avs3 and Avs4-mediated depletion of phage DNA during infection was abolished in gp8 and gp19 knockout phage strains, respectively (**Fig. 3E** and **Fig. 6A-6E**), indicating that gp8 and gp19 are both necessary and sufficient for Avs activation.

[0628] To validate these findings, Applicant transformed plasmids expressing gp8 or gp19 into E. coli harboring SeAvs3 or EcAvs4 and measured cell growth. Consistent with

Applicant's previous results, Applicant observed toxicity following co-expression of SeAvs3 and gp19, as well as co-expression of EcAvs4 and gp8, but not with the reciprocal pairs (**Fig. 3E**). This toxicity depended on the predicted nuclease activity of both Avs systems, and importantly, was not due to any intrinsic features of the natural phage gene sequence, as recoded gene sequences also led to cell death (**Fig. 3F**). Further, the enzymatic activity of the phage terminase, which contains triphosphate (ATPase) and nuclease domains unrelated to those of Avs proteins, was not required for SeAvs3- mediated toxicity (**Fig. 3F**).

Table 10. PhiV-1 phage fragments

						Abundance relative to E. coli NovaBlue control			
#	Gene(s)	DNA	bp	Toxic	Proteins(s)	Empty vector	Empty vector	SeAvs3	EcAvs4
1	gp0.3	ATGATTCACACTAAAGAGCAAGCT AACGTTTTCTATGTACTGGTAAGC GCATTCCGTTCTAACCTGCACCAC TTGGATAACATTAGCCGTCATCGT CGCTTAGTTAACAGCTTACGCAAT GCACCTCTGGAATACGGTAGCGTT GAGTCTACTGATTTAACTGGCTGT TATCATGAGGCTGGCATGGAGTCA CCCTCTGAAGAGCAAACCATCCGA GTACGCTGCACGACCAAGGCGCA AGCTATAAGTGTCGCACGGTTGGC TTGTAACGACTATGAGCAAGACTG TGTACTGGTCTATAAGTCACAGAC TCACAGCGCTGCACTAATCTTTGC TGGTGGCTTGGACGGTTGGAAGAC TGAGCGTATCGGGTCGTTCCAGAT GGTCGCACCTGAAGTTGAACTGAC AGGCAACTACACGGTCGATGAGTT CGGACGTGGCTGGAAGGTAATCT AA (SEQ ID NO: 19)	456	+	Anti-restriction protein Ocr	--	--	--	--
2	0.6A	ATGGTCATGGCACTCTTTGTAATC ACGATGTACGCTCTGATTGGGGCG TACTTCCTAAGAGACTTCCGCAAG GGTCTCAAGGTACACAAGGCTAA CTTCTCGCACTTGAAGTATGGCTT TCTACCACGATTCACAGTCCGTGA ACGTAACGGGCGCTTTAAGGCTAA CAAGGTTGGTATCTTCTATGTTGC AACGCACTGA (SEQ ID NO: 20)	201	–	Hypothetical	0.28	0.26	0.24	0.18
3	gp0.7	ATGAACTACACCGACATTCAGGCA CGTCTGGCAATCATCAAGTCACTC CCGATTAGCGAACTGGACAAGCG CCAGCCTCTTCTGGTTGCACTGGC TGCTGACATCGTGAATGGTGAGAC AAGCGATGGCAACGACACAGACG GGTCGCAGGGTCTTGAGTATCAGG ACTGGTGGCACACCTTGGGTGCAC TCATGCGTGATGCAGGTTTCCGCA TGCTGGGTAATGGTCACTTCAGTG CGGCCTACTCGCACGAAATGTTAC CGGGTCGAGTCATCAAGGTTGGCT TTAAGAAAGAGGATTCAGGCGCT GCATACACTGCATTCTGTAGGATG CACCAAGGCCGTGCTGGTATCCCT AACGTCTATCACGTAGCACGTCAT GCAGGTTGCTATACGGTCGTACTG GACCGCTTAAAAAGCTGTGACCGC TACGCTAACGATGTTCACGCAAAG TATGCAAGCGCTGCGCAGGAGTTC ATTGAGTGCTCTGGAGGTGATGCC GATTGGTACGCACAGTATGTTAGC	1104	+	Protein kinase and inhibitor of host RNA polymerase	--	--	--	--

		AAGGAGTTCATTGAGACGTGCAA GATGATTCGTGAGTTCTTCCACGG TATTGCATCTTTCGACATGCACAG TGGGAACATTATGTTCGATGACAA CGATGTCCCGTACATCACTGACCC GGTTAGTTTCTCACATGACCGTGA ACGTGAGGATGGTTTCCCTCTGGA CCCTGATGCGTTGATTGCCGAGGT CGAGGCGATAGCTAACGAGCGCA TCATTGAACGGTGCCGCAACCGCA AGGCCAAATGTGACCCTAACGGA ACCTTTCAGGTTAACCGCAAGGCC GCAATGAAACGCCGTAAGCGCAA CCGCAAGTTACGTGCTAAACATGC AGCGGAAGGAATGCACCAGTTTG TCAAAGAGCGCCTTAAGCGTAAA GCTATTGACCGAAATGAACGCCGT GCTGAGATGCTCATGGGGTCTGCA TGGCATTACGCATGGTTGCATAAT GGTAACTCAACGTTCCGCAAGGTT GAACAAGTGAACGGCCTCAAGTG GCAACTAGGGGACCGCCTAGCAA TCCAAGCGGGTCTACCTCTGAACA TTGACAAGTTACTTGACGCACAAC TGCAAGGCTGA (SEQ ID NO: 21)							
4	gp1	ATGACTAACGTAATCAACGCACCA AAGAACGACTTCTCTGACATCGCT AACGCAATCCAGCCGTACAACATT CTGGCTGACCATTACGGCGCACAG CTTGCAGCTACGCAGTTACAGCTT GAGCACGAAGCACACACCGAGGG CGAGAAGCGTTTTCTTAAGGCAAT GGAACGTCAAATCAAAGCGGGTG AGTTCGGTGATAACGCAGTGGCTA AACCGCTCCTGTCATCTCTCGCAC CGAAGTTCATCGAGGCGTGGAAC ACTTGGTTCACCGAGGTGGAAGCG AAGCGTGGCAAGCGTCCAGTAGC CTACAATCTGGTACAGAAGGTTGC ACCTGAAGCGGCCGCTTTCATCAC ACTGAAGGTCACACTCGCTTGCTT GACGAAAGAAGAGTTCACCAACC TGCAATCGGTCGCCACTAAGATTG GTCGCAGTATTGAGGATGAACTGC GTTTTGGTCGCATCCGTGACGAAG AAGCGAAGCACTTTAAGAACCAC GTACAGGAGGCACTGAACAAGCG CGTTGGTATCGTCTACAAGAAAGC ATTCATGCAAGCCGTAGAAGGCA AGATGTTAGACGCTGGTCAACTGC AAACCAAGTGGACTACATGGACA CCTGAAGAATCAATTCACGTTGGT GTGCGCATGTTGGAACTGCTCATC GGGTCTACTGGTCTCGTTGAGTTA CACCGACCGTTTGCTGGTAACGTT GAGAAGGATGGCGAGTACATCCA GTTAACAGAACAATATGTAGACCT GCTATCCAAACGTGCTGGCGCACT GGCTGCAATCGCTCCGATGTATCA GCCTTGCGTGGTCCCTCCGAAGCC TTGGACTTCTCCGGTTGGTGGTGG CTATTGGGCTGCTGGTCGCAAGCC TCTCTCTCTGGTTCGCACTGGTAG CAAGAAGGGACTTGAGCGTTACA ACGATGTGTATATGCCTGAAGTCT ACAAGGCGGTGAACATCGCACAG AATACTCCTTGGAAGATTAACAAG AAGGTTCTGGCAGTGGTCAACGA GATTGTGAACTGGAAGCATTGCCC GGTTGAAGATGTTCCTGCTCTGGA ACGTGGTGAACTGCCTGTCAAACC TGATGACATTGACACCAACGAGG	2652	–	RNA polymerase	0.98	0.97	0.63	0.67

		CATCGCTCAAAGCGTGGAAGAAA GCGGCCTCCGCAATCTACCGCAAA GAGAAGGCACGAGTGTCTCGTCGT ATGAGCATGGAGTTCATGTTAGGT CAGGCCAATAAGTTCGCACAGTTT AAAGCTATCTGGTTCCCGATGAAC ATGGACTGGCGCGGCCGAGTGTAT GCGGTCCCAATGTTTAACCCGCAA GGCAACGACATGACCAAAGGCTT GCTGACGCTCGCTAAAGGTAAACC AATCGGTGTCGATGGGTACTACTG GCTGAAGATTCACGGTGCTAACAC TGCTGGTGTGGACAAGGTGGACTT CGCTGAACGCATCAAGTTCATTGA TGACAACCACGAGAACATCATGA GCGTTGCGGCTGACCCGATTGCTA ACACTTGGTGGGCTGAACAAGACT CTCCGTTCTGCTTCCTCGCGTTCTG CTTTGAGTATGCGGGAGTCCAGCA TCACGGGATGAACTACAACTGCTC TTTACCGCTGGCGTTCGATGGCTC TTGCTCTGGTATTCAGCACTTCTCC GCAATGCTCCGCGATGAGGTCGGT GGTCGTGCGGTTAACCTGCTTCCA AGCAAAGAGGTACAAGACATCTA CCGCATCGTTGCTGAACGTGTGAA CGAGATTATGCAACAGGACGTAA TCAACGGAACTGATAACGAAGTG GAAACTGTTATCAACAAGGACACT GGAGAAATCACTGAGAAGCTCAA GCTGGGTACGAAGGAACTGGCGG GTCAATGGTTGGCCTACGGTGTGA CTCGCAAGGTCACTAAGCGTTCGG TCATGACGTTGGCTTACGGCTCCA AAGAGTATGGCTTCCGTGACCAAG TTCTTGAGGACACAATTCAGCCAG CTATCGACGATGGCAAGGGCTTGA TGTTCACTCAACCGAACCAAGCGG CTGGCTACATGGCTAAACTGATTT GGAATGCTGTAACGGTCACAGTTG TCGCAGCGGTTGAAGCGATGAACT GGTTGAAGTCTGCTGCTAAACTTC TGGCTGCTGAAGTCAAGGACAAG AAGACCAAAGAGGTTCTCCGCAA GCGTTGCGCAGTACATTGGGTAAC TCCTGATGGCTTCCCTGTGTGGCA GGAATACAAGAAGCCAGTACAGA CCCGCTTGAATCTGATGTTCTTAG GCCAGATTCGTCTACAGCCTACCG TGAACACCAACAAGGATAGCGGG ATTGATGCTCGTAAACAGGAGTCA GGCATCGCACCTAACTTTGTCCAC TCAATGGATGGAAGCCACCTGCGC ATGACCGTGGTTCGCTCCAATGAG GTCTACGGTGTGGAATCGTTCGCA CTGATTCACGACTCCTTTGGTACT ATCCCTGCTGATGCTGGCAACCTG TTTAAAGCAGTGCGTGAAACGATG GTCAACACTTACGAAGAGAACGA TGTACTTGCAGACTTCTATGAGCA ATTTGCTGACCAGTTGCACGAATC ACAGCTTGACAAGATGCCTGAGAT GCCAGCTAAAGGTTCTCTGGACCT GCAAGAAATCCTTAAGTCAGACTT TGCGTTTGCCTAA (SEQ ID NO: 22)							
5	gp1.05	ATGTTTACTTTCATCACTTCAGTA GGCGCTCTCGTATCAACAATTCTT GCAGCAACATACTATAATCGTTCA AAGGTATTTCTCACTAAGCTAATC AAAGAAGAGAAGAACTCTCATGC ACTCAAAGAGAGACTTGAAGATG GATGGCGAGAGAAGCAGGAACTA	558	–	Hypothetical	0.18	0.18	0.27	0.22

		GCTGGTCGAATCCGTACCTTAGAG TTCGAACAACAATACGACATGTCG AACCAACGAGAACTCATTGAGCG TCTACGCGCTGAGGCAGTGGAGTA CCGCAAGGAACTTAAAGAGGCAA AGAGTACCATCCGTGAGTACACTC ACAGTGACATCACGTCTATCAAGT TTACCGCTCACATCCCGGATGCTG GGTGGAAGCGCACTGAGTTCAAG CTGGGTATAGGGTCTTGTGGCAAA GAGGTTACATGTTTGCACTGGAAA GAACAGTCGGACCGCTATCTGCTG ACACAGACGTGCACTGACGGTGA GCGCAAAGAGTTCACCTATTATAA GAAAGATGTAGCTGGTCGCATTGA GATTGCACGTAAGGTTGCAGACA ATACGACTCACTAA (SEQ ID NO: 23)							
6	gp1.1	ATGCGTACCAACTTTGAACCAACC ACTAAGCGCAACAACGTAATCAA CGAACACGGTGCTGAATGGCAGG AACGTAAAGACCGCATGAAGAAA CGCCACAAGACACAACGTGGCAA CGCTGAGAAACGTAACTGGAAGG AGGCACTCTAA (SEQ ID NO: 24)	150	–	Hypothetical	0.29	0.29	0.40	0.30
7	gp1.2	ATGGGTCGCTTATATAGTGGTAAC TTAATGGCGTTAAAGGATGTCATC GAGAGACTTAAAGAAGACCACGA TGTCAGTGTTGTAACTGAGACTTT CCGCTATGAGAACTACGCACAGC GCCGATGGGTCTCCGGTGAGACTC TCCGCATTGTTCTACGCAATGGCC TCGTACTGACATCCAAGTGTTTCG AACAGTCAGACGAAGATGTCCGG TGCAATGCACAGACTGAGTGGCTC CGCAAGGTACACGCAGACCTGAA GCACTGGAAGTAA (SEQ ID NO: 25)	273	–	Inhibitor of host dGTP triphospho hydrolase	0.49	0.46	0.76	0.6
8	gp1.3	ATGGAAATCACATTCAAGACTAAC CCTTTTAAAGCTGTATCCTACGTT GAGTCTGCCATTAAGAAAGCACTC GACAAAGCTGGCTACCTGATTGCT GACACTAAGCATGATGGTGTACGT GGTAACATCGTGGTTGACAATACG GCTGACTGCCTGTGGCTATCCCGT GTCTCCAAGACCATCCCCGCGCTT GAGCACCTCAATGGGTTCGACCAA CGCTGGGCTAATGTTCTCAAAGAT GACCGCTGGATTTTCCCTGATGGC TTCATGCTGGATGGTGAACTGATG GTCAAAGGCGTAGACTTCAACAC AGGGTCTGGCCTACTGCGTACCAA GTGGCTCAAGAAGACAAACTGTG GATTTGATATTGGTGGAGTAGAGC CTCTCAAGGGGAGTAAAGTAGAC TTCTGCCTAGACCCTAAGAAACTT AAGGTTGTCCTGTATGCCGCTATG CCGTATCACGATATGGTAGAAGGC AAAGAGTACGGCGTTATGTCCCTC CTGATGCAAGAGCATGTGCGTTCC CTCATTCCGGTTCTTAGTGAACAC TTCCCGGAAATCACATGGGAACTC TCTGAGGCCAAAGAAGTATACGA CATGGTAGAACTTCAGGAACTCTA CGAGAAAGCCCGTGCCGATGGTC ATGAAGGTCTGGTCGTCAAAGACC CGCTGGAAGTCTACAAGCGTGGC AAGAAGTCTGGCTGGTGGAAGAT GAAGCCTGAGAATGAAGCTGATG GTATTATCGTTGGTATCAATTGGG GGACTAAAGGTCTCGCTAATGAA GGCAAAGTTATTGGCTTCGAGGTT CTCTTGGAGTCCGGTCGTGTAGTC	1032	–	DNA ligase	0.096	0.087	0.22	0.12

		TCTGCCAACAACATTAGTCAAGCA TTGATGGATGAGTTTACCGCTGCC GTCAAACATGCTGAGTCTGTGGGT ATGGGTATGCCGTACACTGGCTTC GCCTGTCAGGTAGCCTACATGGAG GAAACTCCTGATGGTTCTCTTCGT CACCCATCGTTCGTTATGTTCCGT GGTGTCGAGTCAGACCCAACGGTC AAGATGTAA (SEQ ID NO: 26)							
9	gp1.5	ATGCTATTTCTTATTGCTGCTGTTA TCTTTGCCACGTTTGCCTTCGTTAT CATTGACGACAACTGCTGGCCTGA TTGCTAA (SEQ ID NO: 27)	81	–	Hypothetical	0.15	0.15	0.29	0.20
10	gp1.6-gp1.7	ATGTTACGACTACACTTCAACAAA TCAAATGGCATCTTCTCTGCCCGT GGCGCTGACCGTTCAACCGTTCTG GCCTCTGAGAAACACGCAGTCATC CCTAAAGTGCTTGGCTCGACCATT GAGTTAGCACCTCGTGTTCACGCT GTAATCACCACTGGCTTATATAAA GAAGCGACCACTGGCTCCCGGCCT TTCATTCCGGTTCTCGTCACCCGTT TCCCTAAAGTCCGTCTGGTAATTG CGCGTATCAAGGAGGTGTTCTAAT GGTCTGCATGGCATGTTTAGAGAA TGATGACCGCTACCCACATACATG TGAACCTGTAGCTAAACCACAAG ATATAGAGGGTGTTCGAACTCCGA GCCACTATATGTTGTTTGATGACA TTGAGGCTATCGAAGTGATTGCCC GGTCGATGACCCGTGAACAGTTCA AAGGGTACTGCTTAGGGAACATCC TTAAGTATCGCCTACGTGCTGGTA AGAAATCCGAACTGGCATTCTTAG AGAAGGACATGGCGAAAGCTTGG TTCTACGGTGAACTGTACGAGAAA CATAAGGGTAAGTGCTATGCTTAA (SEQ ID NO: 28)	575	–	Hypothetical; hypothetical	0.34	0.33	0.51	0.42
11	gp1.8	ATGGAAAAGCCTAGCGAATGGTG TGAGCGTATGTTTGAAAAGACAG GCAACACCGACTACCTTGAATTAT ATAATCACTGGAAGGAGAGAGGA CAATG (SEQ ID NO: 29)	98	–	Hypothetical	0.45	0.41	0.74	0.34
12	gp2	ATGAGCACTCTGAATAAAGGGTC ACTCGTAAACGGTGACAAGAAGT TCTTCGCAACGGTCGAAGGCCAAG GCCAATCGTTCGAAGTCCCTGTGT TCGCTGCGTCTCTTGAAGAGGCAA ACGAGTTAGCTGAATGGCAGTATG TTCCGGCTGGCTTCGCAGTGACAC GCATCCGTCCAGACCGAAAGGTTT AA (SEQ ID NO: 30)	192	–	Inhibitor of host RNA polymerase	0.048	0.043	0.10	0.052
13	gp2.5	ATGGCTTTCAACAAACGTAAAGTA TTCACTTCCGCTCGTGGCGTTGCT GAACCGTATTGCTATCTGGCTAAA CCGGACTTCGGTAGCGGTGACTTC AAGAACGAACGTGGTGTCTACAA GGTGTCACTTACTGTATCCAACGA CGACCCCCGCTGTCAGAAAATGAT TTACGAGATTGTCGAAGCCCATGA AGCGGACTATGCTGCTCGTCTCGA AGAGTACGAAGCGAACCCGCCGA AAGTTGTCAAGGGTAAGAAACCG CTGAAGCCGTATGTTGGCGACATG CCTTTCGTGGATAACGAAGACGGA ACCACTACGTTCAACTTCAAGTGC TACGGCTCGTTCACTGACAAGAAG ACTGGTGAGAACAAACCAATCGA CTTGGCTATCGTGGACTCCAAGGG CAAGCGCATTCGTGGTGAACGTCC AGCTATCTCTGGTGGCTCTGAACT GAAAATCAAATACACCCTGTTCCC	699	–	Single-stranded DNA binding protein	0.14	0.14	0.23	0.18

		TTATGGCTGGTCTGCTGTGGCTGG TGCATCCGTTAAACTACAACTGGA TTCTGTGATGCTGATTAAGCTGGT CGAGTTCGGTGGTGGTGATGACGA TTGGGGTGATGAAGTCGAAGAAG ACGGTTATGTAGCTGATGAAGCAC AGACCCGCAAGCCGCAGAACGAC TCTGGCTGGGACGAAGAACCAGA AGAGGACGATGAAGATGAAGACG GTGACTTCTAA (SEQ ID NO: 31)							
14	gp3	ATGGCTAAGGGTTATGCAGGTCGA GGCACTCGTGTGGGTGCCTTTCGC TCTGGCTTGGAGGATAAGGTAAGC AAGCAACTCGAAGGGAAGAACAT TAAGTTCGACTACGAGCGCTGGCG TGTACCATATGTTGTTCCTGCAAG CAACCATACCTATTGCCCGGACTT CCTATTACCTAATGGCATCTTTGT GGAAACAAAGGGTCTGTGGGATA GTGACGACCGCAAGAAACACTTA CTGATTCGTGAACAGTATCCTGAA CTTGATGTGCGCATCGTGTTCTCTT CAAGCCGCTCTAAGCTATACAAAG GGTCTCCGACCAGTTATGGTGAGT GGTGCGAGAAGCATGGCATTCTGT TTGCTGACAAATTAATTCCAGTTG ATTGGCTGAAAGAGCCTTCTAAGG ATGTTCCGTTCGATAAATTTAAAG AGGTGAAGAAGAAATGA (SEQ ID NO: 32)	447	–	Endonuclease I	0.0081	0.0076	0.0081	0.0076
15	gp3.5	ATGAGTAAGGTACAATTCAAACC ACGCGCTGTGACAGAAGCAATCTT TGTCCACTGTAGCGCAACCAAGGC GTCCATGAATGTTGGACTGCGTGA AATCCGTCAGTGGCATAAAGAAC AGGGCTGGCTTGATGTAGGCTACC ACTTCATTATTCGCCGTGATGGTA CAATCGAAGAAGGCCGTCCGGTC GATGTAGTTGGGTCTCACGTTAAG GACTGGAACAGTAAGTCAGTCGG TGTGTGCCTCGTTGGTGGTATCGA CGATAAGGGCAGACACGAAGCTA ACTTTACGCCAGCACAGATGCACT CTCTTAAAGAGAAACTGGCAGAC CTGAAAGACCTGTACCCGGATGCA GAAATCAAAGCACACCACGATGT GGCTCCTAAAGCGTGTCCAAGTTT TAACTTGAGTCGCTGGCTGAAGAC TGGTGAATTAGTAACAAGTGATTG GGGTTAA (SEQ ID NO: 33)	456	–	Lysozyme	0.52	0.55	0.46	0.57
16	hypothetical protein	ATGATGATTATTCATGAGATGGGA ATGTACGATATGATTTACCGCTGT GTATTCGTTCTGGCATTGACCTAT ATCTTAACTCGTAAGTAA (SEQ ID NO: 34)	90	–	Hypothetical	0.16	0.16	0.34	0.23
17	hypothetical protein	ATGCGTAAGCTATTTGGTGAAGCA CTTCCGGTGACACTTTTGGTCATC ATCATTATCGTTTGTGGCGCACTG CGTCCGTTCCTGTAG (SEQ ID NO: 35)	87	–	Transmembrane	0.44	0.41	0.67	0.61
18	gp4	ATGGATATGGAAGAACAACAAGA GAGTATATTCCTCTATCATTTACCT TGTGAGAACTGTGGGTCGTCAGAT GGTAATGGCATGTATTCTGATGGA CACCAGTATTGTTTTGTGTGTCAG AACCATGTGCGTGGAACCGAGGA AACTCGTGAGACCGTGGCGAAGA CTCGCAGAAGCTCATACAACACTG GAGGTGACAAAATGTCTAACCTGT TAAACTTTGGGGATTCAGACGGTC GCTACACGAACCTCAAAGCTCGTG GCCTTATGGAAGCTATCTGCCGTA	1707	+	DNA primase/helicase	--	--	--	--

		AATATGGCTATTGGGTAGCGAAG GTCAACGGTGAGATGCTGCAAGTT GCAAACTACTATGACGTTGAGGG AAATCTCGTAGGTCAGAAGGTAC GCGACAAGCACAAAGAGTTCTCC GCAAAGGGGAAACTTAAAGCTGA CCTATTGTTTGGTAAGCAGTTATG GAATGGTGGCAAGAAGATTGTGG TGACTGAGGGTGAGATTGATTGTC TGACCGTAGCGCAACTTCAAGAA GGGAAGTATCCCGTTGTGAGTCTC CCTATGGGTGCTCAAGCCGCAAAG AAAACATGTGCTGCAAACTATGA ATACTTCGACCAGTTCGATGAGAT TATCCTGATGTTCGACATGGATGA GCCGGGCCGCAAGGCCATCGAAG AGTGTGCTCCTGTCCTGCCTAGCG GGAAAGTCCGAGTCGCTGTACTGC CTCTTAAGGATGCCAATGAGTGTT TGCTCAATGGACAAGCCAAAGCA GTAACCGACCAGATTTGGAATGCG CAACCTTGGGTGCCTGATGGTGTT GTCTCTGCTGTGTCACTTAAAGAC CGTGTGCGTGAAGCGATGGTCAA AGAGGAAACAACAGGTCTACTCTT CACGGGTCAACCGAAGCTCAACG ACATGACACTTGGTGCTCGTGGTG GTGAAGTCATCATGGTAACTTCCG GTTCCGGTATGGGTAAGTCAACAT TCGTTCGTCAACAGATGCTCATGT GGGGCAAGGGTGGTGCGAAGGTA GGCTTGGCGATGCTTGAGGAAGCC GTAGAGGAAACTGTTCAGGACTTG ATGGGCTTGAACAACCATGTCCGT CTACGTCAAGACAAAGAACTCAA GATGCAAATCTTAGAGGATGGAC GCTTCGATGAGTGGTACGATGTAC TGTTCAACACAGATATGTTCCACC TGTATGATTCCTTTGCGGAGTCAC AGGAAGACCGTCTGTTCGCAAAG CTGGCCTACATGGTCGATGGCTTA GACTGTAACGTTATCCTGCTTGAC CATATCTCTATTGTGGTCTCCGGG ATGGAAGATAACTCCGATGAGCG TAAGACGATTGACCGACTGATGAC CAAGCTGAAAGCCTTTGCGAAGA CTAAAGGTGTGGTCGTGGTGGTAA TCTGTCACCTCAAGAATCCTGAGA AGGGGAAAGCACATGAAGAAGGT CGTCCTGTTAGTATTACTGACCTG CGTGGTTCTGGCGCATTACGCCAA CTTAGCGATACTATCATTGCCTTG GAACGTAACCAGCAAGGTGACTA TCCTAATCTTGTCCAGTTGCGTGTT CTCAAGTGTCGTTTTACTGGCGAT ACTGGCGTGGCTGGACACATGGCC TACAACAAAGAAACGGGCTGGCT TGAACCGACTGTCTCACCTGAAGA AGATGGAGGTGGAGATTGGGAAC CGGAAGAAGACGGTCAAGACTTC TGA (SEQ ID NO: 36)							
19	gp4.2	ATGAAGAAGGTCGTCCTGTTAGTA TTACTGACCTGCGTGGTTCTGGCG CATTACGCCAACTTAGCGATACTA TCATTGCCTTGGAACGTAACCAGC AAGGTGACTATCCTAATCTTGTCC AGTTGCGTGTTCTCAAGTGTCGTT TTACTGGCGATACTGGCGTGGCTG GACACATGGCCTACAACAAAGAA ACGGGCTGGCTTGAACCGACTGTC TCACCTGAAGAAGATGGAGGTGG AGATTGGGAACCGGAAGAAGACG	357	–	Transmembrane	0.9	0.92	0.68	0.82

		GTCAAGACTTCTGATAGTAAATCA ATAGAACTCACTAATGGGACAGA CACTCTGTTCCCACTTATCAATCC ACTTAAGGAGAATCACAATGTTTA A (SEQ ID NO: 37)							
20	gp4.3	ATGTTTAAATTCATCAACACTCTT GGTAAGCTGGTAGTTAAAATGTAC TTCCGTGAAGCTAAACGTCTGAAC GACAAGGCACGTAAGGATGCAGA ACAGGCTCAACGCTTGGCGAAAC AGGCGCGTCTACTATCTGAAGATG CAAGTGCTGGCGTTACGTCTGCTG CTAAGATTGCAGCGAAAGCAACC GACCTGAACAAGTTCTTCCTGTGA (SEQ ID NO: 38)	213	–	Hypothetic al	0.15	0.15	0.27	0.23
21	hypotheti cal protein	ATGGTTACGATAACCCTTGTGCTC ACTCTGATAACTGTGAGCGGACAA ACAATAAGTCAAGTCCCTATCACG GCACCGAAGGCTGAGGCATACGA GATATGCGAACAGCAAGGGCGCT CATGGTCTAAGGACTTGATGAAGA ACGGCAGGGTTTTCACTGCCAACT ACAACTGTATCTAA (SEQ ID NO: 39)	180	–	Hypothetic al	0.54	0.49	0.82	0.49
22	gp4.5	ATGGCTATTAAATTCCCCGGCAAC ACAATCCGTCTGTCCGACACCATC GACCAGTATGCCCGTCGAGTGCAC ATCAACGTCCGCAACGGAAAGGT AACACTGGTCTACCGTTGGAAGGA CAGTAAGTCCACCAAGGCGCACA CACAGCGTGTGACTCTGGATGACA CACAGGCAGGTCGTCTGCTGGCCT CCGTTGCTGTAGCCGCTACTGTGG CTGTAGGTGAAGACAATGTTCGTC AGGTTCTGCTGGGTAAAGCCGTGG AGTCTGAGGGCAACATGCTTGCTG ACAAGTGTGAAGAGTAA (SEQ ID NO: 40)	303	–	Hypothetic al	0.089	0.083	0.16	0.099
23	gp4.7	ATGCGTAATCGTAATGTAATCGAA GCTGAAATCGCTAAACTGGAAGCT GAACTGGCAGATGTTAAAGAGTA TGACTGTGCTCGTGACTCCGCTGT CCACATCCTGAAGAACTTAGGGTG GACCCGCAAGTATGGCAAGTGGG TCAAACCAGTAGCCAACTGTGGTA GCGTTGGTATGCAGACTTTCGACG CAGACAAGATGACCCACATTAAG GCTGGGGACTTCTGCACCTCCACC AGCAACGCAGACATTTGGTATGTC CGTAAGGTCGAAGGGACAGACAT TTGGTGCTCCCGTGTGACTCGCAT TGATGCAATGGGTACAGTCGCTGG GATTCAGTACTACAAGATGAACCA CAAGAACCTGCGTGTGGTTGACCC TAAGAACTTCATGGGATACCGCAA ATCTACATTTTAA (SEQ ID NO: 41)	417	–	Hypothetic al	0.31	0.29	0.45	0.36
24	gp5	ATGTTACTATCCGACATCGAAACA AACGGGTTGCTGCACAACGTTAGT CAATTTCACTGTGCGGTGACTTAC GACTACACAACTGACCAGTACATC CGTTACCGACCTGATGACTTTGGG GCGTACATTGATGCACTCGAAGCC GAGGTTGCTCGTGGTGGTCTCATT GTGTTCCATAACGGTCACAAGTAT GACATCCCGGTAATCGAGTTGCTC GCCAAGAAACTGTTAGGCCGCGA CGTTAAGTTCCCGAAAGAGAACGT ACTGGATACCCTCGTGATGTCCCG CTTGATTTACTCTAACATCAAAGA CACCGATGCTGGTCTGCTGCGTTC TGGCAAGCTGCCGGGTAAACGTTT TGGTTCCCACGCTCTGGAGGCTTG	211 5	–	DNA polymeras e	0.55	0.56	0.52	0.51

		GGGTTATCGCTTAGGCGAAATGAA GGGTGAGTACAAGGACGACTTCA AGGCTTCACTGGCTGAGTCTGGTG AAGAATATGTTGATGGTATGGAAT GGCTCCTGTTCAACGAAGATATGA TGGAATATAACGTTCAAGACGTTG TAGTCACCAAGGCTCTCTTCGAGA AACTTTGTAGCAACACATTCTATT TCCCTAACGAGACACCCGCTGGAA GCACAGAAGCAGACCGTTTCTGG AATGGTAGCCTTGAGGCTGTTAAA CTGGAACATGATGCTGCATGGTTG TTAGCCAAGATGGAACGTAACGG TTTCCCGATTGACATTAAGTCTCTT GAGAACCTGTACGCAGAACTCGCT GGTAGACGTGGTGAGCTACTCGTA GAGTTAACCAATACGTTTGGCTCA TGGTACTCCGCTAAGGGCGGTACT GAAGCATTCAGACACCCTCGTACC GGAAAGCCGCTCACCAAGTATCCT CGTGTTAAGTACCCGAAACAGGGT GGCATCTTTAAGAAGCCTAAGAAC AAGAAGCAACGTGAGGGATTAGA GCCGTGCGAGTTGGATACCCGCGA TTACGTTGAGGGTGCTCCGTATAC TCCGGTTGAGCACGTTGTGTTTAA CCCTGCGAGCCGAGACCACATCAC CCGTGTTCTCACCAAGGCTGGTTG GGTTCCCGTTGAGTTCACCCCAAG TGGTGCACCAAAGGTAGACGACG AGGTCTTAGAGCACGTTAAGGTAG ACGACCCTGATGCACAGCGTTGCA TTGAGTTGATTAAAGAATACCTTA TGATTCAGAAACGTATCGGTCAGG TTGCTGAGGGTGATAACGGCTGGC TCCGAATGATTGGAGAAGATGGTC GTATTCACGGAAGCGTTAACCCGA ACGGTGCGGTAACTGGTCGTGCAA CTCATAGTTTCCCTAACGTGGCAC AGGTGCCTTCAATCCGTGCTGCAT ATGGCGAACCTTGCCGTGCTGCTT TCGGTGCCGAGCACAACATGAAA GACGGTAAGCCAGACCCTTGGATT CAAGTTGGTGTGGATGCGTCCGGC CTTGAGCTACGATGCTTAGGTCAC TTCATGTATCGCTATGATGAGGGA GAGTACGTTGAGACTATCCTCACT GGTGACATCCATACGAAGAACCA ATTGGCTGCTGGTCTGCCAACTCG TGACAACGCTAAGACGTTTATCTA TGGGTTCCTCTATGGTGCTGGCGC TGCGAAGATAGGACAGATTGTTG GTGGTACTGCCGAAGACGGTAAG CGACTCATCAAGAACTTCCTTGAG CAAACTCCGGCTATCGCTGCATTA CGTGAGGCCATTACAGGTACTCTC GTCAAGGACTCTAAGTGGGTTGGC GGTGAACAGAAAGTAACTTGGAA ACGCCGTTGGGTTCGTGGATTGGA TGGTCGTAAAGTACACGTCCGTAG TCCTCACGCTGCGTTGAACACCTT GTTGCAGTCTGCTGGTGCTCTGAT TTGTAAACTGTGGATTGTCGAGAC TGAACGACTGTTACTGGAAGCTGG CCTGAAGCATGGCTGGGATGGTG ACTTTGCGTACATGGCATGGGTTC ACGACGAAATACAGGTCGCTGCA AGGACTCAAGAGATTGCCGATAA GATTGTCGAGTTGGCACAGCAAGC AATGCGTAATGTGGGAGACCACTT TAACTTTAGATGTCAGTTAGATAC CGAAGGTAAGACTGGCGCAAACT							

		GGAAGGATTGTCACTAA (SEQ ID NO: 42)							
25	hypothetical protein	ATGGCTTGGTTATATAGAATACGT GAAGTTAGACTTACTACAGGAGA GACACTCTACTACCCTCAATTCCG CCACAAGTGGTTACCGTTTTGGTG CCACTTCCGGGATTATTCAAATGA GAACAAAGTGTCATTCACTTATTA CTCTGAAGCTGTCAACTTTGTTAA CGAGTCGAAAGCAAATGACGAAC CTGAAGTTAAACAGGTTGTCCGTA ATCACTATCTTTAA (SEQ ID NO: 43)	228	–	Hypothetical	0.44	0.43	0.71	0.55
26	gp5.5	ATGGCTATTACTAAACGTTTTAAA GTATCATTTGAAGTTACCGCTGTA ATCAACTCTGAGATGGAGAAAGG CTTCACCGAGGAACTGGTTCATTT GGCTAAGAAGGCTGCGGCAGGTG AGAAGATTAGTCCTCGTGACCGTG AACTTCTGGTACAAGCGCTGACTC ATGGTGTCGGTGGTGCTATCTCCT TTGCCGCCAAACAGGGTATGCGA GAGTTCATCCGTACCGAGTTACCG GAAGATGGTCTGAAGGTTTCTCCA GCAACTATCCGTGAGGTGAAGTA A (SEQ ID NO: 44)	285	–	Binds to host histone-like protein H-NS and transfer RNAs	0.64	0.62	0.82	0.77
27	gp5.7	ATGGGTGAATATCTACGTGTACTG GCGGCTATCAAAAGCTGCCCTAAG ACCTTTCAGTCGAACTATGTGCGC AACAATGCGTCACTGGTGGCTGAG GCTGCGAGTCGAGGTCATATCTCT TGCCTATCAGTCGATGGGCGTAAT GCTGGCGCTTGGGAAGTAACAGG CGCTGGTGTCCGATTCCTGAAGAA GATGGGAGGTTGCGTATGA (SEQ ID NO: 45)	210	–	Inhibitor of host RNA polymerase	0.068	0.06	0.077	0.061
28	gp5.9	ATGAGTCTGCACACAGACAACGTT ACAGTGACCCGTGAGGCGTGGAA CGACATGCAGTCGTACATCCAGTC ACTGGAGAAGAAAGCTGAGAAGC TGGACGCTCTGGAGGCTGGAGGT GTGGACAATAGGTCCGGCTACTGG TTCTCTCTGGAAGAGGCTGGTCTC ACAGGGGAGGACGACTTATGA (SEQ ID NO: 46)	186	–	Inhibitor of host RecBCD	0.25	0.23	0.42	0.20
29	gp6	ATGAGTGCAATTACACTGAAGGA GTTCTACGAGTTACGCCAAGGGAA ATCCGATAAGGGTGTTCTCGTAAT GGATGGTGACTGGTTGGTATTCCA AGCGATGGCAGCAAGTGAAGTCG AGACAGATTGGGGTGACGATGTG TGGACCTTAGAGTGTGACCACGGG AAAGCGTGGAACATTCTGGTTGAC TCTATCAAGTCCTATGCGAGCCGT AAGAAGGCTTGGCGTGATGCACCT ATCGTATTGGCCTTCACTGACGCT GTGAACTGGCGTAAGGAACTGGT GGACCCAACCTATAAGGAGAACC GTAAGGCATCCCGTAAGCCTGTAG GCTACCGTGCGTTCATCGAGCGTG TCCAAGCCTGTGAGGACTGGACGT CTATCCTTGAGGAACGACTTGAGG GTGATGACGTAATGGGAATCATCG GGTCTGGTGCTAAACATTTTGGTT TCAAGAAGGCTGTTCTTGTCTCCT GTGATAAGGATTTTAAGACCATCC CTGATTGTGACTTCTTGTGGTGCA CGACTGGTAACATCTTGACTCAAG ACCAAGCGACTGCTGACTACTGGC ATATCTATCAGACCATTAAAGGTG ACTTGACAGATGGTTATGGTGGCA TTCCGGGCTGGGGCGAGACTACCG	915	–	dsDNA-specific exonuclease	0.082	0.081	0.14	0.12

		CTGAGTGGCTGGAGAACCCATATG CCTTTGAACAGGTTGAGAAAACTT TCAAGTCTGGTGCCCGTAAGGGTC AGACAGTTCTCGAATGGAAGAAG GTTGACATGGGTGAGCGAACCTTA TGGGAAGCCATTGTGACACTTGGT GCGAAAGCTGGGATGACCGAAGA GGAAATCATTAAGCAAGGCCGGA TGGCTCGCATCCTTCGATTCGAAG AGTACAACTTTATCGACAAGGAG ATTTACCTATGGTCGCCGGAACAG TTGCGTTCGTAA (SEQ ID NO: 47)						
30	gp6.5	ATGTTAACCCCTATCAAGAAATAC ATGGAGAACCCACAAGATATTCCT AGCGTTCCACGTAGTGTTATGGAG TATTTACAGGTTCAATATAATGCT GGATATGCCTTGCAGTCTGGTTTA GTTAACCGATTGAAACAGGCTGGT TGGTCTGAAAGCTACATTGCTGGG TTCCTCGCGGGTCTCAACTATGCG TCTCAAACCTTAGACGACATGGAG GCAATCCGTAAGGAACAGGCTGA GTCCTAA (SEQ ID NO: 48)	246	–	DUF2717	0.81	0.78	1.2	1.0
31	gp6.7	ATGTGCTTTTCACCAAAGATTAAG ACACCTAAAGTAAAGACCGATAC TGTCCGAGCTATTGACCCAGCGCC ACTGACTGAGGAACCGAAAGGTA TTCTCTTTGGAGGCGATGAGGACA CAGATAGCGGAACCTCTTCTGAGG TGCCAACGGGTGGCAAGAAGTCC CTTAAGGTGAAACTGGATGACTCC GTTGAGAAGTCCAAGAAGGCTGA TGATGCAGCCAAGAAGAAGTCTA AATCTGGTATCCGTACCAGTATCT TTAAGAAGAAATAA (SEQ ID NO: 49)	273	–	DUF5476 head protein	1.0	1.0	1.3	1.0
32	gp7.3	ATGGGTTGGGGTAAAAAGATTCG AAAATCAGTTAAGAAGGCTGTCA AGAAAGTAACCAAGCCTATCGAG AAGGTAGTCAAGGAAACTGGCAA GGCAGTTGAGAAGGTTGGTAAGG AAGCTGGGAAAGTAGTTGGTGGT GTTGCTGGTACTGTTCAAGCTGCA CCTCCTGCTGCACAGATTATTGAG CCACCGAGTAAAGACGATGTAGA CTCCGATGATGATGCACAGACTGA ATCCGGTAAGAAGAAAGCCCGTG CTGGCGGTAAGAAATCTCTGAGTG TCGCTCGCAGTTCTGGCGGTGGTC TGAATATCTAA (SEQ ID NO: 50)	315	–	Tail assembly protein	1.2	1.2	0.98	1.2
33	gp8	ATGGCAAGCTCACAGAAACGTGA AGGCTTCGCGGAGAATGGTGCCA AGGCGGTGTATGACGCATTGAAG AACGACCGGAACTCCTACGAGAC CCGTGCGGAAAACTGTGCGAAGT ACACCATCCCTTCGTTGTTCCCTA AAGACTCCGACAACGCCTCTACTG ACTACACCACTCCGTGGCAAGCAG TTGGCGCTCGTGGCTTGAACAACT TGGCCTCTAAGCTCATGCTCGCTC TGTTCCCTATGCAGACTTGGATGA AGCTGACCATCTCTGAGTTCGAAG CGAAACAATTGGTTGCACAGCCA GCCGAACTAGCGAAGGTCGAAGA AGGTCTCTCAATGGTCGAGCGAAT CTTGATGAACTACATCGAGTCCAA CTCGTACCGTGTGACACTCTTTGA GACCCTCAAGCAGTTGGTCGTTGC AGGTAACGCTCTGCTCTACATTCC TGAACCTGAAGGTGCTTACAATCC CATGAAGCTATACCGCCTGTCTTC TTATGTTGTCCAACGAGACGCATT	160 8	–	Portal	0.45	0.43	0.65	<0.0001

		CGGCACGGTGCTACAGATTGTAAC CTTAGATAAGACTGCCTATGCAGC ACTTCCAGAAGACGTGCGAAACG CTATGGACTCCGGTCAGGAACATA AAGGCGACGAAATGATTGATGTG TACACTCACATTTATCTTGACGAA GAATCAGGAGAATACCTCAAGTA CGAAGAGATTGATGGTGTTGAAGT CGATGGTACAGATGCTTCTTATCC GGTGGACGCTTGTCCTTATATCCC GGTTCGCATGGTACGCATTGATGG AGAATCATATGGTCGTTCTTACTG TGAGGAATACTTAGGTGACTTACG GTCCCTTGAGAACCTTCAGGAAGC TATCGTTAAGATGTCAATGATTAG TGCAAAAGTTATCGGCTTGGTTAA CCCGGCTGGTATCACACAGGTTCG TCGTTTGACCAAAGCACAGACAG GTGACTTCGTGTCCGGTCGTCCAG AAGACATCTCGTTCCTTCAGCTTG AAAAGGCCGCTGACTTCTCTGTAG CGAAAGCTGTGAGTGAACAGATT GAGGGCCGACTGTCCTATGCCTTT ATGCTGAACTCCGCTGTGCAACGT ACTGGTGAGCGTGTGACTGCTGAA GAGATTCGATATGTCGCATCTGAA CTCGAAGATACCCTTGGTGGTGTC TATTCGATTCTCTCACAGGAACTT CAGTTGCCAATGGTTCGTGTCCTG CTGAAGCAACTTCAAGCAACCAAT CAGATTCCTGAGTTACCGAAAGAG GCCGTTGAGCCTACTATCAGTACC GGGATGGAAGCATTGGGTCGTGG TCAAGACCTCGATAAGCTGGAAC GTTGTATTGCTGCGTGGTCTGCGC TGGCTCCTATGCAGAATGATCCTG ACATCAACATTGCGACCATTAAGT TGCGCATTGCGAACGCTATCGGCA TCGACACTTCTGGTATCCTCAAGA CACCTGAAGAGAAACAACAGGAG ATGGCTGAGGCTGCTCAAGGCACT GCTCTGGAGAACGCTGCTGCGTCT GCTGGTGCTGGTGCTGGTGCACTG GCTACTGCAAGTCCTGAGAACATG GAAGCTGCGGCTGCACAGGCTGG TATGGTTCCTAATTAA (SEQ ID NO: 51)							
34	gp9	ATGGCTGGTGAATCTAACGCTGAC GTATACGCATCTTTCGGTGTGAAC TCTGCGGTTGTTGGTGGGTCTACT CATACTGAACACGAACAGGCTAT GCTTGAACTCGATGTTTCTGCCCG TGATGGCGATGATGCAATCGAGCT AGTTAGCAACGATGACCCTTATGG CAATCCAGACCCATTCGGTGAAGA GGATGAGAGTCGTACTGAGATTCG CATCTCCACTGATGGTGAACCTGA AGTAACCGAAGAAGGTGAAATCG ACTATGAAGGAACCGAAGATGAA GGCTCCGAAGAGTTTGAACCAGTT GGTGACGTTCCAGATGAACTGAAC GAAGCCTCAAGTCAACTCGAAGA GCACGAAGCTGGTTTCGAAGAGA TGGTCAATCAGGCGGCTGAACGTG GTCTGTCAGAAGAAACCATCGTGC GTATCCAACAGGAATACGGAGAG GACGGTCTGTCTGAAGAGTCTTAC GAAGAGTTAGCCAAAGCTGGTTA CTCCAAGTCGTTTGTTGACTCCTA CATCCGTGGTCAGGAAGCACTGGT AGAACAGTATGTTCAGTCCGTTAT GGAATACGCTGGCGGTGAAGCAC	912	–	Capsid assembly scaffolding protein	0.71	0.71	0.95	0.76

		AGTTCCAAGCAATCTACAATCATC TCGAAGTGTCCAACCCGGATGCTG CTGAGTCTCTGATTAGTGCCTTGG AGAACCGTGACCTCGCAACAGTTA AGGCTATCGTCAATCTCGCTGGTG CAAGCCGTGCTAAGACTTTCGGTA AGCCTGCTGCCCGTAGTATCTCTA AACGTGCCGTACCTGCTGCACCTG TACGCACCAAACAGTCTGGCTTTG AGTCACAGGCTGAGATGATTGCTG CAATGTCCGACCCTCGCTATCGCC ACGACAGTAAGTTCCGTGCGGAA GTTGAGCGCAAGATGATGTACAGT CAGTTCTAA (SEQ ID NO: 52)							
35	gp10ab	ATGGCTAACATGAATGGTGGACA GCAGATGGGTAAAGACCAAGGTA AAGGCATGAGCGCTGGTGACAAA CTGGCTCTGTTCCTGAAGGTCTTC GGTGGTGAAGTTCTGACTGCATTT ACCCGTACCTCCGTAACTATGAAC AAACACCTTGTGCGTTCTATTCAG TCCGGTAAGTCCGCGCAGTTCCCT GTGCTGGGTCGTACCAAAGCTGCT TACCTGCAACCGGGTGAGAACCTC GATGACAAACGTAAAGACATGAA GCACACTGAGAAGGTCATCAACA TTGATGGTCTGCTGACTGCTGACG TTCTGATTTACGACATCGAAGACG CAATGAACCACTATGACGTGCGCT CCGAGTATACCGCACAGTTGGGTG AATCTCTCGCGATGGCTGCTGATG GCGCGGTTCTGGCTGAAATGGCTA AACTGTGTAACCTGCCGTCTGCGT CTGATGAGAACATCGCTGGTCTCG GTAAGGCTCATGTACTGGAAGTTG GCAAGCAGTCTGATCTTCAGGGAG ACCAAGTTAAGCTGGGTCAGGCTA TTATCGCACAGTTAACTCTGGCTC GTGCCAAGCTGACTAGCAACTATG TTCCGTCTGCTGACCGTGTGTTCT ACACCACTCCAGACAACTACTCTG CTATTCTGGCTGCTCTGATGCCTA ACGCTGCAAACTATCAGGCTCTGA TTGACCCGTCTACTGGCTCTATCC GCAACGTTATGGGCTTCGAAGTTA TTGAGGTTCCGCATCTGACTGCTG GTGGTGCTGGCGAAGACCGTCCA GAAGAGGGTGTAAATCCGACTGG TCAGAAACATGCGTTCCCTGAGAC TTCCTCTGGTGACACCCGTGTTGC TCTGGACAACGTTGTCGGTCTGTT CAATCACCGCTCTGCCGTTGGTAC TGTTAAGCTGAAAGATATGGCACT GGAACGTGCTCGTCGAGCTAACTT CCAAGCAGACCAGATTATCGCTAA GTACGCTATGGGTCACGGTGGTCT GCGCCCTGAAGCATGTGGCGCGCT GGTTTTCAACAAGGCTTAACTTCT CGCAAACTCGTGGGGGACATTGA GTCCCTTACTGAAGTTGAGCAAGA GCTAACGCCACAACAGAAAGCTG CCCGTACCCGTGCGGCAAACAAG GCTCGAAAGGAAGCTGAGTTAGC TGCTACTGAATGA (SEQ ID NO: 53)	1178	+	Major and minor capsid proteins	--	--	--	--
36	gp11-gp12	ATGCGTTCCTATGAGACTACCCTA GAGACTGGCGAAGAGTTGGCTGC GGTAAATGACATCTTGGCCTCTAT CGGTGAACCACCTGTGTCCACTTT GGAAGGTGATGCTAACGCCGATG TTGCCAATGCTCGTCGAGTTCTCA ATAAGATTAACCGACAGATTCAGT CCAAGGGTTGGACGTTCAACATTG	2995	–	Tail tubular proteins	0.27	0.26	0.3	0.23

		AGGAAGACCAACAGTTGCTACCT							
		GATGTGTTCAACGGATTGATTCCT							
		TATATGTCCGACTACCTGTCTGTG							
		ATGACACCGGGAGGCGCTACTGC							
		CTATGTCAACCGGGGTGGTTACGT							
		GTTTGACCGGACTACTGGAACTGA							
		CGTGTTTAAAGAACCAATCACTGT							
		GACCATTATCAAGCTGAGAGAGTT							
		CTACGAGATGCCTGAATGTTTCCG							
		CTTCTGGATTGTCACCAAAGCTGC							
		CCGTCAGTTTAACAACAGGTTCTT							
		TGGTGCTCCCGAGATTGACGCAGT							
		GTTAGCTGAAGAAGAACAAGAAG							
		CAAAGATGCAGTGTCACGAGTAT							
		GAACTGGACTTTGGTAACTTCAAC							
		ATGCTTGATGGTGATGCCTTCACT							
		GGCGGTCTGCTATCTCGCTAATGA							
		TAACAACTTAAGGAGGTCTCTAAT							
		GGCTCTTATTTCACAATCCGTCAA							
		GAACCTGAAGGGCGGTATCAGTC							
		AACAGCCGGACATCTTAAGGTTCC							
		CCGAACAGGGTTCCGAACAGATT							
		AACGGCTGGTCTTCGGAGACTGAG							
		GGTCTTCAGAAGCGTCCACCTTTT							
		GTCTTCACTAAGACCATTGGAGAC							
		CAGAATGCCCTTGGTGCAAAACCG							
		CTCATCCATTTAATCAACCGTGAT							
		AGCGTAGAACAGTATTACGTTGTG							
		TTCACAGGACAGTCAGTCCGAGTG							
		TTCGACCTCAATGGCAAAGAGTAC							
		ACAGTGAAAGGTGACTTGAGCTA							
		CGTGAAGGTAGGGAACCCGCGTG							
		ATGACCTGCGAATGGTAACGGTTG							
		CTGACTATACGTTTATCGTCAACC							
		GCAACATGGTGGTTCGTGCCGATA							
		AGACACCACTCTATACTCTGAAAG							
		AGAACGGAGACTGTTTGATTAACG							
		TTCGTGGTGGACAGTATGGGAGA							
		ACGCTTGCATTTACAATAAACAAC							
		GTTCGAATTGAGTACAAGATTGCA							
		CAAGGAGATAAGCCTGAACATGT							
		ACAAGAGACCGATGCACAGTGGC							
		TCGTTAAGAAACTGGCAGGTCTTG							
		CTCGTGCTCACGGTTCCTTCAAGG							
		ACTGGAGGTTCGACGAAGGGTCT							
		GGCTTCATCCACGTAATCGCTCCG							
		GGTAACAGTCAGATTAACCAGCTG							
		TCCACTGAGGATGGCTATGCTAAC							
		CAGTTGATGAACGCGGTAATGCAC							
		ACAAGCCAGTCATTCAGTAAGTTG							
		CCAATCGAGGCTCCTAATGGGTAC							
		ACTGTGAAGATTGTTGGTGACACC							
		TCTAAGACTTCCGACATGTTCTAT							
		GTCCAGTATGACAACGTGAAGAA							
		AGTCTGGAAGGAAGTCGCCGGAT							
		GGGGAGTTGAGAAGGGTCTCAAT							
		GGAGGCACTATGCCACACGCATTG							
		GTTCGTCAGTCTGATGGTTCCTTC							
		CAGATGCAGGTGCTACCGTGGGC							
		GCAACGAACTTGTGGTGACATGG							
		ATACTAACCCGACTCCATCCATCG							
		TGGACCAGAAGATTAACGATGTGT							
		TCTTCTTCCGTAACCGCTTAGGGT							
		TCCTCGCTGGTGAGAACATTGTGA							
		TGTCTCGTACCTCTAAGTATTTCTC							
		ACTGTTCCCTGCGTCAGTAGCTAA							
		CCTGTCCGACGATGACCCGATTGA							
		TGTGGCTGTATCTCACAACCGTAT							
		CTCAATCCTGAAGTATGCCGTACC							
		GTTCTCTGAAGAGTTGCTGCTATG							
		GTCTGACCAAGCACAGTTCGTCCT							
		GAGCGCTTCCGGTATCCTATCACC							

		GAAGACAGTTGAGTTGAACCTTAC GACCGAGTTCGATGTGTCCGACCG AGCGAGACCTTTTGGCGTTGGGCG TGGTGTGTACTTTGCGTCACCTCG TGCTTCCTACACGTCACTGAACCG TTACTATGCGGTACAGGATGTTAG CTCCGTGAAGTCTGCTGAGGATAT GAGTGCTCACGTTCCAAGTTACAT TCCGAACGGTGTGTTCTCCATTCG CGGCTCCGGTACTGAGAACTTTAT CTCCGTGCTCTCTGCGAACGCCCC AAGTAAAATCTTCCTATATAAGTT CCTCTACTTGAATGAAGAGATTGC CCAACAGTCGTGGTCTCATTGGGA ACTTGGTAGTAACGTAACGGTTCT GGCTTGTGACTCTATCGGCTCAAC GATGTACCTTGTGTTGCGCAATCA GTCACACACTTGGATGTGCCGAGC ACACTTTACGAAGAACTCCATTGA CTTCAACGATGAACCATATCGGCT GTACATCGACAACAAGGTCAAGC ATGTGATTCCTAAAGGTGCGTACA ACGATGACACCTACACGACCACTA TCAAGCCTGTGGACATCTACGGTA TGAAATATTGGACTGGTAGGTTCT ACATCGTGGCCTCTGATGGTATGG TCTCGTGGTTCGACCCTCCGCGTG GTGGTTGGCCTAATGGTGTCCCTG TGTTGTCAATGAGTGGGAACCGTG AGGATGAGACAATCTACGTCGGCT TGGCTATCAACTTCCGTTATGTGT TCTCTAAGTTCCTCATCAAGAAGA CTGCTGATGATGGCTCGACAGCTA CCGAGGACATTGGGCGCTTGCAAT TACGCCGAGCGTGGGTGAACTATG AGGACTCTGGTGCATTCACTGTAG AAGTGGAGAACACCTCTCGACTGT TCAGCTATGACATGGCGGGTGCTC GCTTGGGTTCCAATGCGTTACGCA CTGGTGGACTTAACGTTGGTACTG GTCAGTTCCGATTCCCGGTTGCTG GCAACGCACAGTTGAACGAAGTA AGGATTATCTCTGACCACACTACA CCACTTAACGTGATTGGCTGTGGA TGGGAAGGTAACTACCTACGACG CTCTTCTAGTATCTAA (SEQ ID NO: 54)							
37	gp13-gp14-gp15-gp16	ATGCTAATAATCAGACCTACTAAA GAATCAGACTTTGCTAGGTTCACT CCGTCACCTGAAGACATCGCTGAG GCAAAAGCCTATGGCATTGAACCT TCATTTCCACCAGCTAACGAGTGC GTCACTATGTCACTTCATGGTATG CCCGTGGCTATCGGTGGTAATTGT GGCGACCAAGTGTGGTTCGTAACG TCCGCTCGTGTTTGGAAACTAAGT ATGAAGACCCGAAGAGAGTTCCG AAAGCTCATTCTGGAGTATCGTGA TACCATGCTCAAACAGTATCCGAT TATCTGGAACTACGTTTGGGTCGG CAACAAGTCCCACATTAGGTTCTT GAAGTCTATTGGTGCGGTGTTCCA TAATGAGTTTACTGGTGACAGTAA TCAATTCCAACTATTCACTATAGG AGGTTAACTATGTGCTGGATGGCT GCAATCCCAATCGCCATGCAAGGT GCGTCTATGGTAATGGGCGGTATG CAGGGTGAACAGGCAAAGGCCGC ACAGATTGACCAAGGTCGTCGAC AAAGCTGGCAGATGATTAAGGAG ATGAACTACAACGATGCTAACCTT AAGTTAGAATCTCGTGACCTCATC	7290	–	Probable scaffold protein (gp13); internal core proteins (gp14, gp15, gp16)	0.29	0.27	0.18	0.32

		GACTCAACTGTCCAAGAGATGACC CAAGCGAATATGAACCGTGTGCG AAACATGGGGACTATCCGTGCGG CAATTGGTGAAGGGATGCTTGAA GGTAACTCAATGGAACGTGTAGCT CGCGTTACTGAGGGTGACTTCCTT CGAGAGTCTCAAGGTCTCACCGAG AACTACCAACGTGACTACAGTGTC ATCTTAGGTAAGCGTTTAGCTAAC CGTGAGAATACCGTAAGTCAAATC AATGAAATCAACAAGTCCGAACC ACGTCTGAAAGGCAAGCTGGAAC AAATCATTGACCCATTAGGCTTAG GCTTAACGAAGCTGGTCGATGTGG CAACTGCTGGTAGCCCTCTGACTA ACAAGCTGGGTAAGAAGGCAGCT GCTAAAGTTAAGGCTTCGGATGCT AAGTCTACCGGACAAGGCAAGTA AACACAGGAGGTAACTGTATGAG TAAGTTAGCACAAGCATTGGGTAA CATGGGTGCTCCTGCTGCAAGCCG ACTCCGTGGTACTGGTCGTATGGA CGTTAAGGCTGCGACTGTCTATGA AGACCCTAGCTATGCCAATAAGTC CGCACTGATTGGGACTGTAGGTAA ACTTGCTGAGATGGGTGCTGATGC GTACATGAAGTATGACCAGAAGC AGAAGGACAAGGCAGATGAGCGC TCCAACGAGATTATCCGCAAGTTG ACTCCTGAGCAACGACGTGAGGC CATTAAGAATGGAACACTTCTGTA TCAGGACGACCCTTACGCGATGGA AGCATTGAAGATTAAGACTGGTCG CAACGCGGCCTACTTGGTTGACGA TGAGGTTGCACAGAAGGTTAAGA ACGGTGAGTTCCGTACTCGTCAAG AGTTGGAAGAGTTCCGTCACAGCC GACTGCAAGAAGCCGCTAAGAAC TATGCGGAACAGTTCGGTATCAAT GAGTCAGATGAGTTTTATCAGAAG GGTTTCAACTCCGACATCACTGAA CGTAACATCGCCATCTATGGTGCT CATGATAACTTCCTGAGTGACCAA GCTAAGAAGGGTGCCGTTATCAAC AGCCGTGTCGAACTGAACTCTGTA CTCAACGACCCTGAGACCCTTCGT TCCCCATATGCTGGTGAGTTCTTT GAGAACTACTTCAGTGCTGGATTG ACGACTGGTAGTATCCCTAGTGAT GACCAAGCGTTCACAATGATTAGC CAAGGTCTGTCTGACGTGGTTAAC CGTGAAGGTGGCGGTCAGTTCCTA CAACAGATTGAGAACCGTAAGGT GAAACTCCACGGAAAAGAGACAA CCTTCAAGGAACTGATGGGTGCCG AACAGTGGAACAATCTGATGGTTA AGGCTCAACATAATGAGTTCCAGT TGAATGCTAAGAAGACCGAGGCG TTCCAACTTAATGTGAACTCTGCG TTGAACCAAGAGAACGTGAACAC TGGCTGGGAACAGTTGCAATCCAT CAAGGCTGAACTGGACACACTGC AACCGGGCGAAGAGATGACACCT GAACGTCAGGCGTTAATCAATGCA CAGACGCAGATGCAGGACAGGAT GAAGCGGGAGACTGCTGAGTTAG CAAAGCAGATGGACAAGCAACAG AAGTCGATGAACAAGATGAACGT CATTGACGCACAGTTCCAGAAGCG ACTCAATGGTCAGTATGTGTCTAC TGCGTATGGCGATATGCCTACCAA CGAGAACACTGGTGAGTTCACCCA							

		TAGTGACATGGTTAACTATGCGAA CAAGAAGTTAGCCGACATTGACG CTATGGACATCCCTGCTGAACAGA AGGACCGAATGAAGCTGGACTAC CTCAAGGCTGACTCCGAGAAGGG TGCTTTCCGTACTGCGGTGGGTGA ACTGATTGGTGACGCTGAGAAGG AATGGACGTCCGCTGTGATTAATG GGAAGATGCCTGAAGATGGCGGT GTTGCCCTTAACGCTCTTCGTCGT GTCCGTAATGCTGACCCTGAGTTA TTTGCTGCGCTGTATCCTGATAAA GCCGAGATGTTCCTCACTATGGAC ATGATGGACAACCAAGGGATTGA CCCGCAGATTCTCTTAGATGCAGA CAAGGCTCGACAATCTCTGACCAA GGATATGCAGTATGAGGACGACA AGGCATGGGCTGCTCTCAAGAAC AACTCACAGTCTCCTGAACTGTCC CGTATGCCAGCCACCTTGGATGCT GGTGCTCGTAAGATTTATGACTCT GTGAAGTACCGTACTGGTAACGCT GACATGGCAATGCAACAGGTCGA TAAGTATCTCAAAGAGAACACAA CGACTCTCACTGGTGATGATGTTG ATGGCGATACCATTGGTGTCTTGA CGAGAAACTCTCTGCGTGTGACCG ATGACCCGGACTCTTGGAAACAA GGTAAGGACATCATCGACCAAGC CGCTAAGAAACTCGCTGAGACGA ACCCTTGGGTGACTAATAAGCAAC TCACTGTGTTCGAGCGTGGTGACT CTATCTACCTGATGGACACTACCG GACAGGTTAACGTAAGGTATGAC AAGCAGTTGCTCTCGAAGATGTAC CAAGAGAATCAGGCCAAGTTGGA TGCGGAAGCACGCGACAAGGCAC TCAAAGATGCCAACAAGCGAACC CTGCATACACGAGCTATGAACCGT AAGCGTGAACGTGAAGCCAAGAA ACCTAAGCGTTCTGGTAGCATGTA TGATTCGGTGAGTGGTAAAGGTAT TCTGGACACGCTGACTGGTAAAGA CTAACAGACCACGATAGGAGGTT CCAAGTGGATAAGTATAACCAGA ACGAACCGCATGAATATGATGCGT TATTTCAACAAGCTGCTGACACAC ATGGAGTCTCCTATGGCCTTCTTC GGAAGGTCGGTTGGGTAGAGTCC CGCTTTAAGCCTACGGCTCAATCA CCAACAGGGCCGAGAGGTGTCAT GCAGTTCACTAAGGCAACTGGTCA GGCTTATGGTCTCCAGAACGATGA GGACTTCGCTGACCCGGCTAAGTC CATTGATGCAGGTGCTCGTTACCT TGCAGACCTCGTTAAGAAATACAA TGGGGATGAACTTAAAGCTGCCCT AGCGTACAACCAAGGTGAGGGAC GTAATGGTAAACCCCAATTGGAA GCCTACGACTCTGGTAACTTTGCT GGCATCGGTGATGAGGGGCGCAA CTATCTCCGCTCACTTCTGGATGT CGCTAAGTCTCCTAAGAGTGGTGA CATTGAGTCATTCGGGGGTATCAC CCCAAAGGGTAAAGGGATTCCGTT CGATACTGCTATGAGTGGTATCGG GAAGAAAGGAAAGGTAACTACAG AACTCCCTGAGTCTCACTCAATGT CCTTTCAAGGTAAAGAACAGTCCG CTCCCAATCAGCCGTTCGGAAAAG ACTATTGGGAAGCAAAAGGAACT ACACTCGACGAAGCTAATGAGCG							

		TTCAACCTTCTTCGGATTCGGTAA TGCTACTGAAGCAGAACTCTCCAA CTCAACCTTGGGTGTGGCCTTCCG TGCTGGTAAGCGTGATAATGGTTT CGATGTTGTTACTGATGTACTTCA ACCGACACGGTTTAACAGCCACAT TTGGTCGCCTGAAGAACTCGAAAG GATTCGCAACGAAGTGAAGAACC CTGCGTACATGAACGTAGTGCTTG GAGGTTCCTCTGAGAATCTGGATG AACTCATCAAGTTAGCCAATGAGA ACTTCGAGTTGGACTCTAAGGCCG CTGATGCTGGTATTGGTGCCAAAC TGTCTGCTGGCCTTATTGGTGCTG GTGTTGACCCATTGACCTACGTTC CTATCGCTGGGACTGCGGCTAAAG GGTTCAAGCTGGTTAACAAGGCAT TGATTGTAGGTGCTCAAGCTGGCG CTCTTAACGTGGCCTCTGAAGGAC TCCGCACATCGGTTGCTGGCGGTG ATGCTCACTATGCTGAAGCTGCCC TTGCTGGTATGCTGTTCGCTGGTG GCCTTACTGCTATCGCTGATGGTG TGGCTGCTGGTCTCCGTAAGTCCG GTGCTGAACAGATTGAGAACCCAT TCGCTGCTGCACAGATGCGCTTTG AGGCTCGTGAGACTGCACGTAAC ACTGGTGGTTATGACGCAAGCCGT ATGCCTCCTAGTGAAGACCGAGTG TTCTCGAATCACAATGGCGTTGAG TATTCACCTCTTGAAACTGAACCG GGTGCCGTAGTGTTGCGTGATGGT TCCATTATCAGTGATACAAACTTG GCTAACCCGATGACAGCCAAAGA GTTCGCAGAAGTAGACCCTGAAC GTGCTGCGTGGGGCTTGCCTATGC GTGGACTGAGTGAAATCGGCTTGA GGACTTTGCGTTCTGAACACGCTG AAATCCGTGGTCTCGCTAAGGACT TAGTGCGTTCACCTACAGGGATGG AGTCTGGCTCACATGGTAAGTTCG GTGCTACGGCCTCTGACATCAAAG AGCGTCTGCACAGTACCAACCAGC GAACCTACAATGACCTCTATGATG CAATGAAGGAAGCAATGGCTGAC CCTGAGTGGTCAGTGGGTATGTTC AAGCGTGGTGCTCAAGGTGCCCGT CAGGAAATCTATCGTCGTGCTGCT ATCGCTATTGAGCGTCCAGAACTT CAAGCCAACTTAACGAAGGCTGA ACGTAAGGTAATGGACATCATGA AGGAACATTTCGACCTCAAGCGTG AGATGATGGAGAACCCGGCTATG TTCGGTAACAAGGCAACTTCCATC TTCCCTAACAGTCGTCACAAAGGG ACTTACGTTCCTCATGTGTACTCA CGGGAAGCCAAGCAGTTGTACGC ACAGGCTCTTGGTGGCTCTGACGG TTTACAGGAAGCTATCGCTGCAAG CTGGATGACCTCTTACCGTTCACG CCCTGAAGTTAAGGCTCGTGTGGA TGAGCATTTAGCGGAGACATTGGG TATCGACCCGAAGGCTGTAACCGA AGAGATGGTGATGAAACATGCCA GCGACAAGGCTTATGGTATTGCGA AGACTGATGAGTTCAACTCTTCGA GTGTCATCGACGATAACATTGAGG GTCTCGTTGGGATTGAGAACAACT CGTTCCTTGAAGCCCGTAACCTGT TTGACTCTGACATGCCAGTCACTC TACCTAATGGACAGCAATTCTCTG TGAATGACCTCCGTGACTTCGATA							

		TGAAGCATGTGATGCCAGCCTATG ACCGCCGTGTTGATGGTGACATTG CTATCATGGGTGGTACAGGAAAG ACTACTGCTGAACTCAAGGATTCC ATTATGGCTCTCGATAAGAAGTCA GAAGGGAAGGGTACGATGAAAGG TGAAGTTGAAGCACTCAAGGATA CCGTTAAGATTCTCACAGGTCGCG CTCGACGCAATCAGGATACTATGG GTGACACTATGGTTCGCGCTCTGT CCGATATGTCCTTCTTTACGAAGA ACGCATACATGGGCTTGCAGAACC TAACTGAAATCTCTGGTCTCTTGG CGAAGGGTAACACCCGTGCGATG CTGCATGGCATTCCTGCTCTGCGT GACCTAGCATTCCGTAACAAGCCA GTGTCAGGTAAGGAACTTAAGGA ACTACACTCTATGGTGTTCGGTAA AGAGTTCGATGACTTAATCAGACC GACACGTCAGGACATCATTCAGCG ACTCCGTGAATCTACGGATACACC TGATGTTGCCGCTAAGGTTGTTGG TACGGTTAAGCACACCACTCAAGA GTTTGCTGCTCGCTCTCCGTTCACT AAGTTCCTGAATGGTACATCAAAC TACATTCTCGACATGGCTCGACAG GGTGTGATGGGTGATGTGGTCACT CATGCTATCACTGGCAAGGGTGCC AACAAGTGGATTAAAGGCGACAT GCTGAAGTCTGCGAGTATCTCAAA GGAACAGTGGGAAGGCATTCAGA ACCTCATCCGTGAGAACGTCACTC AAGGTGCTGACGGTAAGTACACCT TCAAGGACAAACGCAAGCTGGCT AATGACCCTCGTGCTATGGACTTG TGGCGCTTAGCTGATAAGGTTGCC GATGAGACAATGCTGAGACCTCA CAAGGTTTCACTTCAGGATTCCCA TGCGTTTGGTGCTGCCGCTAAGTT GGTCCTTCAGTTCAAGTCTTTCGT AATCAAGTCGATGAACTCTAAGTT TATCCGCTCTGGTCACGAGGCTAT GAAGAATCACCGAGCTATGGACA TGGCATTAACCTATGCAATCTCTG GAGGCATCGCTGGTGCTTACTATG TTGCACAGGCACACATGAAGGCT GCTGGTCTACCTAAAGAGCAACA GAAGGATTACTTGAAGAAAGCTCT GGACCCTAAGATGATTGCGTATGC TGCTGCATCCCGTAGTTCACATTT AGGTTCCCCGTTGAGTATCGCCAA CTTCGCTATGGGTGCTGCTGGTTA CGACCAAGGTCTCATGGTGCGCTC CACGATTCTACCGAAAGGAGAAG ACAAGCGTGACCGGAATAAAGCA GTGACCTCTCGTGATATGGGAGAA TCCATCATGGGCGCTATTGGTGAA CAGATTCCTGCTCTTGGTTTCGCT GGGGCAACTCTTGCTGCTGGTCGA AATGCTTATGGTGTCCTAACGGCT CCTAACAAAGTGACCGAGCGTGA GATGATGACTGGCTTGATGAATGC TCATCGTGAGATGGTTCCTAACGA CCCTATCACTCAACAGATGCTCAT TAAGTTCTATGAGGCTAACGGTGT TCACCTTAAGTCTGACAAGAAGTA A (SEQ ID NO: 55)							
38	gp17	ATGGCTAACAAGATTCGTACTGTG ATGACTTACCCACTTAATGGTGCC GTGAACTTCACGATTACCTTTGAG TATCTGGCACGTAAGTTCATCCAA GTGACACTCATTGGTAAAGACCGT	1968	–	Tail fiber protein	0.45	0.44	0.66	0.40

		AAAGAACTGGTTCTCAATCAGGAC TACCGTTTTACCAGTAAGACTCAA ATTACTACTACCCGTGCGTGGACT GCTGCTGATGGCTACCAGATGATT GAGATTCGTCGATTCACTTCGGCT ACTGACCGTCTGGTTGACTTTGCC GATGGCTCAATCCTCCGTGCGTAT GACCTGAACATCTCGCAGATTCAA ACACTGCATGTTGCCGAAGAAGC ACGAGACCTCACTGCGGATACCAT TGGTGTCAACAATGATGGTCACTT AGATGCTCGTGGTCGTAAGATTGT TAACTTGGCGTATGCCACAAGTGA CTATGATGCCGTTCCTTTAAAGCA GGTTAAAGAGCGTGAATCCTCTGC ATGGAATGCTGTGACAAAAGCTA ACGAACACGCTGACAGGGCCAAT AAGGAAGCAAACCGTTCTCGTGAT GAGGCCAACCGTGCCCAACGTGA AGCTGACCGCTCAACGCAACAGG CAGGTGTCTCTACACAACAAGCAG TTGAGGCCAAGAAGCAAGCTGAC CGTTCTAACTCTGAGGCTAACCGA GCTAAAGGTTATGCTGATAGTATG ACCGCAAGTGTCGCTGAGGCCAA GAAGCAAGCTGACCGTTCCACCTC TGAGGCCAATCGTGCCCGTGATGA GGCCAATCGTGCTGCCGGAGAAG TGACTAAAGCTGCTGCTGAGGTTG CCAAGGCTGCTGCACACGTTGAGA CCGCTAAGAGTCACGCCAATCGTG CTAACACTGAGGCAGAACGCTCTA AGAGTGAAGCCGACCGTGCTAAA ACCGAGGCTGACAAGCTGGGTAA TATGAATGAATTAGCTGAGGCCAT CGACAGGGTTGATGCTCCTGATGT TTACTTCAAGAGGAATGTTATAGC CGGACACAAATTGGAGACTTCTGG AGGCTCGCATCTTCACCCTAATGA GATTAAATTGGTTAACAATAAATC TCCAGACCAGCACTTCATTGTACA AGCGTGGGGAACTGGAGACCCTA ATAATAGGTCGCAGGTGTTTGAGT TAAAGGATAACTTGGGTCCGTACG CATTCTATATCCAGCGCCAGAAGA AGACAGACCAAACTTGGCAAATG AGTGTGAACGGTACTGTTAGCTGC GTATCCATCGAGGCTGGCAAGGGT TTCATCTCTAAGGACAGTACAGGC TCCATTCAGATGAAGACTTGGGGG GGTGGTTGGAACCAATGGACTTCT AATGCTGTTCCAGACATTCAGCAC GATATTGAGGCCCAACATGATGCT GCTTATCTTCTGGAGAGGGCTACC ATTTGGGGTAAGGAGCATTTATAC GCAAAACAACTGTACAATAATGG CACTGGCGCAAACTTCTTTATTCG CCACTACCTACGAGGCTCTTACTT CGACTTCCGCTCTGATGGTACTCT TATCAGTTCCACCGGATGGAATAT CCCGGCAGATGGAAACCTGTACAT CAAGAAATATGGACTATATATTGA TGCGTGGGTTAACAAGCGCTTGGG CGAACACGCCTACAACAAGAGCG AGACAAACAACCTAGTGAATGGT CGATTGACACAAGCGCAAGCTGA TGGTCGCTACATTCGCAAGAACGC TGGTTGGGCTGAGGTCTGGAGCGG CTCTGCTGGTGGTGGTGTTTCCGT AGGTCTCTCACAGGATGTTCGTTG GAGAACCATCTGGATTCAAACTCA AGGCAGATGGAACCCTGTTCAGAT							

		TGGTGATAATGGTGCTTACTACAT CTCGTCTAATGGTGGGTGGTTGAA GTTCACCATCTCTAATAACGGTAG GACGTTCCGTAACGATGAGGACA GGACTTCTGTTCCTACGAGAATCT TGATTCAAAACCAATAA (SEQ ID NO: 56)							
39	gp17.5-gp18	ATGTTGTCATTGGATTTTAACAAC GAGGTAGTGAAGGCTGCTCCGATT GTAGGGACAGGGGTTGCCGATGG TGCTTCCAGACTGTTCTTCGGCCT ATCACTTAACGAGTGGTTTTATGT AGCCGCAATTGCCTACACAGTGGC TCAAATTGGTGCCAAGGTAGTCGA CGTGGTTATCAAATGGAAGAAGG AGGGTAGAGATGTCGGATAAATC TTTGATTCAATTCCTTGAGATGCT GGACACTGAGATGGCTCAACGTAT GCTTGCTGACTTGCAGAACGATGA ACGCCGTACACCTCAACTGTACAA CGCAATCGGGAAACTGTTAGAAC GCCACAAGTTCCAAGTCTCCAAGT TACAGCCAGACCGTAACATCCTTG GTGGACTGGCAGATGGACTTCAG GAGTATCAGAACCTAGTAGGTGA CGATGGTCTCACTGAAGATGAGA AGTACACGCTGAACTGA (SEQ ID NO: 57)	466	–	Holin; small terminase	0.060	0.060	0.10	0.059
40	gp18.5-gp18.7	ATGCTCAAGAATTTAAAGAGCTAC GTCATAATCGCTGCGTTTGCCTTG GGGATGACCCTTGCGTACAACCAC GGCTATGACAAGGCTAACACTAA ATGGGAACAGGAGGTGCACAATG AGTACGTTAAGAAAGTTAAGGCG ACTGAGGATAAACAGGCTGCTATC AACGAAGTCTCACGACAATACCA AGAAGACCTTGCGGCGCTGGAAG GCAGCACTGATAGGGTCATTAATG ATTTGCGTAACGATGGTAAGCGGC TGCGTGTCAAACTATCAGCCACCT CACGAGAACTCCAAGCTAACGGT GGATGCCTCGTTGATGGTCGAGCC GAACTTGACGAAGAGTTTAGTAA GCGTCTTATCGGAGTAACACAAAG AGGGGATGCGTGGATTAAAGCGC TCCAAAATACAATAACTGAAATGC AATCTAAGAAGGAGAAATAA (SEQ ID NO: 58)	444	–	Spanin, inner membrane and outer lipoprotein subunits	0.046	0.054	0.16	0.077
41	hypothetical protein	ATGGCTACTACAAACTCAAGAACC CGCACACTCTGGACCATCAATGGT CGCTCACCGGACAGCCGGGATAT GTACGACCGAAACAATGTTGGCCT TCGTGTTGGGCGCATCGACATCCC AGCGTCAGAAATGGTCAAGGTAG ATGGCTACGACAAGGTTGAATAC GTGGTGCAACTGAGTGATGGCAC AACCGAGCAACATGCTACGTTTGA GACCTTGATGTCCACTTGGGGTGG CTCCGTAGAGTCTAAACCGGGTAG TGGGTTCATGACGGTATTTCTGAA TGAGCAAAGCACTGAGCGGCTCG GTATGGTTCGAATCGACGTATACA AAGAGTGGACTGGTGCTGACCCTG AGCCGGAAGGTGGTGTGGCTTCAT TAGTTAGTGCAACGCTTTCTGCGT CTGATAGTGCATTTGGTACACAGT TCGTGGATAGCGAAGGGAAGAGC GCTTCCGTTATGTTCCAACCTAGT AACATCGACGGGGCTGTATCTAAG ACCGACACTATCAAGGCTCTTGAG AAGCGCATCGCTGCCTTAGAGAA ATCTGCCGAGTAA (SEQ ID NO: 59)	558	–	Hypothetical	1.0	1.0	1.1	1.2

42	gp19	ATGTCACAATCCCAAGAAGCTAA GAATGCGCTCATCATTGCGCAACT CAAGGGTGACTTCGTGGCCTTCCT GTTCGTCTTATGGAAGGCTCTGAA CTTACCTAAACCAACCAAGTGTCA GATTGATATGGCACGAACACTCGC TAATGGAGACCACAAGAAGTTTAT CCTTCAGGCGTTCCGTGGTATCGG CAAGTCCTTTATCACCTGTGCGTT CGTTGTGTGGGTCTTATGGCGTGA CCCTCAACTTAAGGTGCTCATTGT GTCTGCCTCTAAGGAACGTGCGGA TGCTAACTCCATCTTCATCAAGAA CATCATCGACCTCTTACCGTTCCT CTCTGAGTTGAAACCTCGACCGGG ACAGCGAGACTCCGTGATTAGCTT TGATGTTGGCTTGGCTAAACCCGA CCACTCACCTTCTGTGAAGTCTGT GGGTATTACTGGTCAGTTGACTGG TAGCCGTGCTGACATCATTATTGC AGATGACGTTGAGGTTCCCGGTAA TAGCTCTACGAGTTCTGCCCGTGA GAAGCTCTGGACGTTGGTAACTGA GTTTGCTGCATTGCTTAAGCCGCT GCCTACCTCTCGTGTTATCTACCTT GGCACACCTCAAACCGAGATGAC GCTCTACAAGGAACTTGAGGATA ACAAAGGGTACAGCACAGTAATC TGGCCAGCACAGTATCCTCGTAAT GATGCTGAGGCTCTCTACTATGGG GACCGCTTGGCTCCGATGCTTAAG GCTGAGTACGACGAGGGCTTTGA GTTACTCCGTGGTCAACCTACTGA CCCTGTTCGATTCGACATGGATGA CCTCCGTGAACGTGAACTCGAATA TGGCAAAGCTGGCTACACTCTTCA GTTCATGCTTAACCCGAACCTTAG TGATGCCGAGAAGTACCCACTACG ACTCCGTGACGCTATCGTGTGCGC TGTGGACCCTGAACGTGCCCCATT GTCTTACCAGTGGTTGCCGAACCG TCAGAACCGCAATGAGGAACTCC CTAACGTTGGTCTGAAGGGTGATG ACATCCATGCGTTCCATACGTGCT CTTCTCGTACCGCTGAGTATCAGT CTAAGATTCTGGTCATTGACCCAA GTGGTCGTGGTAAGGATGAGACC GGGTATGCAGTCCTCTACTCACTG AACGGTTACATCTACCTGATGGAA GTCGGAGGGTTCCGTGGTGGCTAC GATGATGCTACCCTAGAGAAGCTC GCTAAGAAGGCCAAGCAGTGGAA GGTACAGACAGTCGTCCACGAGTC CAACTTCGGTGATGGTATGTTCGG TAAGATTTTCTCTCCGATTCTCCTG AAGCACCATAAGTGTGCCTTAGAG GAAATCCGTGCTAAGGGCATGAA GGAGATGCGAATCTGCGATACTAT CGAGCCGTTGATGGGTGCACACA AGCTGGTCATCCGTGATGAGGTCA TCCGTGAGGACTACCAGACTGCCC GTGACCTTGATGGGAAACATGATG TCCGCTACAGTGCGTTCTATCAGA TGACACGTATGACCCGTGAGCGA GGCGCTGTGGCACACGATGACCG CATTGATGCTATCGCATTGGGTAT CGAATATCTCCGCGAGGGGATGCT TGTGGACAGCCGTGTAGGTGAAG AAGAGATGACCCTTGAGTTCCTTG AACACCACATGGAGAAACAGACC ATTGGTGGCGACCAGATTCATAGC TTTGATGTTGGGGGTGTGGACATC	1761	–	Large terminase	0.11	0.11	<0.0001	0.11

		TACTATGAGGATGAGGACGGAGG TTCTAGCTTCATTGAGTGGTAA (SEQ ID NO: 60)							

Avs proteins are pattern recognition receptors capable of recognizing a diverse range of terminase and portal proteins

[0629] To investigate the specificity of Avs activation, Applicants cloned the portal and large terminase subunit genes from 24 tailed phages, spanning nine major phage families, and co-expressed these genes in E. coli with 12 Avs systems spanning all four Avs families (**Tables 1-2** and **Fig. 4A-4B**). Applicants quantified cellular toxicity and depletion of specific Avs-phage protein pairs for all 576 combinations by deep sequencing (**Fig. 7A** and **Data S2**). These experiments revealed precise target specificity: Avs1-3 were activated only by large terminase subunits, whereas Avs4 recognized only the portal protein (**Fig. 7B** and **Fig. 8**). To assess the robustness of the assay, Applicant repeated these experiments varying the Avs promoter or the amount of terminase and portal induction, obtaining similar results (**Fig. 9A-9B**). Surprisingly, Avs1 and Avs2 also recognized terminases despite the lack of substantial sequence similarity among the C-terminal TPRs of Avs1, Avs2, and Avs3, although Applicant detected a structurally similar domain at the end of the TPR arrays in all three proteins (**Fig. 10A-10C**). These findings demonstrate remarkable conservation of target recognition across Avs families and suggest that the portal and large terminase subunit are key PAMPs recognized by prokaryotic NLR homologs. Moreover, Avs systems recognize PAMPs from diverse phages; for example, SeAvs3 and EcAvs2 are strongly activated by 20 of 24 and 19 of 24 tested terminases, respectively, and EcAvs4 is strongly activated by 15 of 24 tested portals (>100-fold depletion) (**Fig. 3F**). Because the portals and terminases from different phage families have limited sequence similarity, with less than 5% pairwise sequence identity in some cases (**Fig. 7C**), but share the same core fold (**Figs. 11-13**), this broad range of activity implies that Avs proteins are triggered by conserved structural features rather than by specific peptide sequences. Consistent with this idea, EcAvs2 and EcAvs4 displayed weak but clear recognition of the terminase and portal, respectively, of human herpesvirus 8 (**Fig. 7D**), which is a highly diverged evolutionary derivative of tailed phages (24) and does not infect prokaryotes.

SeAvs3 and EcAvs4 are phage-activated DNA endonucleases

[0630] SeAvs3 and EcAvs4 contain predicted N-terminal PD-DExK-family 139 nuclease domains (**Fig. 14A**), which Applicant hypothesized degrade phage and cellular DNA upon target recognition. The nuclease domain of SeAvs3 is most similar to the recently reported

Cap4 effector nucleases of cyclic-oligonucleotide based defense systems (25, 26), whereas the nuclease domain of EcAvs4 is related to Mrr-like restriction endonucleases (27). Both Avs proteins contain conserved D-QxK or E-QxK catalytic motifs (**Fig. 14B**), and in addition to the STAND NTPase gene, the SeAvs3 system contains a small ORF, the deletion of which reduced anti-phage activity in E. coli (**Fig. 15A**). Applicant sought to biochemically reconstitute Avs activity in vitro, and Applicant purified recombinant SeAvs3, the protein encoded by the small ORF, EcAvs4, and the PhiV-1 portal (gp8) and terminase (gp19) proteins (**Fig. 16A**). Applicant incubated SeAvs3 and the small ORF product with linear double stranded DNA (dsDNA) and observed progressive degradation of the substrate in the presence of gp19, but not gp8 (**Fig. 14C-14D**). This nuclease activity was dependent on the catalytic residues of SeAvs3 but did not require the small ORF product (**Fig. 14C**). Applicant further investigated the substrate specificity of Avs systems and found that the nuclease activity was specific for dsDNA, whereas ssDNA and RNA were not cleaved (**Fig. 16B**). Moreover, SeAvs3 cleaves both linear and circular dsDNA, including E. coli genomic DNA (**Fig. 16C**), indicative of endonuclease activity with no specificity for phage DNA, which is consistent with an abortive infection defense mechanism.

[0631] Applicant next investigated cofactor requirements of SeAvs3 and found that in vitro activity depends on both Mg^{2+} and adenosine triphosphate (ATP); however, ATP hydrolysis is not strictly required because nuclease activity was observed at a reduced level in the presence of the nonhydrolyzable ATP analogue AMP-PNP (**Fig. 14E**). Applicant also reconstituted the nuclease activity of EcAvs4 and found that it is activated by gp8, but not by gp19, and was abolished in a Q63A/K65A (Gln^{63}→Ala/ Lys^{65}→Ala) EcAvs4 mutant (**Fig. 14F-14G**). Similar to SeAvs3, nuclease activity required the presence, but not the hydrolysis, of ATP (**Fig. 14H**), consistent with phage plaque assays of SeAvs3 and EcAvs4 ATPase active-site mutants (**Fig. 15B**). Together, these experiments indicate that SeAvs3 and EcAvs4 are promiscuous DNA endonucleases that are activated by distinct, highly conserved phage proteins in an ATP-dependent manner.

Structural basis for Avs binding and target recognition

[0632] To investigate how Avs systems recognize and bind their cognate phage proteins, Applicant solved cryo-electron microscopy (cryo-EM) structures of the SeAvs3-terminase and EcAvs4-portal complexes in the presence of ATP and Mg^{2+} (**Figs. 29A-29F, 30A-30B, 31A-31C,** and **32A-32C** and **Table S1** of Gao et al., Science, 377, eabm4096 (2022), which is

incorporated herein by reference as if expressed in its entirety herein). A reconstruction at 3.4-Å resolution revealed that SeAvs3 forms a tetramer, with each C-terminal TPR domain gripping the ATPase and nuclease domains of the gp19 terminase (**Fig. 27A-27B**). These TPR lobes are flexible and required symmetry expansion to improve their local resolution to 3.4 Å (see Materials and Methods discussed herein and **Figs. 30A-30B, 31A-31C,** and **32A-32C**) For EcAvs4 bound to the PhiV-1 gp8 portal (**Fig. 27C-27D**), image processing revealed equal abundances of a teterameric complex and an octameric complex corresponding to tetramer head-to-head dimerization (**Fig. 29A-29F**). At lower protein concentrations, however, Applicant observed only the tetramer, indicating that it most likely the functional complex (**Fig. 29A-29F**). Negative-stain and cryo-EM imaging of SeAvs3 and EcAvs4 in the absence of phage proteins revealed only smaller monomeric particles (**Fig. 29A-29F**), indicating that phage protein binding is required for the assembly of SeAvs3 and EcAvs4 into tetramers.

[0633] Tetramerization of both SeAvs3 and EcAvs4 is mediated through their STAND ATPase domains, which interact in a manner distinct from each other and from other characterized STAND ATPase oligomers like Roq1 resistosome tetramer (59) or the Apaf1 apoptosome heptamer (60) (**Fig. 33A-33B**). The SeAvs3 ATPase domain forms a C4-symmetric tetramer by interactions between the nucleotide-binding domain (NBD) and winged-helix domain (WHD) subdomains and the NBD subdomain of the adjacent protomer, whereas the EcAvs4 ATPase domain forms a C2-symmetric dimer of dimers with a tighter interface (1232 $Å^2$ of buried surface area, compared to 436 $Å^2$ of for SeAvs3), with adjacent WHDs and NBDs both interacting (**Fig. 33A-33B**). The smaller interface in SeAvs3 is compensated for by additional contacts between its C-terminal TPR domains (**Fig. 27A-27B**). SeAvs3 and EcAvs4 both maintain ATP in their active sites with and adjacent magnesium ion coordinated by the canonical Walker A and B motifs (**Fig. 27E-27F**). Notably, in both cases, tetramerization of the STAND ATPAse domains brings adjacent N-terminal nuclease domains close together, forming two nuclease dimers with overall C2 symmetry (**Fig. 27G and 27L**).

[0634] SeAvs3 and EcAvs4 contain nuclease effectors of the PD-DExK superfamily. Conventional PD-DExK nucleases (e.g., restriction endonucleases) use a pair of acidic residues to coordinate at least one metal ion and a conserved lysine residue to bind the scissile phosphate and stabilize the transition state for nucleolytic cleavage (**Fig. 27J**) (60, 61). In the SeAvs3 Cap4 tetramer, this arrangement of residues is found in all four promoters; however, in the two "outward-facing" protomers, an extended β strand makes a steric block for DNA binding

and/or metal coordination (**Fig. 27I**). **Fig. 27H** shows Cap4 nuclease inward-facing active site. Based on the crystal structure of the Hind III restriction endonuclease (62), the inward-facing protomers can be predicted to form a cavity for DNA binding, with each protomer likely cleaving opposite strands of the DNA (**Fig. 27K**). The EcAvs4 Mrr tetramer shows a similar principle, whereby the two inward-facing protomers contain active sites that resemble canonical PD-DExK nucleases, but in the outward-facing protomers, Glu^{49}, which is part of the conserved trio of active-site residues, is displaced (**Fig. 27M-27N**). Glu^{49} is found in the loop that spans residues 33 to 52, and interactions between this loop on an "inward" protomer and an adjacent "outward" protomer likely stabilize its position in the inward protomer. Like the SeAvs3 Cap4 tetramer, these two inward protomers form a cavity that accommodates DNA in a manner similar to HindIII (**Fig. 27K**).

[0635] SeAvs3 and EcAvs4 both contain extensive TPR domains for binding their cognate phage proteins, which we confirmed using a bacterial two-hybrid system and protein copurification (**Figs. 17A-17F, 17I-17K**, and **19A-19G**). The SeAvs3 TPR domain forms a left-hand-like structure capped by a b sheet–rich C-terminal domain (**Fig. 28A**). This domain has two cavities in which the terminase ATPase and nuclease domains are nestled. Consistent with the ability of SeAvs3 to bind terminases with little sequence similarity (**Fig. 7C**), there are few specific residue-residue pair contacts between SeAvs3 and the PhiV-1 terminase. Instead, binding is deter-mined by shape and charge complementarity between the two proteins, burying more than 3700 Å2 of solvent-accessible surface area. This complementarity is maintained across a diverse range of experimental structures and AlphaFold models of phage terminases (**Fig. 28B**). Additionally, SeAvs3 directly recognizes residues within the two terminase active sites. In particular, Asp^{1710} in SeAvs3 forms a salt bridge with the highly conserved Arg^{61} within the Walker A motif of the terminase ATPase (**Fig. 28C** and **28E**). An arginine in this position is found in most terminase ATPases that activate SeAvs3 but not in nonactivating terminases (**Figs. 28E** and **34**). These observations suggest that Arg^{61} within the Walker A motif is a determinant of recognition specificity, and indeed, mutation of the cognate arginine in the T4 terminase ATPase domain substantially reduced SeAvs3 activation by the ATPase domain (**Fig. 28F**). Notably, an arginine is not typically found in this position in endogenous cellular ATPases (62), suggesting a possible mechanism for avoiding off-target activation. Furthermore, Arg^{1196} and Lys^{1198} in SeAvs3 form salt bridges to the four conserved aspartates that make up the active site of the terminase nuclease (**Fig. 28D-28E**), and mutation

of Asp^{365} in the PhiV-1 terminase nuclease notably reduced SeAvs3 activation by the nuclease domain (**Fig. 28F**). Thus, SeAvs3 directly reads the active-site residues of both domains of the terminase. Furthermore, the ATP ligand bound by the terminase is detected by interactions between the gamma phosphate and His^{1770} and Tyr^{1714} of SeAvs3 (**Fig. 28C**). Targeting this ligand presumably helps avoid phage escape mutations, because ATP binding is required for the function of the terminase.

[0636] Because SeAvs3 detects both domains of the terminase, Applicant hypothesized that there might be some functional redundancy in these inter-actions. Indeed, SeAvs3 was activated by the nuclease domain alone from some phages, including T7, but was also activated by the ATPase domain alone from T4 and ZL19, a T1 family phage (**Fig. 17L**). Likewise, SeAvs1 was activated by the nuclease domain from T7, but in the case of ST32, both the nuclease and ATPase domains were required. These results suggest that Avs1 and Avs3 recognize both the nuclease and ATPase domains but differ in the extent of activation by either domain, depending on the terminase. By contrast, deletion of the nuclease domain had no impact on Avs2 activity for any of the five tested terminases, suggesting that Avs2 recognizes the ATPase domain only (**Fig. 17L**). This pattern of recognition is consistent with the larger size of Avs1 and Avs3 compared with Avs2.

[0637] The TPR domain of EcAvs4 also binds the PhiV-1 portal with a large interface, burying 5800 Å2 of solvent-accessible surface area, that includes notably few residue-residue contacts (**Fig. 28G**). The portal protein is recognized through its stem, clip, and part of its wing domain. In an assembled dodecameric portal complex, these regions are found toward the interior and are therefore more constrained in their fold requirements (**Fig. 28I**). Consistent with this observation, Applicant performed random mutagenesis by polymerase chain reaction (PCR) to screen for portal mutations that abrogate Avs4 activation and found that all 29 identified mutations were nonconservative and located in the core wing or stem regions (**Figs. 17G-17H, 18A-18C**), possibly disrupting the core portal fold. The clip domain, which contains a con-served antiparallel b sheet with an intervening α helix, is recognized by β-sheet augmentation with a hairpin of EcAvs4 (**Fig. 28H**), a mode of fold recognition that does not depend on the amino acid sequence of the target. Because portal proteins are not enzymes, there are no active-site residues to target as in the SeAvs3-terminase complex. Finally, portal oligomerization is not compatible with the Avs4-bound state (**Fig. 28H**),suggesting that Avs4 recognizes portal monomers before they assemble into the procapsid.

Avs proteins are widespread in prokaryotes and possess diverse, modular N-terminal effector domains

[0638] To assess the diversity of avs genes in prokaryotes, Applicant collected all intact homologs from each of the four families present in the National Center for Biotechnology Information (NCBI) nonredundant protein sequence database (**Tables 3-6** herein and **Data S8** of Gao et al., 2022. "Prokaryotic innate immunity via pattern recognition of conserved viral proteins" Science 377, eabm4096 (2022), which is incorporated by reference as if expressed in its entirety herein. The avs genes were identified in approximately 4-5% of sequenced prokaryotic genomes and are broadly distributed across phyla (**Fig. 20A**), with at least one avs gene detected in 27 of 29 and 3 of 10 well represented bacterial and archaeal phyla, respectively (**Fig. 20B** and **Fig. 21A**). Each Avs family has a characteristic protein size (**Fig. 20C**), consistent with their distinct mechanisms of target recognition. Applicant next constructed phylogenetic trees of each of the four families (**Fig. 20D-20E,** and **Fig. 21B-21C**), and found that these trees did not follow bacterial and archaeal phylogenies, suggesting extensive horizontal gene transfer, particularly for avs2 and avs4, in agreement with previous analyses 226 of STAND NTPases (13, 14). Furthermore, Applicant detected at least 18 distinct types of N-terminal effector domains present in Avs proteins, including non-nuclease domains such as proteases, nucleosidases, sirtuins (SIR2), Toll/interleukin-1 receptor homology (TIR) domains, cytidine monophosphate (CMP) hydrolases, transmembrane helices, and domains with unknown functions (**Tables 3-6** herein and **Data S8** of Gao et al., 2022. "Prokaryotic innate immunity via pattern recognition of conserved viral proteins" Science 377, eabm4096 (2022), which is incorporated by reference as if expressed in its entirety herein). Some less common variants are predicted to participate in intracellular signaling networks via effector-associated domains (EADs) that recruit a caspase-like enzyme by protein-protein interaction (29, 30) (**Fig. 22A-22B**), reminiscent of animal NLRs.

[0639] The apparent frequent exchange of N-terminal domains in the evolution of the Avs families emphasizes the modular organization characteristic of STAND NTPases (14) and implies that closely related ATPase and TPR domains can activate a wide range of effector functions beyond DNA cleavage. To test this hypothesis, Applicant chose an Avs4 homolog from Sulfurospirillum sp. that contains ATPase and TPR domains similar to those of EcAvs4, with 44% overall amino acid identity, but encompasses an N-terminal region with predicted transmembrane helices instead of a nuclease (**Fig. 23A**). Applicant generated a chimeric Avs4

protein by transplanting the transmembrane domain to EcAvs4 and found that the chimera conferred protection against T7, PhiV-1, and ZL19 (**Fig. 20F** and **Fig. 23B**) while retaining the ability to recognize the portal proteins from diverse phages (**Fig. 23C**).

Phage inhibit Avs systems through diverse anti-defense proteins.

[0640] Bacterial and archaeal viruses have evolved diverse mechanisms to counteract defense systems (31), including numerous anti-restriction and anti-CRISPR proteins (32, 33). Applicant hypothesized that Avs inhibitors might be found among phage early genes, which are expressed before the portal and terminase genes during the phage life cycle. Focusing on the Autographiviridae family of T7-like coliphages, which have readily identifiable early genes, as well as portals and terminases that strongly activate Avs proteins, Applicant identified a set of 122 representative early genes that typically encode small proteins (median length 77 amino acids), characteristic of anti-defense (**Table 7**). Applicant performed a genetic screen for suppressors of Avs toxicity by co-expressing these early genes with SeAvs3, EcAvs4, or KpAvs4 and their cognate phage trigger (**Fig. 24A**). Applicant identified several early genes which rescued cell growth (**Fig. 24B**), most of which originate from a hypervariable region within a group of closely related phages isolated from wastewater (**Fig. 24C**) (34, 35).

[0641] To validate these observations, Applicant produced three of the Avs inhibitors by cell-free translation and observed inhibition of SeAvs3 nuclease activity in vitro by Lidtsur-17, and to a lesser degree by Forsur-7 (**Fig. 24D**). Lidtsur-17, Forsur-7, and Lidtsur-6 were also active in phage plaque assays and restored phage propagation on Avs-containing E. coli (**Fig. 25A-25B**). Surprisingly, these inhibitors were active against different Avs families, including the chimeric EcAvs4, where the effector nuclease was replaced with a transmembrane domain. Furthermore, the lack of detectable sequence similarity between these inhibitors suggests distinct modes of action, which resembles the case of the highly diverse anti-CRISPRs (32, 33). Further studies will be required to elucidate the mechanism of how these phage proteins block Avs activity.

Table 7. Anti-defense candidates tested.

					Abundance relative to no-Avs control			
#	Accession	Amino Acids	Phage	Group	EcAvs4 + portal	KpAvs4 + portal	SeAvs3 + terminase	Note
1	QLF86333.1	168	LS3	--	<0.001	<0.001	<0.001	Anti-restriction Ocr-like
2	NP_041954.1	117	T7	--	<0.001	<0.001	<0.001	Anti-restriction Ocr
3	YP_007500976.1	95	UAB Phi78	1	<0.001	<0.001	<0.001	
4	YP_424974.1	75	K1E	1	<0.001	<0.001	<0.001	
5	AQN32039.1	75	vB_EcoP_R	1	<0.001	<0.001	<0.001	
6	YP_654100.1	74	K1-5	1	<0.001	<0.001	<0.001	
7	YP_009812367.1	69	LL11	1	<0.001	<0.001	<0.001	
8	YP_007500980.1	69	UAB Phi78	1	<0.001	0.0022	<0.001	
9	YP_009812365.1	64	LL11	1	<0.001	<0.001	<0.001	
10	YP_009795380.1	61	mutPK1A2	1	0.0021	<0.001	<0.001	
11	YP_007500978.1	59	UAB Phi78	1	<0.001	<0.001	0.0029	
12	QBQ76569.1	58	vB_EcoP_R4596	1	<0.001	<0.001	<0.001	
13	YP_006987767.1	55	vB_EcoP_ACG-C91	1	<0.001	<0.001	<0.001	
14	YP_424971.1	54	K1E	1	<0.001	<0.001	<0.001	
15	YP_424968.1	51	K1E	1	0.0063	<0.001	<0.001	
16	YP_009796071.1	41	VEc3	1	<0.001	<0.001	<0.001	
17	YP_009796066.1	32	Vec3	1	<0.001	<0.001	0.0013	
18	QHR73015.1	175	Smaasur	2	<0.001	<0.001	<0.001	
19	YP_009821615.1	171	Lidtsur	2	0.0013	0.37	0.22	Lidtsur-17
20	QHR65291.1	170	Altidsur	2	<0.001	0.0019	0.0015	
21	YP_009821618.1	150	Lidtsur	2	<0.001	<0.001	<0.001	
22	QHR70507.1	146	Megetsur	2	<0.001	<0.001	<0.001	
23	QHR65282.1	113	Altidsur	2	<0.001	<0.001	<0.001	
24	QHR67393.1	112	Glasur	2	<0.001	0.034	0.0016	
25	QHR65405.1	111	Mellemsur	2	<0.001	<0.001	<0.001	
26	QHR73014.1	103	Smaasur	2	0.0059	<0.001	0.0064	
27	QHR71835.1	102	Forsur	2	<0.001	0.031	0.46	Forsur-7

28	YP_009821614.1	99	Lidtsur	2	<0.001	<0.001	0.0016	
29	YP_009821603.1	98	Lidtsur	2	<0.001	<0.001	0.0044	
30	YP_009821604.1	97	Lidtsur	2	<0.001	0.91	0.028	Lidtsur-6
31	QHR73008.1	94	Smaasur	2	<0.001	<0.001	<0.001	
32	QHR72976.1	94	Usur	2	<0.001	0.19	0.016	Usur-3
33	QHR68511.1	94	Aldrigsur	2	<0.001	0.02	0.0015	
34	QHR65280.1	92	Altidsur	2	<0.001	<0.001	<0.001	
35	QHR70495.1	91	Megetsur	2	<0.001	0.06	0.031	
36	QHR73012.1	91	Smaasur	2	0.0078	0.011	0.0065	
37	QHR73013.1	89	Smaasur	2	<0.001	<0.001	<0.001	
38	QHR73011.1	87	Smaasur	2	<0.001	0.56	<0.001	Smaasur-6
39	QHR72971.1	83	Usur	2	<0.001	0.0023	<0.001	
40	QHR65404.1	81	Mellemsur	2	<0.001	<0.001	<0.001	
41	QHR67402.1	81	Glasur	2	<0.001	<0.001	<0.001	
42	QHR65287.1	81	Altidsur	2	0.0011	0.0011	0.0039	
43	QHR73005.1	80	Smaasur	2	<0.001	<0.001	<0.001	
44	YP_009821616.1	79	Lidtsur	2	<0.001	<0.001	<0.001	
45	QHR67395.1	76	Glasur	2	<0.001	<0.001	<0.001	
46	QHR65403.1	76	Mellemsur	2	<0.001	0.17	<0.001	Mellemsur-6
47	QHR68502.1	67	Aldrigsur	2	<0.001	<0.001	<0.001	
48	QHR70492.1	64	Megetsur	2	<0.001	<0.001	<0.001	
49	YP_009821607.1	58	Lidtsur	2	<0.001	<0.001	<0.001	
50	YP_009821608.1	57	Lidtsur	2	<0.001	<0.001	<0.001	
51	QHR73009.1	50	Smaasur	2	<0.001	0.0014	<0.001	
52	QHR71836.1	37	Forsur	2	<0.001	<0.001	<0.001	
53	QEG09774.1	174	Penshu1	3	<0.001	<0.001	<0.001	
54	YP_009812836.1	140	IMM-002	3	0.0012	<0.001	0.0031	
55		77	CLB_P1	3	<0.001	<0.001	<0.001	
56	QEG09666.1	92	Peacock	3	<0.001	<0.001	<0.001	
57	QEG09778.1	86	Penshu1	3	0.15	0.027	0.071	Penshu1-7
58	QOI58465.1	81	vB_EcoP-U8	3	0.0011	<0.001	0.0029	
59	QLF86388.1	75	LS2	3	<0.001	<0.001	<0.001	

60	QNH91713.1	63	DY1	3	<0.001	<0.001	<0.001	
61	QEG09773.1	62	Penshu1	3	<0.001	<0.001	<0.001	
62	YP_009324935.1	58	GA2A	3	<0.001	<0.001	0.0016	
63	YP_009787259.1	55	ZG49	3	<0.001	<0.001	<0.001	
64	YP_009789078.1	52	vB_EcoP_F	3	<0.001	<0.001	<0.001	
65	YP_002003738.1	47	EcoDS1	3	<0.001	<0.001	<0.001	
66	YP_009818255.1	41	Ro45lw	3	<0.001	<0.001	<0.001	
67	QEG09670.1	29	Peacock	3	<0.001	<0.001	<0.001	
68	YP_009786124.1	94	ECA2	4	<0.001	<0.001	<0.001	
69	YP_004678726.1	77	K30	4	<0.001	<0.001	<0.001	
70	YP_009200622.1	66	P483	4	<0.001	<0.001	<0.001	
71	NP_523299.1	55	T3	4	<0.001	<0.001	<0.001	
72	YP_009204329.1	54	P694	4	<0.001	<0.001	<0.001	
73	YP_009804640.1	53	SRT7	4	<0.001	<0.001	<0.001	
74	YP_009795801.1	42	EG1	4	<0.001	<0.001	<0.001	
75	YP_009804639.1	40	SRT7	4	<0.001	0.0011	<0.001	
76	YP_009146374.1	86	ECBP5	5	<0.001	<0.001	<0.001	
77		78	ECBP5	5	0.0025	0.0014	0.0016	
78	YP_009146376.1	61	ECBP5	5	<0.001	<0.001	<0.001	
79	YP_009821042.1	191	Minorna	6	<0.001	<0.001	<0.001	
80	YP_009821041.1	73	Minorna	6	<0.001	<0.001	0.0074	
81	YP_009795443.1	169	PGT2	7	<0.001	<0.001	<0.001	
82	YP_009795442.1	111	PGT2	7	0.0019	<0.001	0.0031	
83	YP_009795446.1	101	PGT2	7	<0.001	<0.001	<0.001	
84	YP_009795445.1	96	PGT2	7	<0.001	<0.001	<0.001	

85	YP_009795447.1	91	PGT2	7	<0.001	<0.001	<0.001	
86	YP_009795436.1	88	PGT2	7	<0.001	<0.001	<0.001	
87	YP_009795441.1	82	PGT2	7	<0.001	<0.001	0.0014	
88		80	phiKT	7	<0.001	<0.001	<0.001	
89	YP_009795435.1	79	PGT2	7	<0.001	<0.001	<0.001	
90		77	phiKT	7	<0.001	0.0012	<0.001	
91	YP_009795440.1	59	PGT2	7	<0.001	<0.001	<0.001	
92		56	phiKT	7	<0.001	<0.001	<0.001	
93	YP_009795439.1	37	PGT2	7	<0.001	<0.001	<0.001	
94	YP_009821043.1	71	Minorna	6	--	--	<0.001	
95	YP_424970.1	70	K1E	1	Excluded due to toxicity			
96	YP_654099.1	59	K1-5	1	Excluded due to toxicity			
97	YP_007500983.1	51	UAB Phi78	1	Excluded due to toxicity			
98	YP_009812370.1	37	LL11	1	Excluded due to toxicity			
99	QHR72969.1	154	Usur	2	Excluded due to toxicity			
100	YP_009821599.1	130	Lidtsur	2	Excluded due to toxicity			
101	QHR72972.1	104	Usur	2	Excluded due to toxicity			
102	QHR71845.1	56	Forsur	2	Excluded due to toxicity			
103	QEG09776.1	115	Penshu1	3	Excluded due to toxicity			
104	QEG09672.1	108	Peacock	3	Excluded due to toxicity			
105	QEG09556.1	106	Pisces	3	Excluded due to toxicity			
106	QEG09664.1	71	Peacock	3	Excluded due to toxicity			
107	QLF86337.1	65	LS3	3	Excluded due to toxicity			
108	QEG09668.1	51	Peacock	3	Excluded due to toxicity			
109	NP_523296.1	152	T3	4	Excluded due to toxicity			
110	YP_009820202.1	147	HZP2	4	Excluded due to toxicity			
111	YP_009804637.1	118	SRT7	4	Excluded due to toxicity			
112	YP_009813769.1	72	N30	4	Excluded due to toxicity			
113		60	K30	4	Excluded due to toxicity			
114	NP_523298.1	65	T3	4	Excluded due to toxicity			

115	QOV06375.1	46	JB01	4	Excluded due to toxicity	
116	YP_009152463.1	45	CICC80001	4	Excluded due to toxicity	
117	YP_009146373.1	46	ECBP5	5	Excluded due to toxicity	
118	YP_009821040.1	87	Minorna	6	Excluded due to toxicity	
119	YP_009821038.1	67	Minorna	6	Excluded due to toxicity	
120	YP_009821039.1	59	Minorna	6	Excluded due to toxicity	
121	YP_009795437.1	164	PGT2	7	Excluded due to toxicity	
122	YP_009795444.1	103	PGT2	7	Excluded due to toxicity	
See also Data S9 of Gao et al., Science 377, eabm4096 (2022), which is incorporated herein by reference as if expressed in its entirety herein.						

Multiple Sequence Alignments

[0642] Multiple sequence alignments of Avs1, Avs2, Avs3, Avs4, 24 tested terminases, and 24 tested portals were generated and are shown in Data S10, Data S12, Data S14, Data S16, Data S18, and Data S19, respectively, of Gao et al. Science 377, eabm4096 (2022), which are incorporated by reference as if expressed in their entireties herein.

Maximum likelihood phylogenetic trees

[0643] Avs1, Avs2, Avs3, and Avs4 maximum likelihood phylogenetic trees were generated and are shown in Data S11, Data S13, Data S15, and Data S17, respectively, of Gao et al. Science 377, eabm4096 (2022), which are incorporated by reference as if expressed in their entireties herein.

Discussion

[0644] Here Applicant characterize four families of prokaryotic STAND NTPases and demonstrate that they act as pattern recognition receptors against two hallmark phage proteins, the large terminase subunit and the portal. These proteins, along with the major capsid protein, are the signature proteins of the virus realm Duplodnaviria, which uintes taild phages and tailed archaeal virus (24). Members of this realm, particularly tailed phages, are the most abundant among known viruses (63, 64). The portal protein nucleates virion assembly, occupying the unique pentameric vertex of capsids and providing the attachment site for the phage tail, and serves as the channel for genome entry into and exit from the capsid (36). The terminase is the motor that packages the phage genome into the capsid at high density and pressure, using the

energy of ATP hydrolysis, and cleaves DNA concatemers into genome-size units (37). The universal, complex molecular functions of these proteins engender strong selective constraints and hence evolutionary conservation. It is therefore no surprise that these particular phage proteins were selected as the targets for pattern recognition during the coevolution of prokaryotes with viruses. Furthermore, the three groups (Avs1-3) that recognize terminases do not form a clade in the phylogeny of the STAND domain (**Fig. 4A-4B**), suggesting that defense based on the recognition of this phage protein evolved independently on multiple occasions.

[0103] The reconstitution of two Avs systems in vitro described here provides insight into their mechanism of defense, including promiscuous DNA endonuclease activity (**Fig. 24E**). Although many Avs proteins contain predicted nucleases, Applicant identified diverse N-terminal effector domains throughout Avs families, indicating unique mechanisms of defense that remain to be characterized. The demonstration that at least some of these effectors can be swapped without compromising the defense function of the Avs highlights the modular functionality of these proteins, which appears important for the diversification of defense mechanisms. The effectors of SeAvs3 and EcAvs4 are both activated by tetramerization suggesting that diverse Avs effectors are unified by oligomerization for activity, a common mechanism for signal transduction by STAND NTPases (38). Indeed, oligomerization is also involved in the activation of Cap4 nucleases in cyclic oligonucleotide-based antiphage signaling systems (25).

[0104] Notably, Applicant demonstrated that Avs proteinsrecognize conserved structural features of their cognate targets across an extreme variety of amino acid sequences, including those originating from both tailed phages and archael viruses, as well as eukaryotic herpesviruses, which are only distantly related and do not infect prokaryotes. Structural analysis of Avs3 revealed that it di-rectly detects the active-site residues and ATP ligand of the terminase, thereby targeting the moieties that are the most difficult for phages to mutate without abrogating function.

[0105] Notable similarities, but also several differences, exist between eukaryotic NLRs and prokaryotic Avs proteins. Both are intracellular receptors of the STAND superfamily that detect PAMPs through C-terminal repetitive structures. Both exhibit triggered oligomer-ization, but with distinct interfaces between the central ATPase domains (**Fig. 33A-33B**). Similar to plant NLRs like RPP1 and the Roq1 resistosome, both Avs3 and Avs4 form tetramers, with the effector domains activated by forming a twofold symmetric dimer of dimers

(59, 65). In the absence of their ligands, animal and plant NLRs have autoinhibited states that prevent oligomerization and effector activation (18). In these complexes, the effector (e.g., caspase-1) is a separate protein rather than a domain of the NLR. By contrast, Avs effectors are usually the N-terminal domain of the STAND NTPase. This simpler organization might be advantageous because counteracting phage replication requires a rapid, direct cellular response. This contrast parallels the distinction between the mechanisms of prokaryotic and eukaryotic STING proteins, whereby bacterial STING homologs directly activate TIR domain effectors rather than regulate transcription, as mamma-lian STINGs do (18, 39).

[0106] Bacteria and archaea encode numerous diverse STAND NTPases beyond the four families characterized in this study (14). Although some are not involved in defense, such as the well-characterized transcriptional regulators MalT (21), AfsR (22), and GutR (23), several are confirmed defense genes or are predicted to have a defense function based on their enrichment in genome regions adjacent to known defense systems (4, 40). Applicant investigated several of these other defense-associated systems (**Table 2**) but observed no detectable toxicity when co-expressed with any of the 48 tested terminases or portals. In light of the high success rate of this assay for Avs homologs, these results suggest that they are triggered by other pathogen-related patterns that remain to be identified. Further investigation will shed light on whether these triggers are proteins, and whether they are phage-specific or endogenous to the host. For instance, most characterized plant and fungal NLRs sense the state of host pathways rather than pathogen-specific proteins, and it remains a possibility that other groups of prokaryotic STAND NTPases function similarly.

[0645] Given the extensive sequence divergence among the STAND NTPases, it is unclear whether Avs proteins are direct evolutionary ancestors of their eukaryotic NLR homologs, although this remains a possibility. Alternatively, or additionally, the characteristic tripartite domain architectures of diverse STAND NTPases could have evolved convergently, suggesting that this modular organization is a facile way to create allosterically activated enzymes that can respond to PAMPs and could inspire the design of engineered molecular sensors. Overall, the results of this work advanceunderstanding of host-virus interactions in diverse microbes and extend the paradigm of pattern recognition of pathogen-specific proteins to all domains of life.

[0646]

Materials and Methods

Phylogenetic analysis of STAND NTPases

[0647] For STAND phylogenetic analysis (**Fig. 4A-4B**), PSI-BLAST searches (41) against the database of complete bacterial and archaeal genomes (extracted from Genbank, March 2019) were performed for three iterations using ATPase domains of seven Avs1-3 homologs (WP_126523998.1, WP_115407481.1, WP_084007836.1, WP_060615938.1, WP_139964370.1, WP_063118745.1, WP_001017806.1) investigated experimentally. The 2000 best hits from each run were taken and combined with 949 Avs4 homologs found in the NCBI nonredundant(nr) database in 2021. A non-redundant set of sequences (4843) was used for phylogenetic reconstruction using a hybrid UPGMA/FastTree approach as follows. At the first step, sequence clusters were obtained using MMseqs2 (42) with a sequence similarity threshold of 0.5, and the sequences within each cluster were aligned using MUSCLE (43). At the second step, cluster-to-cluster similarity scores were obtained using HHSEARCH (44) (including trivial clusters consisting of a single sequence each) and normalized by the minimum of the self-scores. Relative similarity scores (s) were converted to distances (d) defined as $d = -\ln s$, and a UPGMA (unweighted pair group method with arithmetic mean) dendrogram was constructed from the distance matrix. At the third step, sequence-based trees were constructed from the cluster alignments using FastTree (45) (WAG evolutionary model, gamma-distributed site rates) and rooted by midpoint; these trees were grafted onto the tips of the profile similarity-based UPGMA dendrograms. FastTree was also used to calculate support values. Only the second step of the above procedure was applied to reconstruct the UPGMA dendrogram using multiple alignments of selected well-supported branches identified by the first procedure.

Construction of avs phylogenetic trees

[0648] Homologs of each of the four clades of avs genes were identified using PSI-BLAST searches against the non-redundant protein sequence database (nr) in June 2021 using position-specific scoring matrices for each clade derived from manually curated multiple sequence alignments (MSAs) of conserved regions. After a round of curation to remove false positives hits and partial proteins, referencing the corresponding genome assemblies to correct misannotated start codons, a list of 1584, 2342, 1018, and 1813 non-redundant full-length proteins were obtained for Avs1, Avs2, Avs3, and Avs4, respectively. To reduce sampling bias, sequences were then clustered at 95% sequence identity (minimum 80% coverage) using

MMseqs2 with parameters --min-seq-id 0.95 -c 0.8 --cov-mode 1. One representative from each cluster was selected for subsequent analyses, resulting in 843, 1255, 630, and 1089 sequences for Avs1, Avs2, Avs3, and Avs4, respectively.

[0649] MSAs of each Avs clade, excluding the variable N-terminal domains, were generated using MAFFT v7.450 (46) with global pairwise alignment (parameters --maxiterate 1000 --globalpair). Alignments were trimmed using trimAl 1.2 with a gap threshold of 0.25 (-gt 0.25). Phylogenetic trees were built from the trimmed MSAs using IQ-TREE 1.6.12 (47) with the LG+G4 model and 2000 ultrafast bootstrap replicates (parameters -nstop 500 -bb 2000 -m LG+G4). To categorize the N-terminal domains, the N-terminal sequences were clustered using MMseqs2 with parameters --min-seq-id 0.4 -c 0.8, and a representative sequence from each cluster was analyzed using HHpred (48). Phyla classification was determined from the NCBI taxonomy database, and trees were rooted by midpoint and visualized using iTOL (49).

[0650] The phylogenetic tree comparing the ATPase domains of NLR-like genes across model organisms (**Fig. 3A**) was constructed in a similar manner, incorporating the set of 23 human NLRs and the best characterized NLRs from Arabidopsis thaliana (50) and Neurospora crassa (19, 51).

Taxonomic distribution of avs genes

[0651] To determine the taxonomic distribution of avs genes, genome assemblies containing one or more full-length Avs homologs were identified via the NCBI Identical Protein Groups (IPG) database (**Data S8** of Gao et al., 2022. "Prokaryotic innate immunity via pattern recognition of conserved viral proteins" *Science 377 eabm4096*, which is incorporated by reference as if expressed in its entirety herein). Redundant assemblies were removed on the basis of their nine-digit accession numbers. To determine the percentage of genomes containing avs genes, the list of all available prokaryotic assemblies was downloaded from ftp.ncbi.nlm.nih.gov/genomes/GENOME_REPORTS/prokaryotes.txt.

Construction of terminase and portal alignments

[0652] Structures of all tested terminase and portal proteins were predicted using AlphaFold2 (52), and structures were aligned and visualized using PyMOL 2.3.4. Representatives of the predicted structures were used as input for MSA construction using PROMALS3D (53). Prior to computing pairwise sequence identity, MSAs were trimmed to retain only the regions corresponding to the core terminase or portal fold.

Cloning

[0653] Genes were chemically synthesized or amplified with Q5 (New England Biolabs) or Phusion Flash (Thermo Scientific) polymerase. Plasmids were assembled using the Gibson Assembly or NEBuilder HiFi DNA Assembly mix (New England Biolabs). Plasmid sequences were verified by Tn5 tagmentation and high-throughput sequencing, as previously described (4, 54).

Bacterial Strains

[0654] E. coli NovaBlue and NovaBlue(DE3) were obtained from Millipore Sigma. E. coli K-12 (ATCC 25404) and strain C (ATCC 13706) were obtained from the American Type Culture Collection. All genetic assays were performed with E. coli NovaBlue(DE3) unless indicated otherwise.

Competent cell production

[0655] E. coli strains were cultured in ZymoBroth with 25 μg/mL chloramphenicol and made competent using Mix & Go buffers (Zymo) according to the manufacturer's recommended protocol.

PhiV-1 fragment screen

[0656] DNA fragments consisting of intact open reading frames (ORFs) were amplified from phage PhiV-1 and cloned into expression plasmids after a LacI-repressed T7 promoter. Plasmids were pooled with an mNeonGreen-expressing control plasmid and transformed into E. coli NovaBlue(DE3) containing either SeAvs3, EcAvs4, or an Avs-free pACYC184 empty vector. An additional sample consisting of the plasmidpool transformed into empty vector-containing E. coli NovaBlue, which lacks the ability to express from T7 promoters, was also included to assess the basal toxicity of the phage genes. After 1 hr outgrowth in S.O.C. (super optimal broth with catabolite repression) media (at 37 degrees C, cells were plated on LB agar plates containing 25 μg/mL chloramphenicol and 100 μg/mL ampicillin in the absence of IPTG. Plates were incubated for an additional 12h at 37 degrees C, after which surviving plasmids were isolated by miniprep (Qiagen). A total of 200 ng of plasmid for each condition was tagmented with Tn5 to yield an average fragment size of about 500 bp. Following addition of 0.5 volumes of 0.1% SDS and column purification, tagmented fragments were amplified over 8 cycles by Q5 DNA polymerase (NEB) with unique i5 and i7 index primers. Amplicons were gel extracted and sequenced on a NextSeq (Illumina) using 150 cycles for the forward read. Reads were mapped to reference sequences using Geneious Prime. The read coverage of

each sample was then normalized to the read coverage of the mNeonGreen control within the same sample. Finally, for each sample, the read coverage per base was divided by the corresponding read coverage per base for the empty vector NovaBlue(DE3) control (**Fig. 3D**), or by that of the empty vector NovaBlue control (**Fig. 5A-5B**).

Terminase and portal depletion screens

[0657] Terminase and portal genes were amplified directly from phage samples or chemically synthesized (Twist Bioscience) with codon optimization for E. coli. Genes were expressed under the control of a pBAD promoter. Plasmids were pooled with an mCherry-expressing control plasmid and transformed into E. coli NovaBlue (DE3) containing an Avs homolog or a pACYC184 empty vector. After 1 hr outgrowth in S.O.C. at 37 degrees C, cells were plated on LB agar plates containing 25 μg/mL chloramphenicol and 100 μg/mL ampicillin with 0.002% arabinose, or in some cases with 0.2% arabinose as detailed in the figures. After an additional 12h at 37 degrees C, plasmids were isolated and sequenced, and depletion values were computed as described for the PhiV-1 depletion screen (**Fig. 7A-7D, Fig. 9A-9B,** and **Fig. 23C**).

Portal and terminase mutant screens

[0658] Two synonymous versions of the T7 portal gene were randomly mutagenized by PCR using KAPA HiFi HotStart ReadyMix DNA polymerase (Roche) and cloned via Gibson assembly into a plasmid backbone containing a LacI-repressed T7 promoter. Plasmids were column purified and transformed into E. coli NovaBlue(DE3) containing EcAvs4, KpAvs4, or CcAvs4. Cells were plated on LB agar plates containing 25 μg/mL chloramphenicol and 100 μg/mL ampicillin in the absence of IPTG. After overnight growth, surviving colonies were sampled at random, cultured, and sequenced, and those containing single amino acid substitutions in the portal were retained for subsequent analysis. To reduce the number of stop codons and frameshift mutants sampled, a fluorescent protein (mNeonGreen) was included in the plasmid backbone immediately after the portal ORF with a single nucleotide overlap, such that mNeonGreen was translated only if the portal ORF remained intact (55) (**Fig. 18A-18B**). Both portal and mNeonGreen were translated as separate polypeptides. To quantitatively assess the effect of each mutant on Avs4-mediated toxicity (**Fig. 18C**), mutant plasmids were pooled and re-transformed into Avs4-containing E. coli as described for the PhiV-1 depletion screen. Fold depletion was also quantified as described, with the exception that only reads containing 20-mer sequences unique to one mutant (i.e., mapping to the mutation site) were counted in

the analysis. A similar procedure was followed to quantify the effect of truncation of the terminase or portal (**Fig. 17G** and **17L**) , as well as terminase-domain mutations (**Fig. 28F**).

Antidefense screen

[0659] Putative early genes from Autographiviridae coliphages were tabulated clustered at 50% sequence identity and 50% coverage using MMseqs2 (--min-seq-id 0.5 -c 0.5), resulting in 120 clusters. One representative was selected from each cluster, along with two additional sequences, for a total of 122 initial candidates (**Table 7**). Genes were synthesized by Twist Bioscience and cloned via Gibson assembly into expression vectors containing either the portal or terminase from phage PhiV-1 driven by a pBAD promoter. Antidefense candidates were expressed under the control of a lac promoter. Plasmids were pooled and transformed into E. coli containing SeAvs3, EcAvs4, KpAvs4, or an empty vector. Cells were grown at 37 degrees C for 16h on LB agar plates containing 25 μg/mL chloramphenicol and 100 μg/mL ampicillin with no added arabinose. Following plasmid isolation, anti-defense candidates were amplified over two rounds of PCR to attach 8-nucleotide i7 and i5 index barcodes and sequenced with a 600 cycle MiSeq kit to ensure maximal coverage of each ORF. Reads containing mutations were discarded in the subsequent analysis.

Phage plaque assays

[0660] E. coli host strains were grown to saturation at 37 degrees C in Luria Broth (LB) or terrific broth (TB). To 10 mL top agar (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 7 g/L agar) was added chloramphenicol to a final concentration 25 μg/mL and, if needed, ampicillin to a final concentration of 100 μg/mL. Ten-fold dilutions of phage in phosphate-buffered saline were spotted on the plates. After overnight incubation at 37 degrees C, plates were photographed in a dark room with a white backlight.

Construction of mutant phages

[0661] PhiV-1 gp8 and gp19 knockout phages were constructed as previously described (56, 57) using plasmid donors with homology arms to gp8 or gp19 in an E. coli trxA– strain (JW5856 from the Keio collection (58)). The sequence encoding TrxA was inserted into the PhiV-1 genome as a selection marker.

Protein purification

[0662] Avs NTPases were cloned into pCDF-Duet expression plasmids containing a C-terminal 6xHis tag. PhiV-1 gp8 and gp19 genes were cloned into TwinStrep-SUMO expression plasmids. Proteins were expressed in BL21(DE3) cells (NEB #C2527H). Cells were grown in

Terrific Broth to mid-log phase and the temperature lowered to 18 degrees C. Expression was induced at OD_{600} 0.6 with 0.25 mM IPTG for 16-20 h before harvesting and freezing cells at -80 degrees C. Cell paste was resuspended in lysis buffer (50 mM Tris pH 7.5, 500 mM NaCl, 5% glycerol) supplemented with EDTA-free cOmplete protease inhibitor (Roche). Cells were lysed using a LM20 microfluidizer device (Microfluidics) and cleared lysate was bound to either Strep-Tactin Superflow Plus (Qiagen) or Ni-NTA Superflow resin (Qiagen). For TwinStrep-SUMO phage proteins, the resin was washed with lysis buffer and proteins eluted with lysis buffer supplemented with 5 mM desthiobiotin. The TwinStrep-SUMO tag was removed by overnight digest at 4 degrees C with homemade SUMO protease Ulp1 at a 1:100 weight ratio of protease to target. Cleaved proteins were run on a Superose 6 Increase column (GE Healthcare Life Sciences) with a final storage buffer of 25 mM Tris pH 7.5, 500 mM NaCl, 10% glycerol, 1 mM DTT. Avs proteins containing 6xHis tags were bound to Ni-NTA resin in the presence of 25 mM imidazole, washed with lysis buffer containing 50 mM imidazole, and eluted with lysis buffer containing 300 mM imidazole. SeAvs3 was diluted to a final concentration of 100 mM NaCl and purified using a Resource Q column on an AKTA Pure 25 L (GE Healthcare Life Sciences) with a 100 mM-1M NaCl gradient. EcAvs4 was further purified by diluting to a final concentration 100 mM NaCl and absorbing contaminants by flowing the protein over a Resource Q and Heparin HP column. SeAvs3 and EcAvs4 were concentrated and loaded onto a Superose 6 Increase column with a final storage buffer of 25 mM Tris pH 7.5, 500 mM NaCl, 10% glycerol, 1 mM dithiothreitol (DTT).

[0663] Avs proteins containing 6xHis tags were bound to Ni-NTA resin in the presence of 25 mM imidazole, washed with lysis buffer containing 50 mM imidazole, and eluted with lysis buffer containing 300 mM imidazole. SeAvs3 was diluted to a final concentration of 100 mM NaCl and purified using a Resource Q column on an AKTA pure 25 L (GE Health-care Life Sciences) with a 100 mM-1M NaCl gradient. EcAvs4 was further purified by di-luting to a final concentration of 100 mM NaCl and absorbing contaminants by flowing the protein over a Resource Q and Heparin HP column. SeAvs3 and EcAvs4 were concentrated and loaded onto a Superose 6 Increase column with a final storage buffer of 25 mM Tris-HCl pH 7.5, 500 mM NaCl, 10% glycerol, and 1 mM DTT. SeAvs3 for cryo-EM analysis was purified in the same buffer without glycerol and only 300 mM NaCl, then concentrated to 1.4 mg/ml in a 500-µl 100,000 molecular weight cutoff (MWCO) Amicon spin concentrator.

Avs complex purification

[0664] Avs-TwinStrep constructs were co-transformed with plasmids expressing either gp8 or gp19 into electrocompetent BL21 cells (Sigma Aldrich CMC0016) and grown and induced as before. Avs pulldowns using Strep-Tactin resin were run on SDS-PAGE gels and bound gp8 and gp19 bands excised and sent for confirmation by mass spectrometry (Taplin Biological Mass Spectrometry Facility, Harvard Medical School). For tandem affinity purification, plasmids containing SeAvs3-6xHis and gp19-StrepTag were cotransformed into electrocompetent E. coli BL21(DE3) and grown and induced as before. An SeAvs3-gp19 complex was purified using Ni-NTA followed by Strep-Tactin SuperFlow Plus resin. The final elution was run on a Superose 6 Increase column and yielded a peak elution at 13 ml containing a 1:1 ratio of SeAvs3 and gp19, as determined by SDS-PAGE band intensity analysis. A standard curve was generated using the Bio-Rad Gel Filtration Standard (1511901), and the gel-phase distribution coefficient (K_{ave}) was cal-culated as (elution volume – void column)/(column volume – void volume).

SeAvs3-terminase complex formation for cryoEM

[0665] A total of 20 mg of SeAvs3 was mixed with 8.3 mg of PhiV-1 gp19 terminase in a total volume of 24 ml in the presence of 17 mM Tris-HCl pH 7.5, 280 mM NaCl, 0.8 mM DTT, 2% glycerol, 5mM $MgCl_2$,and1 mMATP. The reaction was incubated at 37°C for 30 min and placed on ice for about 1 hour before cryo-EM grid prepa-ration. Cryo-EM grids were prepared on a Thermo Scientific Vitrobot Mark IV at 4°C and 100% humidity. A total of 3 mlofreaction was applied to a freshly glow-discharged (12 s at 15 mA) Cu 300 R2/2 holey carbon grid with a 2-nm layer of amorphous carbon (Quantifoil). After 30 s, the grid was manually blotted with Whatman Grade 1 filter paper and plunged into liquid ethane.

EcAvs4-portal complex formation for cryoEM

[0666] PhiV-1 gp8 was cloned into an MBP-bdSUMO expression plasmid, and EcAvs4 was cloned into a pCDF-Duet plasmid with an internal TwinStrep tag added between residues 114 and 115. The EcAvs4 Mrr-like nuclease active site was mutated (Q63A/K65A) to allow co-expression with the portal. These two plasmids were cotransformed into E. coli BL21(DE3). A total of 6 liters of culture was grown in Terrific Broth to mid–log phase, and the tem-perature was lowered to 18°C. IPTG (0.25 mM) was added to induce expression, and growth was continued overnight. Cell paste was re-suspended in lysis buffer (50 mM Tris-HCl pH 7.4, 250 mM NaCl, 5% glycerol, 5 mM β-mercaptoethanol, 2 mM $MgCl_2$,0.1 mM ATP) supplemented

with EDTA-free cOmplete pro-tease inhibitor (Roche). Cells were lysed using a LM20 microfluidizer (Microfluidics) and cleared lysate was bound to Amylose Resin High Flow (NEB). After extensive washing with lysis buffer, the resin was eluted overnight at 4°C by addition of 10 mgofhomemade bdSENP1 protease. Eluted protein was in-cubated with Strep-Tactin Superflow Plus resin, washed with lysis buffer, then eluted with lysis buffer supplemented with 5 mM desthiobiotin. The eluate was concentrated in a 6-ml Vivaspin spin concentrator (30,000 MWCO) and run on a Superose 6 Increase column using 20 mM Tris-HCl pH 7.4, 200 mM NaCl, 2 mM $MgCl_2$,and 0.1 mMATP.Peak fractions containing EcAvs4 and gp8 were concentrated to 1.7 mg/ml using a 0.5 ml Amicon spin concentrator (100,000 MWCO) then immediately used for cryo-EM grid prep-aration. Cryo-EM grids were prepared on an Thermo Scientific Vitrobot Mark IV at 4°C and 100% humidity. A total of 3 mlofsamplewas applied to a freshly glow-discharged (60 s at 15 mA) Cu 300 R1.2/1.3 holey carbon grid (Quantifoil). The grid was blotted for 4 s with blot force +5 and drain time 1 s, then plunged into liquid ethane.

Cryo-EM data collection

[0667] All data were collected using the Thermo Sci-entific Titan Krios G3i cryo TEM at MIT.nano using a K3 direct detector (Gatan) operated in super-resolution mode with twofold binning and an energy filter with slit width of 20 eV.

[0668] For SeAvs3-gp19, 15,422 movies were collected at 105,000× magnification giving a real pixel size of 0.8697 Å, with defocus ranging from 1 to 3.5 mm with an exposure time of 1.15 s, fractionated into 30 frames and a flux of 19.7 e–/pix/s giving a total fluence per micro-graph of 30 e–/Å2. For EcAvs4-gp8, 22,902 movies were collected at 130,000× magnification giving a real pixel size of 0.6788 Å, with defocus ranging from 1 mmto 2.5 mmwithanexposure timeof 0.6 s, fractionated into24 framesand afluxof 23.6 e–/pix/s giving a total fluence per micro-graph of 30.8 e–/Å2.

Cryo-EM data processing

[0669] All cryo-EM data were processed using RELION-4.0 (66). Movies were corrected for motion using the RELION implementation of MotionCor2, with 5-by-5 patches and dose-weighting. Contrast transfer function (CTF) parameters were estimated using CTFFIND-4.1. All reported resolutions use the gold-standard Fourier shell correlation with a cutoff of 0.143. For the SeAvs3-gp19 dataset, particle picking was first carried out on 800 micrographs using the Topaz general model (67). A good subset of these particles, as determined by three-

dimensional (3D) classification, was used to train Topaz, and this trained model was used to pick 128,500 particles from the entire dataset. Extracted particles, downscaled fourfold, were subjected to 3D classification without imposing symmetry using a reference derived from a preliminary dataset. A total of 44,489 particles, corresponding to 34.5% of picked particles, showed sharp features and apparent C4 symmetry and were reextracted without binning and refined with C4 symmetry imposed. After refining per-particle defocus and global magnification, beamtilt, trefoil, and performing Bayesian polishing, a reconstruction was yielded at 3.8-Å resolution with clear density for theSeAvs3ATPase domain but blurred density for both the N-terminal nuclease and C-terminal TPR+terminase domains.

[0670] To improve density for the N-terminal nuclease domains, 3D classification without alignment was performed while imposing C2 symmetry. This revealed two equal populations of particles each with clear density for the nuclease domains, related by a 90° rotation in the z axis. In the refinement STAR file, the parameter *rlnAngleRot* was therefore incremented by 90° for one of these populations, before focused refinement starting at 1.8° local angular searches with a soft mask around the nuclease domains. This produced a reconstruction at 3.4 Å, measured using the same soft mask.

[0671] To improve density for the C-terminal TPR+terminase domains, C4 symmetry expansion was performed on the C4 refinement data.star file, followed by particle subtraction with recentering using a mask around one of the four asymmetric units. This generated four subparticles for each original particle. Refinement starting at 1.8° local angular searches with a soft mask, followed by CTF refinement and another round of refinement, produced a reconstruction at 3.4-Å resolution.

[0672] For the EcAvs4-gp8 dataset, 1825 particles from 80 micrographs were manually picked and used to train Topaz. The trained Topaz model then picked 444,626 particles from the entire dataset, which were extracted with fourfold binning and subjected with 3D classification without imposing symmetry using the octameric (pseudo-D4) EcAvs4-gp8 reference derived from a preliminary dataset. A total of 133,133 particles (29.9%) showing the same pseudo-D4 symmetry were reextracted at 1.034 Å/pix and refined with D4 symmetry imposed. After Bayesian polishing, this yielded a 3.7-Å resolution reconstruction. D1 symmetry expansion followed by particle subtraction was then used to convert these particles to 266,266 subparticles that correspond to the tetrameric complex. Like SeAvs3-gp19, these also had blurry density for the N-terminal nuclease and C-terminal TPR+terminase domains

but additionally had poor density for the ATPase domains, suggesting a C2 reconstruction might be suitable for the whole tetramer.

[0673] To improve overall density, a 3D classification without alignment was first performed with C2 symmetry imposed. This produced two equally occupied classes, collectively rep-resenting a 169,977-particle subset (63.8%), that appeared identical but for a 90° rotation, but less clearly distinguished than the same analysis on SeAvs3-gp19. Therefore, they were refined together with local 1.8° angular searches and C2 symmetry but with "Relax symmetry: C4" to account for the pseudo-C4 symmetry. This produced a consensus C2 refinement but still with relatively blurred densities for the nuclease and C-terminal TPR domains. The nuclease domain density was improved by focused refinement with a soft mask, followed by refining anisotropic magnification, per-(sub)particle defocus, and beamtilt, trefoil and fourth-order aberrations, and second refinement, yielding a 2.9-Å resolution re-construction. The C-terminal TPR domains were improved by C4 symmetry expansion, followed by C1 focused refinement with a soft mask and CTF refinement, but still had unclear density at the periphery at the site of an important EcAvs4-portal contact. There-fore, a final 3D classification was performed with a soft mask just around this contact and a regularization parameter (T) of 20. A total of 500,066 selected subparticles (73%) were then focus-refined with the same mask to yield a reconstruction at 3.0-Å resolution with better density for this region.

Model Building

[0674] Initial models for SeAvs3, EcAvs4, PhiV-1 gp8, and PhiV-1 gp19 were generated using AlphaFold and fit into the cryo-EM maps using ISOLDE (68) with adaptive distance restraints, followed by manual rebuilding in Coot (69) and further refinement in ISOLDE. Coordi-nates were refined in real space using PHENIX (70), performing one macrocycle of global minimization and atomic displacement pa-rameter (ADP) refinement and skipping local grid searches.

In vitro cleavage reactions

[0675] Purified Avs proteins were incubated with nucleic acid substrates in reaction buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 1 mM DTT, 5% glycerol final concentration). Typical reactions contained approximately 100 ng of DNA substrate, 100 ng Avs protein, and 100 ng gp8 or gp19 in a 10 μL reaction volume. MgCl2 was added at 5 mM where indicated, ATP and AMP-PNP at 1 mM. Reactions were carried out at 37 C for the indicated time and products

purified using a PCR Purification column (Qiagen) before agarose gel analysis on a 1% E-Gel EX (Thermo Fisher Scientific).

Bacterial two-hybrid assays

[0676] Expression plasmids were cloned by fusing either the T18 or T25 fragments of CyaA from *Bordetella pertussis* to nuclease-deficient Avs proteins as well as PhiV-1 gp8 portal and gp19 terminase. BTH101 cells (F–, cya-99, araD139, galE15, galK16, rpsL1 (Strr), hsdR2, mcrA1, mcrB1) were cotransformed with pairs of T18 and T25 containing plasmids. Overnight cultures were diluted 1:20 and plated on indicator plates containing 50 mg/mL ampicillin, 25 mg/mL kanamycin, 500 μg/mL ammonium iron (III) citrate, 300 μg/mL S-gal, and 0.5 mM IPTG. Cells were grown at 30 C overnight before imaging.

Southern blot analysis

[0677] E. coli K-12 (ATCC 25404) cultures were grown to mid-log phase (OD600 0.5), and for each sample, 6 mL of culture was infected with wild type or mutant PhiV-1 at a multiplicity of infection of 1. After 20 min at 37 degrees C, prior to cell lysis, infected cells were pelleted and resuspended in 200 μL of media. After further incubation at 37 degrees C, for a total of 90 min, samples were frozen in liquid nitrogen. DNA was extracted from 200 μL cultures by adding 200 μL lysis buffer (10 mM Tris-Cl, pH 8.0, 1 mM EDTA, 100 mM NaCl, 1% SDS, 2% Triton X-100), 100 μL of glass beads, and 200 μL phenol-chloroform (1:1) followed by brief vortexing. Samples were centrifuged at 4 degrees C and DNA from the upper layer, extracted with chloroform, and precipitated with the addition of 1 mL ice-cold 100% ethanol and centrifugation at 4 degrees C. DNA pellets were resuspended in 200 μL TE with 300 μg RNAse A (Sigma-Aldrich) and incubated at 37 degrees C for 1 hour. DNA was precipitated with the addition of 1 mL ice-cold 100% ethanol and 20 μL of 4 M ammonium acetate, centrifuged, dried, and resuspended in TE. DNA was digested with Eco47III and run on a 1% agarose gel in 1x TBE at 100 V. The gel was denatured with 0.5 M NaOH and 1.5 M NaCl for 30 min, and neutralized with 1.5 M NaCl and 0.5 M Tris- Cl pH 7.5 for 30 min. DNA was transferred to Hybond N + membrane (GE Healthcare Life Sciences) using overnight capillary flow and 10X SSC buffer (1.5 M NaCl, 150 mM sodium citrate, pH 7). Membranes were UV-crosslinked (Stratalinker 1800, Agilent) and blocked at 61 degrees C with Church hybridization buffer (250 mM NaPO4 pH 7.2, 1 mM EDTA, 7% SDS). Radiolabeled probes complementary to the gp13 gene were generated from purified PCR products using the Prime-It Random labeling kit (Agilent) and α32- dCTP. Membranes were probed overnight, washed

three times with 61 degrees C Church hybridization buffer and exposed overnight with X-ray film (GE Healthcare Life Science) before developing. Quantification of phage DNA bands was performed in Fiji with background signal subtracted.

References for Example 1

[0678] 1. K. S. Makarova, Y. I. Wolf, S. Snir, E. V. Koonin, Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. J. Bacteriol. 193, 6039–6056 (2011).

[0679] 2. E. V. Koonin, K. S. Makarova, Y. I. Wolf, Evolutionary Genomics of Defense Systems in Archaea and Bacteria. Annu. Rev. Microbiol. 71, 233–261 (2017).

[0680] 3. S. Doron, S. Melamed, G. Ofir, A. Leavitt, A. Lopatina, M. Keren, G. Amitai, R. Sorek, Systematic discovery of antiphage defense systems in the microbial pangenome. Science. 359 (2018), doi:10.1126/science.aar4120.

[0681] 4. L. Gao, H. Altae-Tran, F. B hning, K. S. Makarova, M. Segel, J. L. Schmid-Burgk, J. Koob, Y. I. Wolf, E. V. Koonin, F. Zhang, Diverse enzymatic activities mediate antiviral immunity in prokaryotes. Science. 369, 1077–1084 (2020).

[0682] 5. D. Klaiman, E. Steinfels-Kohn, G. Kaufmann, A DNA break inducer activates the anticodon nuclease RloC and the adaptive immunity in Acinetobacter baylyi ADP1. Nucleic Acids Res. 42, 360 328–339 (2014).

[0683] 6. C. K. Guegler, M. T. Laub, Shutoff of host transcription triggers a toxin-antitoxin system to cleave phage RNA and abort infection. Mol. Cell. 81, 2361–2373.e9 (2021).

[0684] 7. R. Cheng, F. Huang, H. Wu, X. Lu, Y. Yan, B. Yu, X. Wang, B. Zhu, A nucleotide-sensing endonuclease from the Gabija bacterial defense system. Nucleic Acids Res. 49, 5216–5229 (2021).

[0685] 8. R. Bingham, S. I. Ekunwe, S. Falk, L. Snyder, C. Kleanthous, The major head protein of bacteriophage T4 binds specifically to elongation factor Tu. J. Biol. Chem. 275, 23219–23226 (2000).

[0686] 9. A. Millman, A. Bernheim, A. Stokar-Avihail, T. Fedorenko, M. Voichek, A. Leavitt, Y. Oppenheimer-Shaanan, R. Sorek, Bacterial Retrons Function In Anti-Phage Defense. Cell. 183, 1551–1561.e12 (2020).

[0687] 10. S. Kronheim, M. Daniel-Ivad, Z. Duan, S. Hwang, A. I. Wong, I. Mantel, J. R. Nodwell, K. L. Maxwell, A chemical defence against phage infection. Nature. 564 (2018), pp. 283–286.

[0688] 11. A. Bernheim, A. Millman, G. Ofir, G. Meitav, C. Avraham, H. Shomar, M. M. Rosenberg, N. Tal, S. Melamed, G. Amitai, R. Sorek, Prokaryotic viperins produce diverse antiviral molecules. Nature. 589, 120–124 (2021).

[0689] 12. K. S. Makarova, Y. I. Wolf, E. V. Koonin, Comparative genomics of defense systems in archaea and bacteria. Nucleic Acids Res. 41, 4360–4377 (2013). 13. E. V. Koonin, L. Aravind, Origin and evolution of eukaryotic apoptosis: the bacterial connection. Cell Death Differ. 9, 394–404 (2002).

[0690] 13. Koonin et al., Origin and evolution of eukaryotic apoptosis: The bacterial connection. Cell Death Differ. 9, 394–404 (2002).

[0691] 14. D. D. Leipe, E. V. Koonin, L. Aravind, STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. J. Mol. Biol. 343, 1–28 (2004).

[0692] 15. Y. Zhao, J. Yang, J. Shi, Y.-N. Gong, Q. Lu, H. Xu, L. Liu, F. Shao, The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. Nature. 477 (2011), pp. 596–600.

[0693] 16. E. M. Kofoed, R. E. Vance, Innate immune recognition of bacterial ligands by NAIPs determines inflammasome specificity. Nature. 477, 592–595 (2011).

[0694] 17. R. Caruso, N. Warner, N. Inohara, G. N ez, NOD1 and NOD2: signaling, host defense, and inflammatory disease. Immunity. 41, 898–908 (2014).

[0695] 18. J. D. G. Jones, R. E. Vance, J. L. Dangl, Intracellular innate immune surveillance devices in plants and animals. Science. 354 (2016), doi:10.1126/science.aaf6395.

[0696] 19. J. Heller, C. Clav , P. Gladieux, S. J. Saupe, N. L. Glass, NLR surveillance of essential SEC-9 SNARE proteins induces programmed cell death upon allorecognition in filamentous fungi. Proc. Natl. Acad. Sci. U. S. A. 115, E2292–E2301 (2018).

[0697] 20. S. Bauernfried, M. J. Scherr, A. Pichlmair, K. E. Duderstadt, 394 V. Hornung, Human NLRP1 is a sensor for double-stranded RNA. Science. 371 (2021), doi:10.1126/science.abd0811.

[0698] 21. O. Danot, A complex signaling module governs the activity of MalT, the prototype of an emerging transactivator family. Proc. Natl. Acad. Sci. U. S. A. 98, 435–440 (2001).

[0699] 22. S. Horinouchi, M. Kito, M. Nishiyama, K. Furuya, S. K. Hong, K. Miyake, T. Beppu, Primary structure of AfsR, a global regulatory protein for secondary metabolite formation in Streptomyces coelicolor A3(2). Gene. 95, 49–56 (1990).

[0700] 23. R. Ye, S. N. Rehemtulla, S. L. Wong, Glucitol induction in Bacillus subtilis is mediated by a regulatory factor, GutR. J. Bacteriol. 176, 3321–3327 (1994).

[0701] 24. E. V. Koonin, V. V. Dolja, M. Krupovic, A. Varsani, Y. I. Wolf, N. Yutin, F. M. Zerbini, J. H. Kuhn, Global Organization and Proposed Megataxonomy of the Virus World. Microbiol. Mol. Biol. Rev. 84 (2020), doi:10.1128/MMBR.00061-19.

[0702] 25. B. Lowey, A. T. Whiteley, A. F. A. Keszei, B. R. Morehouse, I. T. Mathews, S. P. Antine, V. J. Cabrera, D. Kashin, P. Niemann, M. Jain, F. Schwede, J. J. Mekalanos, S. Shao, A. S. Y. Lee, P. J. Kranzusch, CBASS Immunity Uses CARF-Related Effectors to Sense 3'-5'- and 2'-5'-Linked Cyclic Oligonucleotide Signals and Protect Bacteria from Phage Infection. Cell. 182, 38–49.e17 (2020).

[0703] 26. K. S. Makarova, A. Timinskas, Y. I. Wolf, A. B. Gussow, V. Siksnys, Č. Venclovas, E. V. Koonin, Evolutionary and functional classification of the CARF domain superfamily, key sensors in prokaryotic antivirus defense. Nucleic Acids Res. 48, 8828–8847 (2020).

[0704] 27. J. Heitman, P. Model, Site-specific methylases induce the SOS DNA repair response in Escherichia coli. J. Bacteriol. 169, 3243–3250 (1987).

[0705] 28. G. Karimova, J. Pidoux, A. Ullmann, D. Ladant, A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc. Natl. Acad. Sci. U. S. A. 95, 5752–5756 (1998).

[0706] 29. G. Kaur, L. M. Iyer, A. M. Burroughs, L. Aravind, Bacterial death and TRADD-N domains help define novel apoptosis and immunity mechanisms shared by prokaryotes and metazoans. Elife. 10 (2021), doi:10.7554/eLife.70394.

[0707] 30. G. Kaur, A. M. Burroughs, L. M. Iyer, L. Aravind, Highly regulated, diversifying NTP-dependent biological conflict systems with implications for the emergence of multicellularity. Elife. 9 (2020), doi:10.7554/eLife.52696.

[0708] 31. J. E. Samson, A. H. Magad n, M. Sabri, S. Moineau, Revenge of the phages: defeating bacterial defenses. Nat. Rev. Microbiol. 11, 675–687 (2013).

[0709] 32. A. Pawluk, A. R. Davidson, K. L. Maxwell, Anti-CRISPR: discovery, mechanism and function. Nat. Rev. Microbiol. 16, 12–17 (2018).

[0710] 33. J. Bondy-Denomy, A. Pawluk, K. L. Maxwell, A. R. Davidson, Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature. 493, 429–432 (2013).

[0711] 34. N. S. Olsen, L. Forero-Junco, W. Kot, L. H. Hansen, Exploring the Remarkable Diversity of Culturable Phages in the Danish Wastewater Environment. Viruses. 12 (2020), doi:10.3390/v12090986. 35. M. Schmerer, I. J. Molineux, J. J. Bull, Synergy 432 as a rationale for phage therapy using phage cocktails. PeerJ. 2, e590 (2014).

[0712] 35. M. Schmerer, I. J. Molineux, J. J. Bull, Synergy as a rationale for phage therapy using phage cocktails. PeerJ 2, e590 (2014).

[0713] 36. C. L. Dedeo, G. Cingolani, C. M. Teschke, Portal Protein: The Orchestrator of Capsid Assembly for the dsDNA Tailed Bacteriophages and Herpesviruses. Annu Rev Virol. 6, 141–160 (2019).

[0714] 37. S. R. Casjens, The DNA-packaging nanomotor of tailed bacteriophages. Nat. Rev. Microbiol. 9, 647–657 (2011).

[0715] 38. O. Danot, E. Marquenet, D. Vidal-Ingigliardi, E. Richet, Wheel of Life, Wheel of Death: A Mechanistic Insight into Signaling by STAND Proteins. Structure. 17 (2009), pp. 172–182.

[0716] 39. B. R. Morehouse, A. A. Govande, A. Millman, A. F. A. Keszei, B. Lowey, G. Ofir, S. Shao, R. Sorek, P. J. Kranzusch, STING cyclic dinucleotide sensing originated in bacteria. Nature. 586, 429– 433 (2020).

[0717] 40. S. A. Shmakov, K. S. Makarova, Y. I. Wolf, K. V. Severinov, E. V. Koonin, Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. Proc. Natl. Acad. Sci. U. S. A. 115, E5307–E5316 (2018).

[0718] 41. S. F. Altschul, T. L. Madden, A. A. Sch ffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402 (1997).

[0719] 42. M. Steinegger, J. S ding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat. Biotechnol. 35, 1026–1028 (2017).

[0720] 43. R. C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 5, 113 (2004).

[0721] 44. M. Steinegger, M. Meier, M. Mirdita, H. V hringer, S. J. Haunsberger, J. S ding, HH-suite3 for fast remote homology detection and deep protein annotation, , doi:10.1101/560029.

[0722] 45. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 5, e9490 (2010).

[0723] 46. K. Katoh, K. Misawa, K.-I. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066 (2002).

[0724] 47. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274 (2015).

[0725] 48. L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. S ding, A. N. Lupas, V. Alva, A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J. Mol. Biol. 430, 2237–2243 (2018).

[0726] 49. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49, W293–W296 (2021).

[0727] 50. A.-L. Van de Weyer, F. Monteiro, O. J. Furzer, M. T. Nishimura, V. Cevik, K. Witek, J. D. G. Jones, J. L. Dangl, D. Weigel, F. Bemm, A Species-Wide Inventory of NLR Genes and Alleles in Arabidopsis thaliana. 469 Cell. 178, 1260–1272.e14 (2019).

[0728] 51. W. Dyrka, M. Lamacchia, P. Durrens, B. Kobe, A. Daskalov, M. Paoletti, D. J. Sherman, S. J. Saupe, Diversity and variability of NOD-like receptors in fungi. Genome Biol. Evol. 6, 3137–3158 (2014).

[0729] 52. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Ž dek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. Nature. 596, 583–589 (2021).

[0730] 53. J. Pei, B. H. Kim, N. V. Grishin, PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res. 36 (2008), doi:10.1093/nar/gkn072.

[0731] 54. S. Picelli, A. K. Bj rklund, B. Reinius, S. Sagasser, G. Winberg, R. Sandberg, Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 24, 2033–2040 (2014).

[0732] 55. M. Huber, G. Faure, S. Laass, E. Kolbe, K. Seitz, C. Wehrheim, Y. I. Wolf, E. V. Koonin, J. Soppa, Translational coupling via termination-reinitiation in archaea and bacteria. Nat. Commun. 10, 4006 (2019).

[0733] 56. U. Qimron, B. Marintcheva, S. Tabor, C. C. Richardson, Genomewide screens for Escherichia coli genes affecting growth of T7 bacteriophage. Proc. Natl. Acad. Sci. U. S. A. 103, 19039–19044 (2006).

[0734] 57. A. M. Grigonyte, C. Harrison, P. R. MacDonald, A. Montero-Blay, M. Tridgett, J. Duncan, A. P. Sagona, C. Constantinidou, A. Jaramillo, A. Millard, Comparison of CRISPR and Marker-Based Methods for the Engineering of Phage T7. Viruses. 12 (2020), doi:10.3390/v12020193.

[0735] 58. T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, H. Mori, Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2, 2006.0008 (2006).

[0736] 59. R. Martin et al., Structure of the activated ROQ1 resistosome directly recognizing the pathogen effector XopQ. Science 370, eabd9993 (2020).

[0737] 60. Steczkiewicz et al., Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. Nucleic Acids Res. 40, 7016–7045 (2012).

[0738] 61. Pingoud et al., Type II restriction endonucleases: Structure and mechanism. Cell. Mol. Life Sci. 62, 685–707 (2005).

[0739] 62. Burroughs et al., Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. Genome Dyn. 3,48–65 (2007).

[0740] 63. Hatfull et al., Bacteriophages and their genomes. Curr. Opin. Virol. 1, 298–303 (2011).

[0741] 64. Ackermann et al., 5500 Phages examined in the electron microscope. Arch. Virol. 152, 227–243 (2007).

[0742] 65. Ma et al., Direct pathogen-induced assembly of an NLR immune receptor complex to form a holoenzyme. Science 370, eabe3069 (2020).

[0743] 66. Kimanius et al., New tools for automated cryo-EM single-particle analysis in RELION-4.0. Biochem. J. 478, 4169–4185 (2021).

[0744] 67. Bepler et al., Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. Nat. Methods 16, 1153–1160 (2019).

[0745] 68. Croll, T.I., ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. Acta Crystallogr. D Struct. Biol. 74, 519–530 (2018).

[0746] 69. Casañal et al., Current developments in Coot for macromolecular model building of electron cryo-microscopy and crystallographic data. Protein Sci. 29, 1069–1078 (2020).

[0747] 70. Liebschner et al., Macromolecular structure determination using x-rays, neutrons and electrons: Recent developments in Phenix. Acta Crystallogr. D Struct. Biol. 75, 861–877 (2019).

***

[0748] Various modifications and variations of the described methods, pharmaceutical compositions, and kits of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific embodiments, it will be understood that it is capable of further modifications and that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the art are intended to be within the scope of the invention. This application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure come within known customary practice within the art to which the invention pertains and may be applied to the essential features herein before set forth.

[0749] Further attributes, features, and embodiments of the present invention can be understood by reference to the following numbered aspects of the disclosed invention. Reference to disclosure in any of the preceding aspects is applicable to any preceding numbered aspect and to any combination of any number of preceding aspects, as recognized by appropriate antecedent disclosure in any combination of preceding aspects that can be made. The following numbered aspects are provided:

[0750] 1. An engineered protein comprising an effector domain, an effector activation domain, and a recognition domain, wherein binding of a target polypeptide to the recognition domain leads to activation of the effector domain via the effector activation domain, and wherein at least one of the effector domain, effector activation domain, and/or recognition domain is derived from a STAND NTPase protein.

[0751] 2. The engineered protein of aspect 1, wherein the STAND NTPase protein is an antiviral STAND (Avs).

[0752] 3. The engineered protein of aspect **Error! Reference source not found.**, wherein the Avs is an Avs1, Avs2, Avs3, or Avs4.

[0753] 4. The engineered protein of any one of aspects 1-3, wherein the effector domain is an endonuclease, a protease, a nucleosidase, hydrolase, or caspase-like domain.

[0754] 5. The engineered protein of any one of aspects 1-4, wherein the effector activation domain is an NTPase.

[0755] 6. The engineered protein of any one of aspects 1-5, wherein the recognition domain is engineered to recognize a target polypeptide other than a target polypeptide of a wild-type STAND NTPase protein.

[0756] 7. The engineered protein of aspect 6, wherein the recognition domain comprises one or more tetratricopeptide repeat (TPR) domains.

[0757] 8. The engineered protein of any one of aspects 1-7, wherein a microbe comprises the target polypeptide.

[0758] 9. The engineered protein of aspect 8, wherein the microbe is part of a microbiome.

[0759] 10. The engineered protein of any one of aspects 1-9, wherein the target polypeptide is a phage polypeptide.

[0760] 11. An oligomer comprising two or more engineered proteins of any one of apects 1-10.

[0761] 12. The oligomer of aspect 11, wherein the oligomer is a tetramer, a trimer, or a dimer.

[0762] 13. The oligomer of any one of aspects 11-12, wherein each of the two or more engineered proteins are the same.

[0763] 14. The oligomer of any one of aspects 11-12, wherein at least two of the two or more engineered proteins are different.

[0764] 15. The oligomer of any one of aspects 11-12, wherein each of the two or more engineered proteins are different.

[0765] 16. A detection composition comprising:

a. an engineered protein of any one of aspects 1-10 or an oligomer thereof;

b. a detection construct,

wherein binding of a target polypeptide to the recognition domain activates the effector domain and mediates effector domain modification of the detection construct resulting in generation of a detectable signal.

[0766] 17. A polynucleotide encoding the engineered protein of any one of aspects 1-10.

[0767] 18. A polynucleotide encoding component (a), component (b), or both of the detection composition of aspect 16.

[0768] 19. A vector or vector system comprising the polynucleotide of any one of aspects 17-18.

[0769] 20. A cell or cell population comprising an engineered protein of any one of aspects 1-10, an oligomer of any one of aspects 11-15, a detection composition of aspect 16, a polynucleotide of any one of aspects 17-18, a vector or vector system of aspect 19, or any combination thereof.

[0770] 21. A formulation comprising an engineered protein of any one of aspects 1-10, an oligomer of any one of aspects 11-15, a detection composition of aspect 16, a polynucleotide of any one of aspects 17-18, a vector or vector system of aspect 19, a cell or cell population of aspect 20, or any combination thereof; and optionally a pharmaceutically acceptable carrier.

[0771] 22. A method of modifying a target molecule and/or cell comprising:

delivering an engineered protein of any one of aspects 1-10, an oligomer of any one of aspects 11-15, a polynucleotide of aspect 17, a vector or vector system of aspect 19, a formulation thereof, or any combination thereof to the target molecule and/or cell, wherein the target molecule and/or cell is or comprises a target polypeptide; and

activating an effector domain of the engineered protein by allowing binding of the target polypeptide to the recognition domain thereby activating the effector domain via the effector activation domain, wherein effector domain activity modifies the target molecule and/or cell.

[0772] 23. The method of aspect 22, wherein delivering comprises *in vitro*, *ex vivo*, or *in vivo* delivery.

[0773] 24. A method of detecting a target molecule and/or cell, the method comprising:

combining a detection composition of aspect 10 or a formulation thereof and a sample or component thereof; and

activating an effector domain of the engineered protein via binding of a target polypeptide in the sample to the recognition domain thereby mediating effector domain modification of the detection construct and generation of a detectable signal.

[0774] 25. The method of aspect 18, wherein the method is performed in whole or in part *in vitro*, *ex vivo*, or *in vivo*.

[0775] 26. A method of modifying a microbiome structure comprising:

introducing an engineered protein of any one of aspects 8-10 into a microbiome, wherein activation of the effector domain via binding of a target polypeptide of one or more microbes in the microbiome to the recognition domain results in modification of the one or more microbes thereby modifying the microbiome structure.

[0776] 27. A method of engineering phage-resistant bacteria comprising:

expressing an engineered protein of aspect 10 or an oligomer comprising one or more engineered proteins of aspect 10 in a bacterium or a bacteria population.

[0777] 28. A method of cargo delivery comprising:

delivering to a cell

(a) an engineered protein of any one of aspects 1-10; and

(b) a cargo;

(c) a detection composition; or

(d) any combination thereof;

wherein the engineered protein comprises the cargo or wherein the cargo comprises the target polypeptide, and

wherein activation of the effector domain by binding of the target polypeptide to the recognition domain results in delivery of the cargo and optionally activation of the detection construct thereby monitoring cargo delivery.

[0778] 29. The method of cargo delivery of aspect 28, wherein the cell comprises the target polypeptide.

CLAIMS

What is claimed is:

1. An engineered protein comprising an effector domain, an effector activation domain, and a recognition domain, wherein binding of a target polypeptide to the recognition domain leads to activation of the effector domain via the effector activation domain, and wherein at least one of the effector domain, effector activation domain, and/or recognition domain is derived from a STAND NTPase protein.

2. The engineered protein of claim 1, wherein the STAND NTPase protein is an antiviral STAND (Avs).

3. The engineered protein of claim **Error! Reference source not found.**, wherein the Avs is an Avs1, Avs2, Avs3, or Avs4.

4. The engineered protein of claim 1, wherein the effector domain is an endonuclease, a protease, a nucleosidase, hydrolase, or caspase-like domain.

5. The engineered protein of claim 1, wherein the effector activation domain is an NTPase.

6. The engineered protein of claim 1, wherein the recognition domain is engineered to recognize a target polypeptide other than a target polypeptide of a wild-type STAND NTPase protein.

7. The engineered protein of claim 6, wherein the recognition domain comprises one or more tetratricopeptide repeat (TPR) domains.

8. The engineered protein of claim 1, wherein a microbe comprises the target polypeptide.

9. The engineered protein of claim 8, wherein the microbe is part of a microbiome.

10. The engineered protein of claim 1, wherein the target polypeptide is a phage polypeptide.

11. An oligomer comprising two or more engineered proteins of any one of claims 1-10.

12. The oligomer of claim 11, wherein the oligomer is a tetramer, a trimer, or a dimer.

13. The oligomer of claim 11, wherein each of the two or more engineered proteins are the same.

14. The oligomer of claim 11, wherein at least two of the two or more engineered proteins are different.

15. The oligomer of claim 11, wherein each of the two or more engineered proteins are different.

16. A detection composition comprising:

a. an engineered protein of any one of claims 1-10 or an oligomer thereof;

b. a detection construct,

wherein binding of a target polypeptide to the recognition domain activates the effector domain and mediates effector domain modification of the detection construct resulting in generation of a detectable signal.

17. A polynucleotide encoding the engineered protein of any one of claims 1-10.

18. A polynucleotide encoding component (a), component (b), or both of the detection composition of claim 16.

19. A vector or vector system comprising the polynucleotide of any one of claims 17-18.

20. A cell or cell population comprising an engineered protein of any one of claims 1-10, an oligomer of any one of claims 11-15, a detection composition of claim 16, a polynucleotide of any one of claims 17-18, a vector or vector system of claim 19, or any combination thereof.

21. A formulation comprising an engineered protein of any one of claims 1-10, an oligomer of any one of claims 11-15, a detection composition of claim 16, a polynucleotide of any one of claims 17-18, a vector or vector system of claim 19, a cell or cell population of claim 20, or any combination thereof; and

optionally a pharmaceutically acceptable carrier.

22. A method of modifying a target molecule and/or cell comprising:

delivering an engineered protein of any one of claims 1-10, an oligomer of any one of claims 11-15, a polynucleotide of claim 17, a vector or vector system of claim 19, a formulation thereof, or any combination thereof to the target molecule and/or cell, wherein the target molecule and/or cell is or comprises a target polypeptide; and activating an effector domain of the engineered protein by allowing binding of the target polypeptide to the recognition domain thereby activating the effector domain via the effector activation domain, wherein effector domain activity modifies the target molecule and/or cell.

23. The method of claim 22, wherein delivering comprises *in vitro*, *ex vivo*, or *in vivo* delivery.

24. A method of detecting a target molecule and/or cell, the method comprising:

combining a detection composition of claim 10 or a formulation thereof and a sample or component thereof; and

activating an effector domain of the engineered protein via binding of a target polypeptide in the sample to the recognition domain thereby mediating effector domain modification of the detection construct and generation of a detectable signal.

25. The method of claim 18, wherein the method is performed in whole or in part *in vitro*, *ex vivo*, or *in vivo*.

26. A method of modifying a microbiome structure comprising:

introducing an engineered protein of any one of claims 8-10 into a microbiome, wherein activation of the effector domain via binding of a target polypeptide of one or more microbes in the microbiome to the recognition domain results in modification of the one or more microbes thereby modifying the microbiome structure.

27. A method of engineering phage-resistant bacteria comprising:

expressing an engineered protein of claim 10 or an oligomer comprising one or more engineered proteins of claim 10 in a bacterium or a bacteria population.

28. A method of cargo delivery comprising:

delivering to a cell

a. an engineered protein of any one of claims 1-10; and
b. a cargo;
c. a detection composition; or
d. any combination thereof;

wherein the engineered protein comprises the cargo or wherein the cargo comprises the target polypeptide, and

wherein activation of the effector domain by binding of the target polypeptide to the recognition domain results in delivery of the cargo and optionally activation of the detection construct thereby monitoring cargo delivery.

29. The method of cargo delivery of claim 28, wherein the cell comprises the target polypeptide.

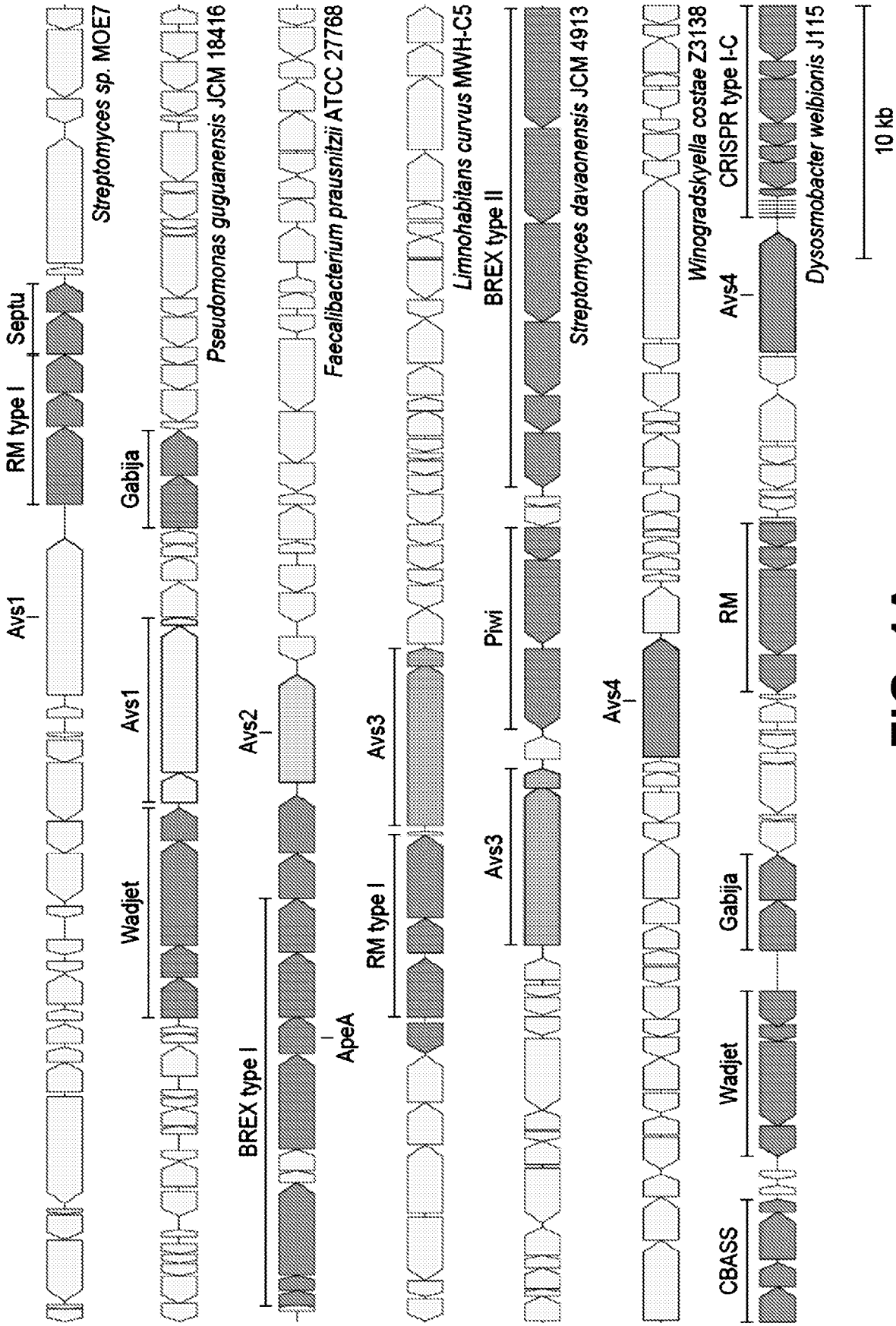

Streptomyces sp. MOE7
Avs1
RM type I
Septu
Pseudomonas guguanensis JCM 18416
Gabija
Avs1
Wadjet
Faecalibacterium prausnitzii ATCC 27768
Avs2
BREX type I
ApeA
Limnohabitans curvus MWH-C5
Avs3
RM type I
BREX type II
Piwi
Streptomyces davaonensis JCM 4913
Avs3
Avs4
Winogradskyella costae Z3138
Avs4
RM
CRISPR type I-C
Dysosmobacter welbionis J115
Gabija
Wadjet
CBASS
10 kb

FIG. 1A

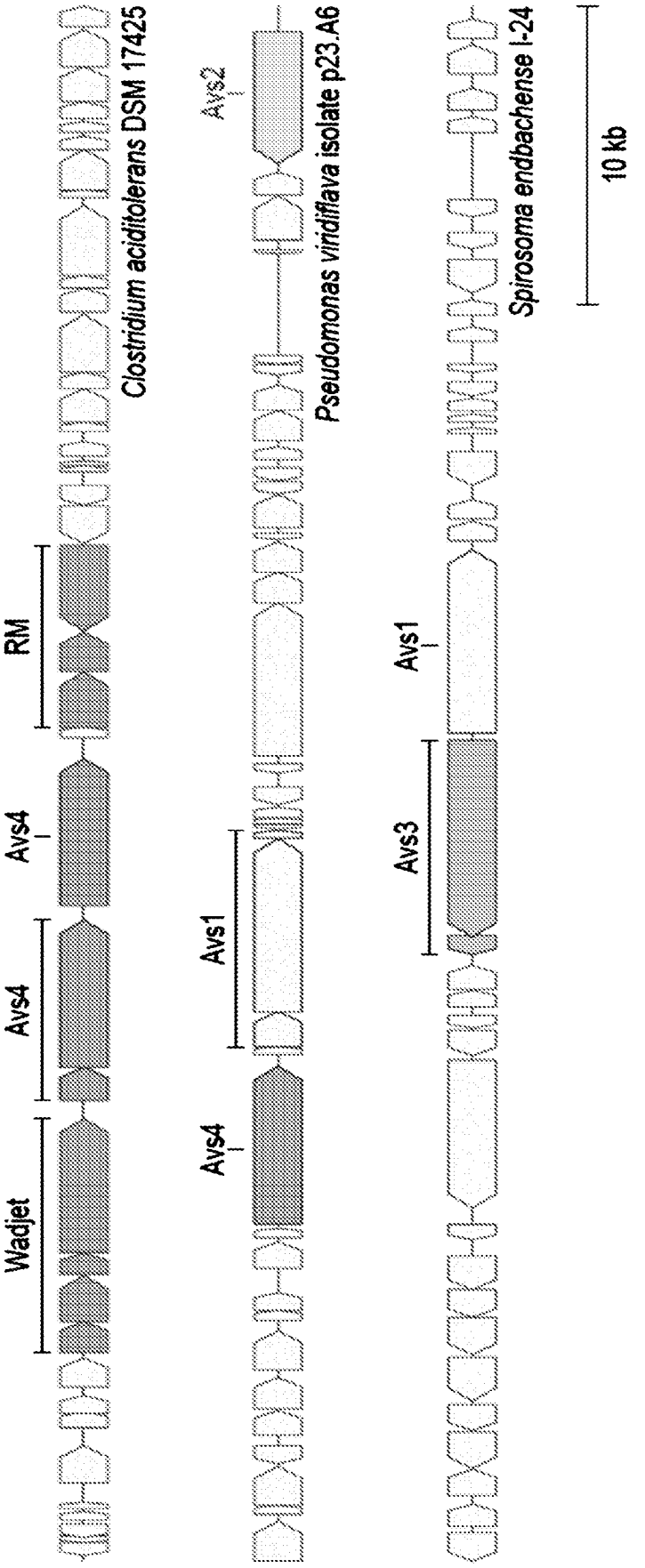

Clostridium aciditolerans DSM 17425
Wadjet
Avs4
Avs4
RM
Pseudomonas viridiflava isolate p23.A6
Avs4
Avs1
Avs2
Spirosoma endbachense I-24
Avs3
Avs1
10 kb

FIG. 1B

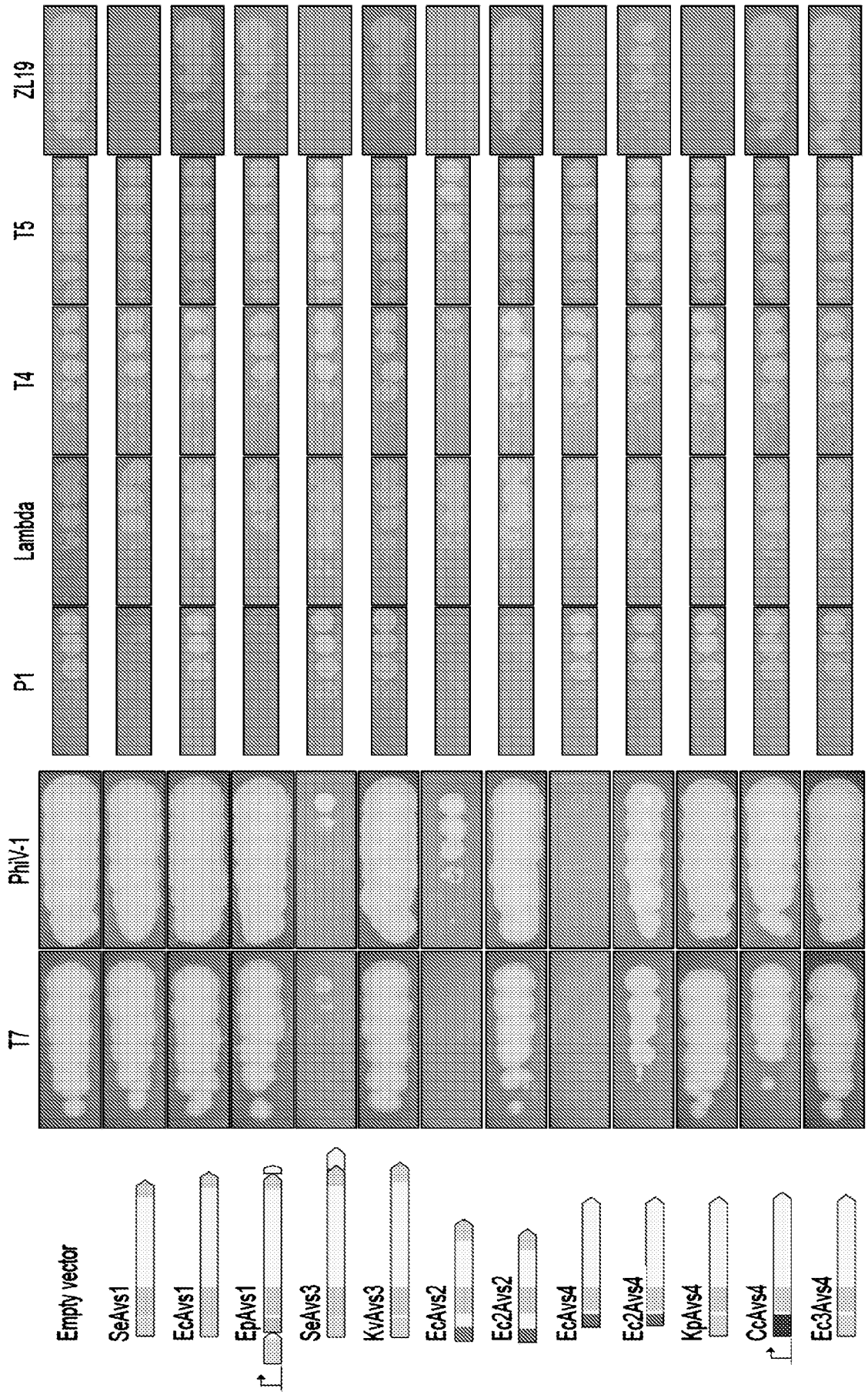

T7
PhiV-1
P1
Lambda
T4
T5
ZL19
E. coli strain C
E. coli K12 (ATCC 25404)
Empty vector
SeAvs1
EcAvs1
EpAvs1
SeAvs3
KvAvs3
EcAvs2
Ec2Avs2
EcAvs4
Ec2Avs4
KpAvs4
CcAvs4
Ec3Avs4
1000 aa

FIG. 2

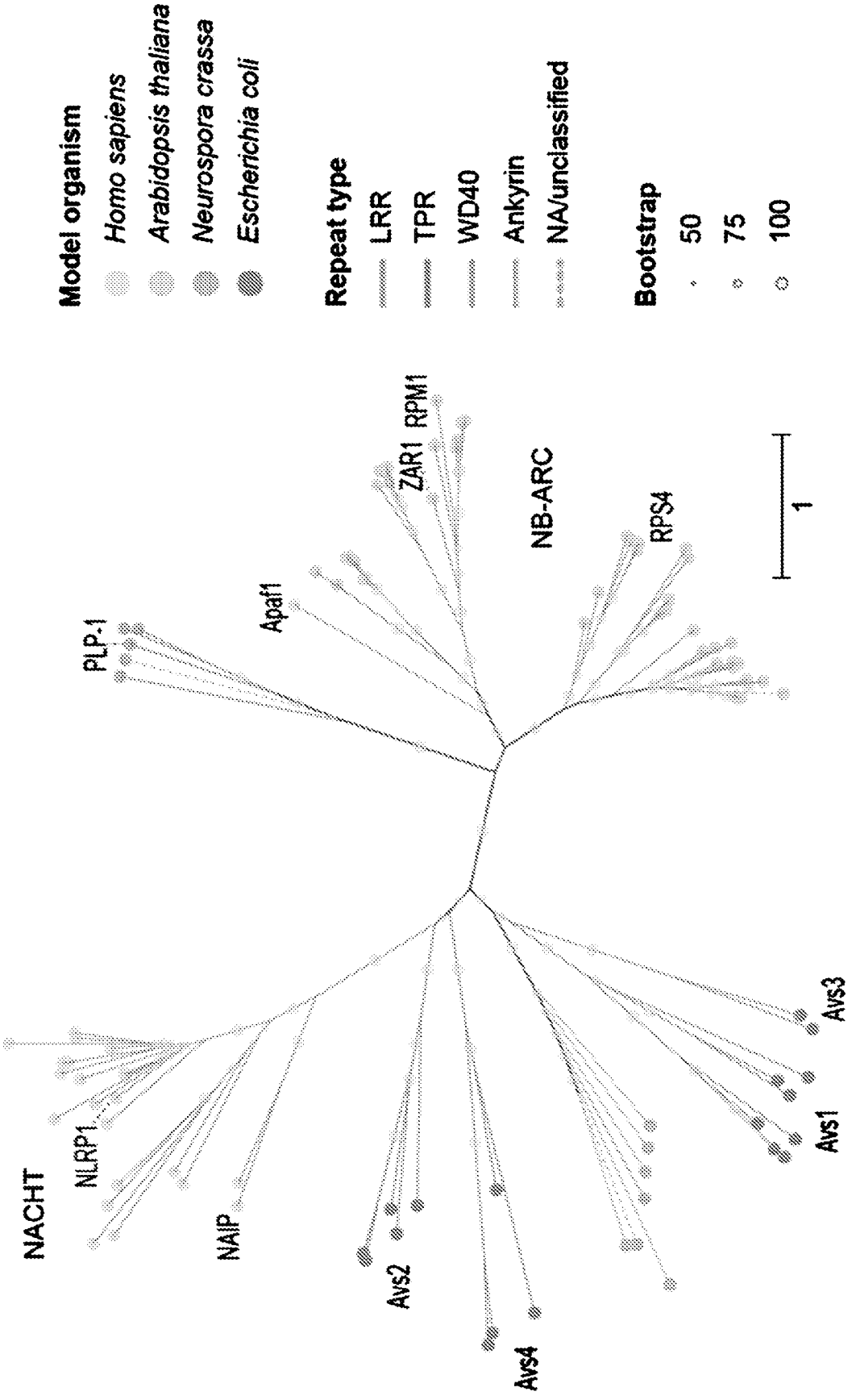

Model organism
Homo sapiens
Arabidopsis thaliana
Neurospora crassa
Escherichia coli
Repeat type
LRR
TPR
WD40
Ankyrin
NA/unclassified
Bootstrap
50
75
100
NACHT
NLRP1
NAIP
Avs2
Avs4
Avs1
Avs3
PLP-1
Apaf1
ZAR1
RPM1
NB-ARC
RPS4
1

FIG. 3A

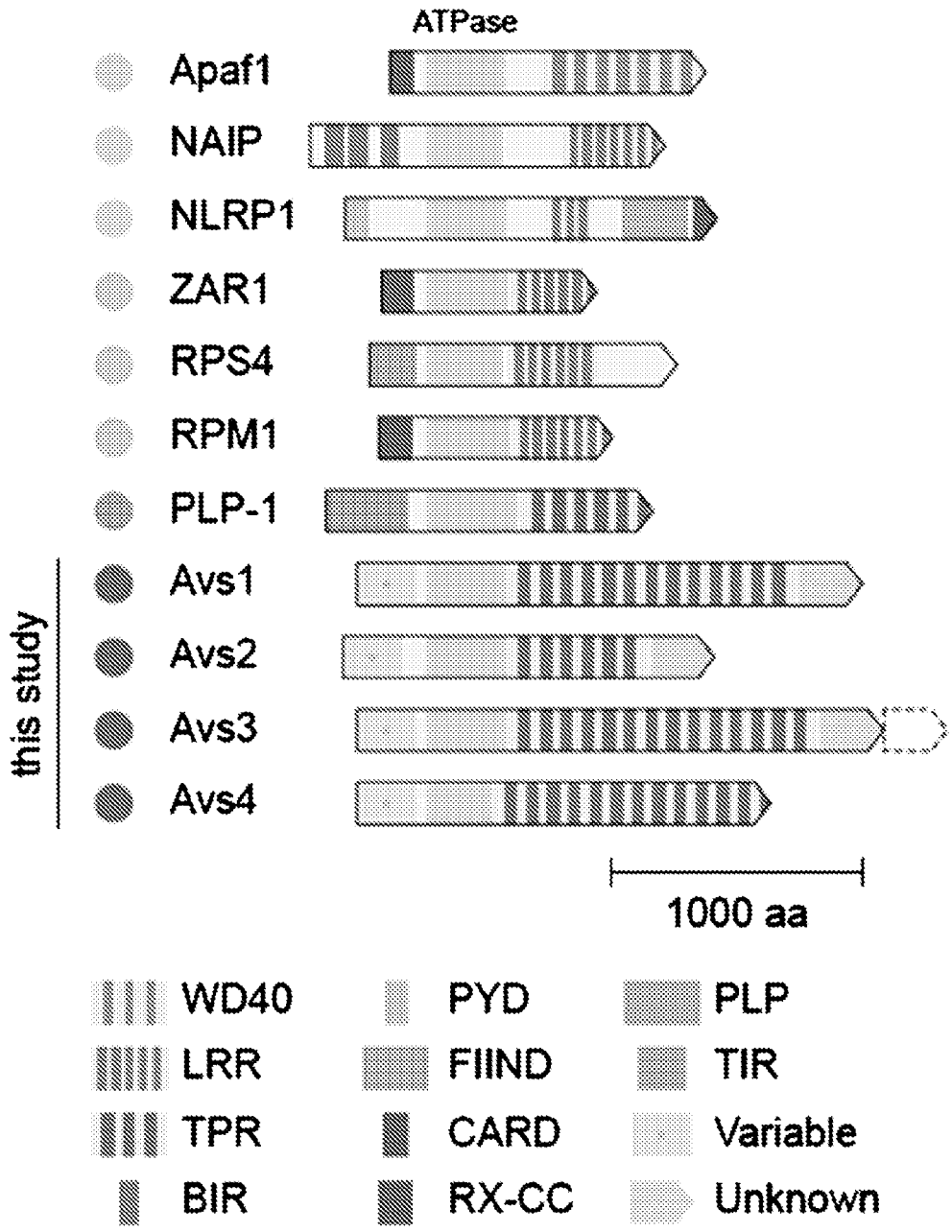

ATPase
Apaf1
NAIP
NLRP1
ZAR1
RPS4
RPM1
PLP-1
this study
Avs1
Avs2
Avs3
Avs4
1000 aa
WD40
LRR
TPR
BIR
PYD
FIIND
CARD
RX-CC
PLP
TIR
Variable
Unknown

FIG. 3B

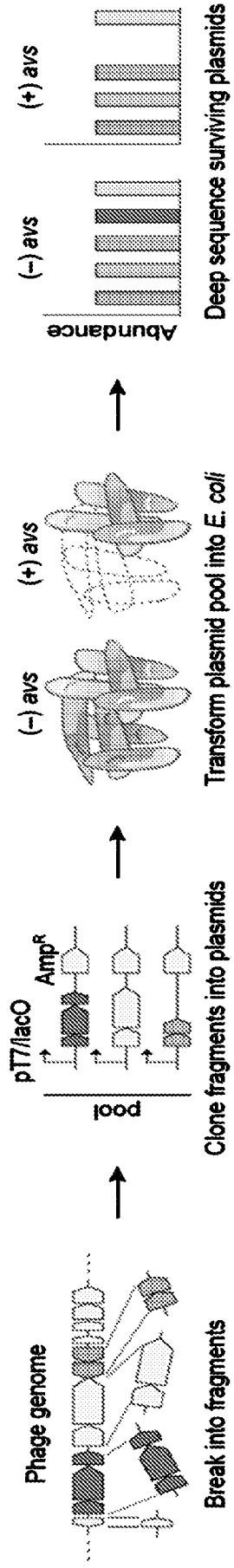

Phage genome
Break into fragments
pool
pT7/lacO
Amp^R
Clone fragments into plasmids
(–) avs
(+) avs
Transform plasmid pool into E. coli
Abundance
(–) avs
(+) avs
Deep sequence surviving plasmids

FIG. 3C

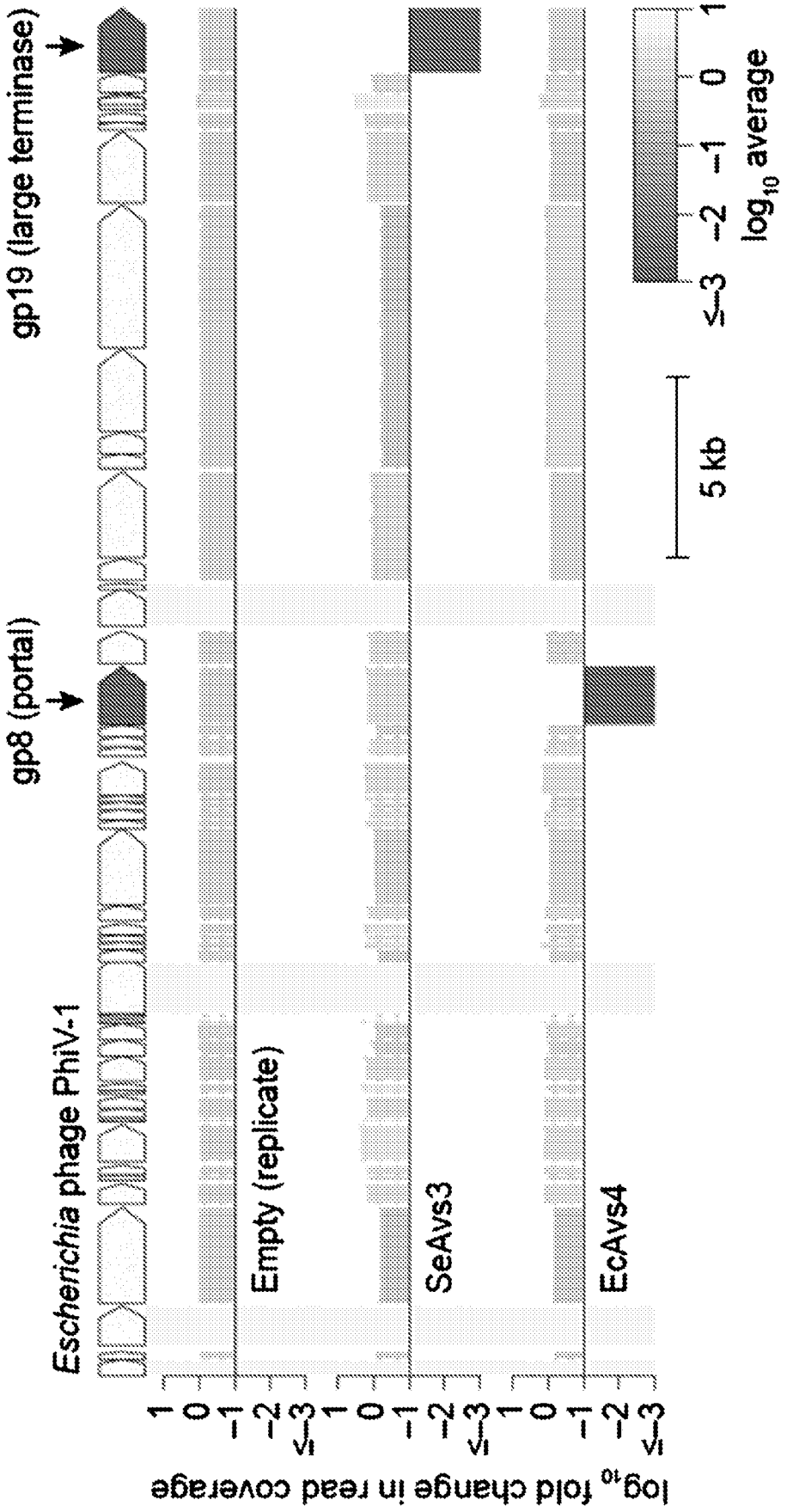

gp19 (large terminase)
gp8 (portal)
Escherichia phage PhiV-1
Empty (replicate)
SeAvs3
EcAvs4
log10 fold change in read coverage
1
0
−1
−2
≤−3
5 kb
log10 average

FIG. 3D

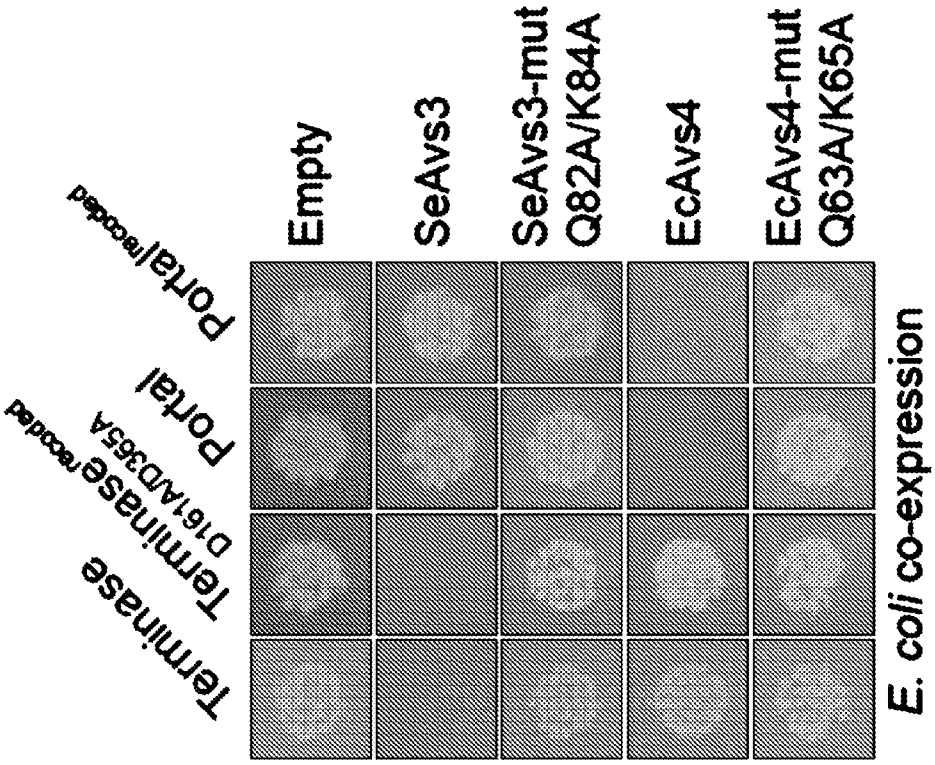

Empty
SeAvs3
SeAvs3-mut Q82A/K84A
EcAvs4
EcAvs4-mut Q63A/K65A
E. coli co-expression
Terminase
Terminase recoded D16A/D365A
Portal
Portal recoded

FIG. 3F

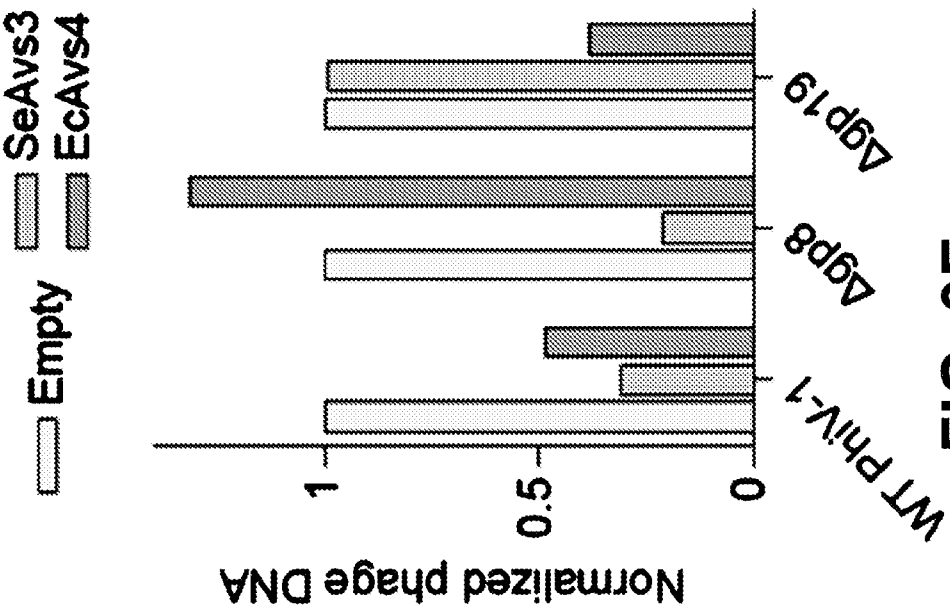

Empty
SeAvs3
EcAvs4
Normalized phage DNA
0
0.5
1
WT PhiV-1
Δgp8
Δgp19

FIG. 3E

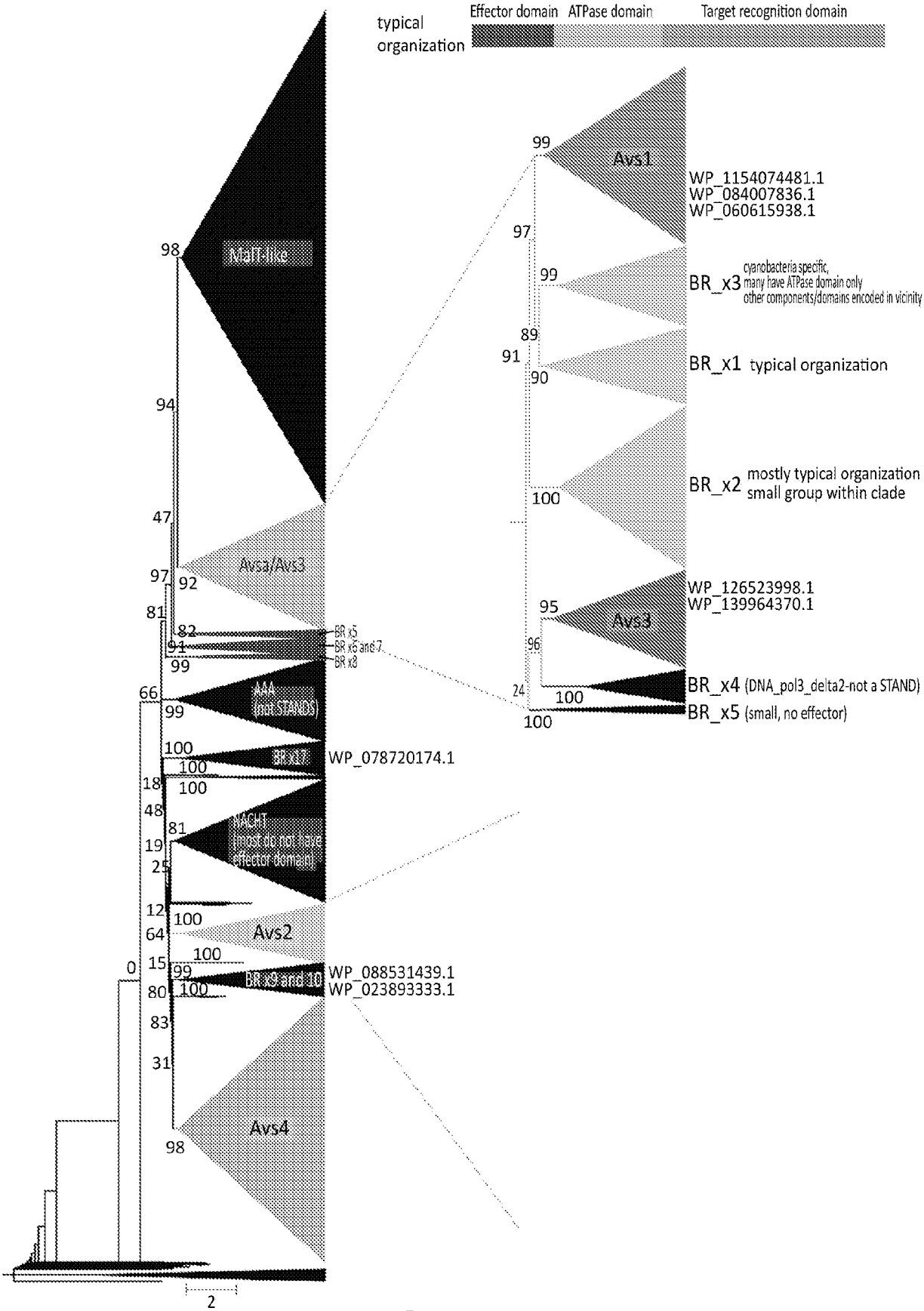

typical organization
Effector domain
ATPase domain
Target recognition domain
Avs1
WP_1154074481.1
WP_084007836.1
WP_060615938.1
BR_x3 cyanobacteria specific, many have ATPase domain only other components/domains encoded in vicinity
BR_x1 typical organization
BR_x2 mostly typical organization small group within clade
Avs3
WP_126523998.1
WP_139964370.1
BR_x4 (DNA_pol3_delta2-not a STAND)
BR_x5 (small, no effector)
Maff-like
Avs3/Avs3
BR x5
BR x6 and 7
BR x8
AAA (not STANDS)
BR x11
WP_078720174.1
NACHT (most do not have effector domain)
Avs2
BR x9 and 10
WP_088531439.1
WP_023893333.1
Avs4
2

FIG. 4A

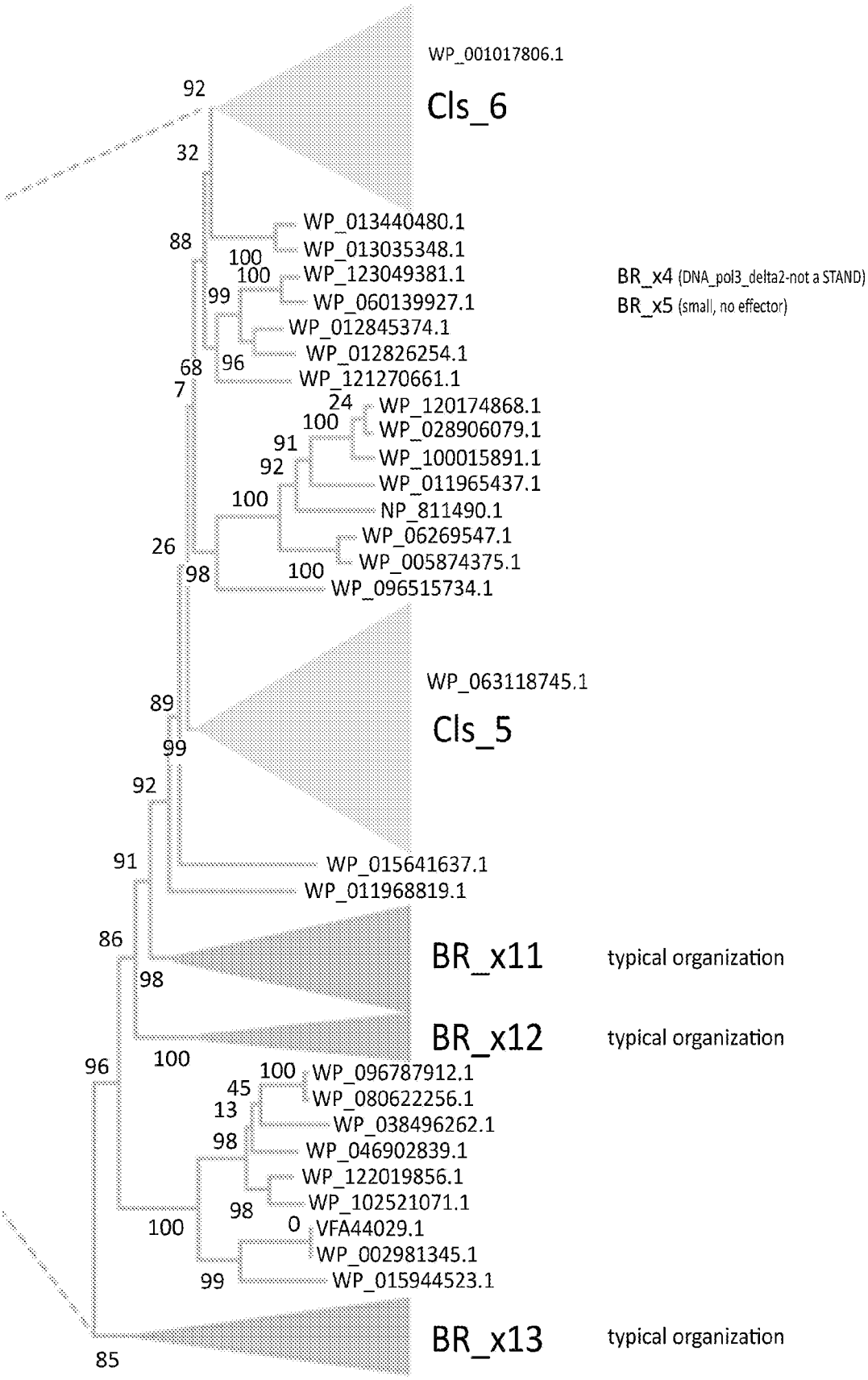


FIG. 4A Continued

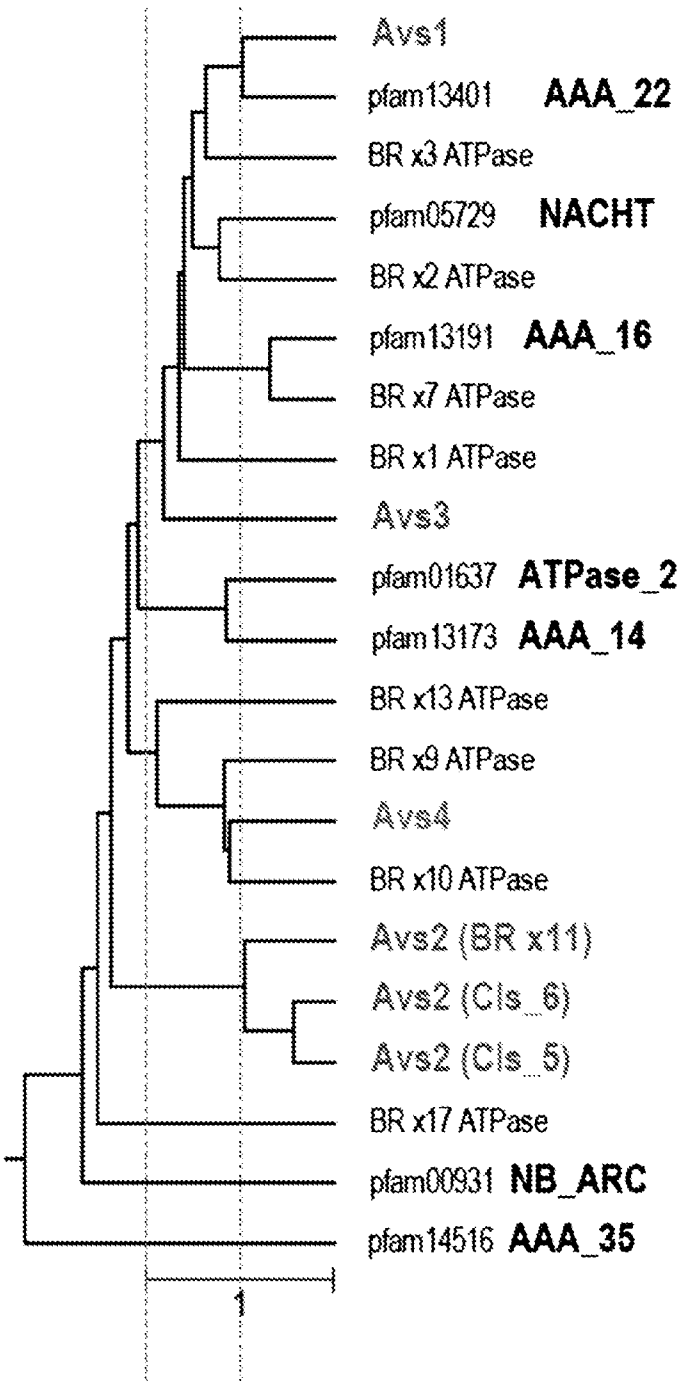

Avs1
pfam13401 AAA_22
BR x3 ATPase
pfam05729 NACHT
BR x2 ATPase
pfam13191 AAA_16
BR x7 ATPase
BR x1 ATPase
Avs3
pfam01637 ATPase_2
pfam13173 AAA_14
BR x13 ATPase
BR x9 ATPase
Avs4
BR x10 ATPase
Avs2 (BR x11)
Avs2 (Cls_6)
Avs2 (Cls_5)
BR x17 ATPase
pfam00931 NB_ARC
pfam14516 AAA_35
1

FIG. 4B

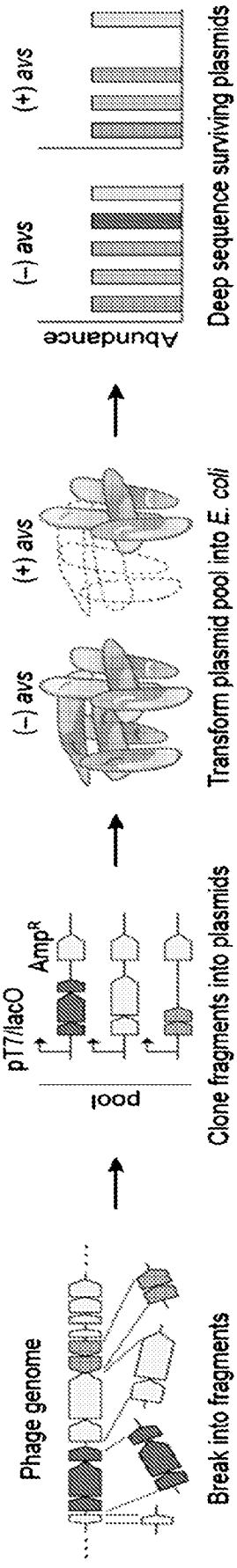

Phage genome
Break into fragments
pool
pT7/lacO
Amp^R
Clone fragments into plasmids
(−) avs
(+) avs
Transform plasmid pool into E. coli
Abundance
(−) avs
(+) avs
Deep sequence surviving plasmids

FIG. 5A

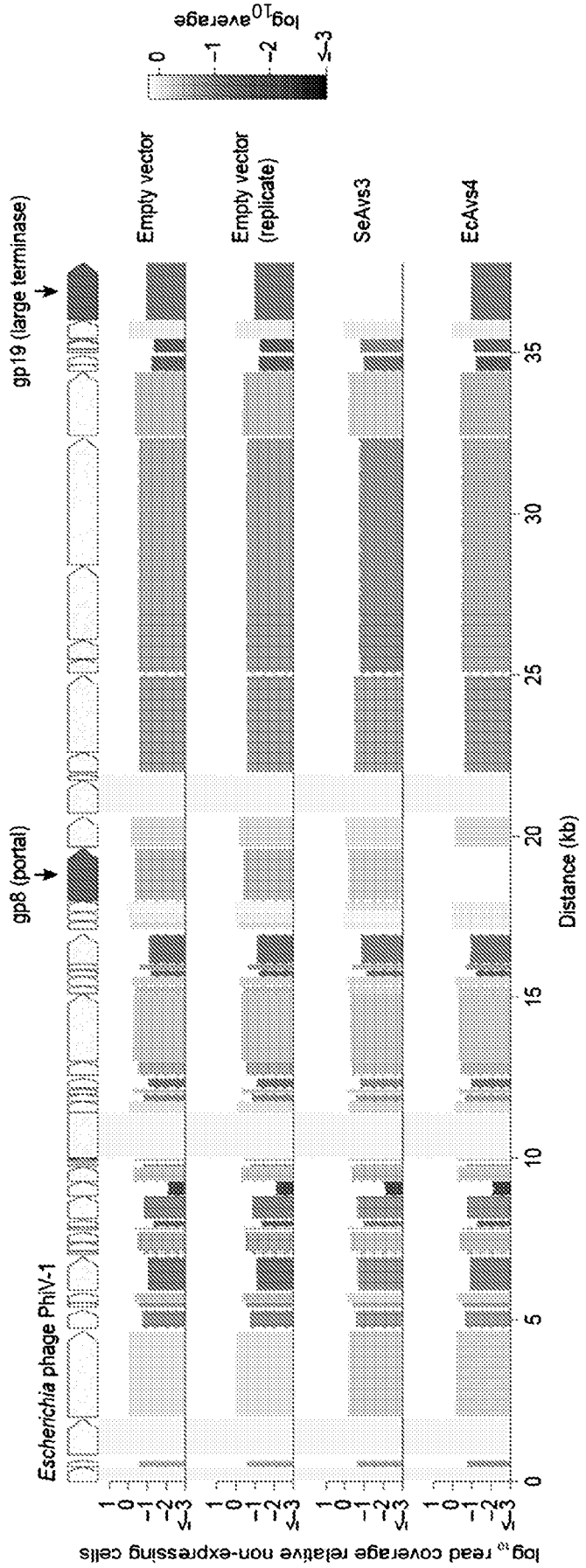

log10 average
0
-1
-2
≤-3
Empty vector
Empty vector (replicate)
SeAvs3
EcAvs4
gp19 (large terminase)
gp8 (portal)
Escherichia phage PhiV-1
Distance (kb)
0
5
10
15
20
25
30
35
log2 read coverage relative non-expressing cells
1
0
-1
-2
≤-3

FIG. 5B

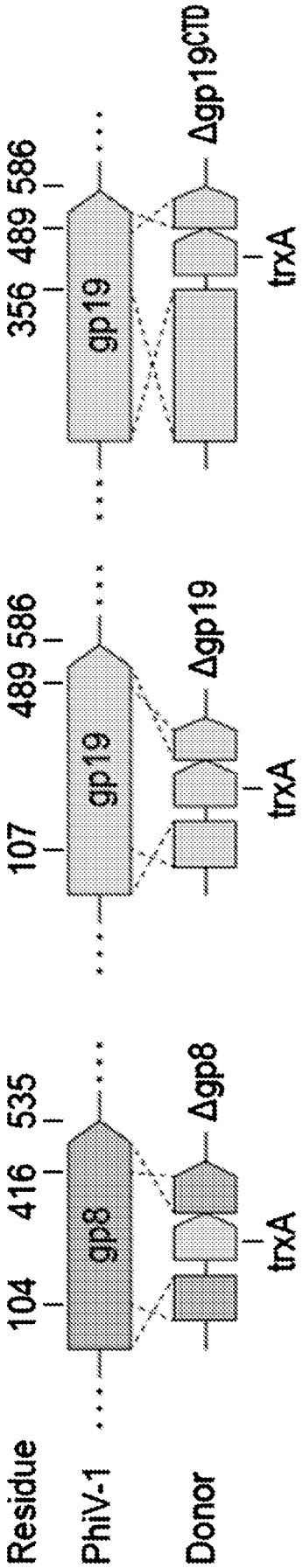

Residue
PhiV-1
Donor
104
416
535
gp8
Δgp8
trxA
107
489
586
gp19
Δgp19
trxA
356
489
586
gp19
Δgp19CTD
trxA

FIG. 6A

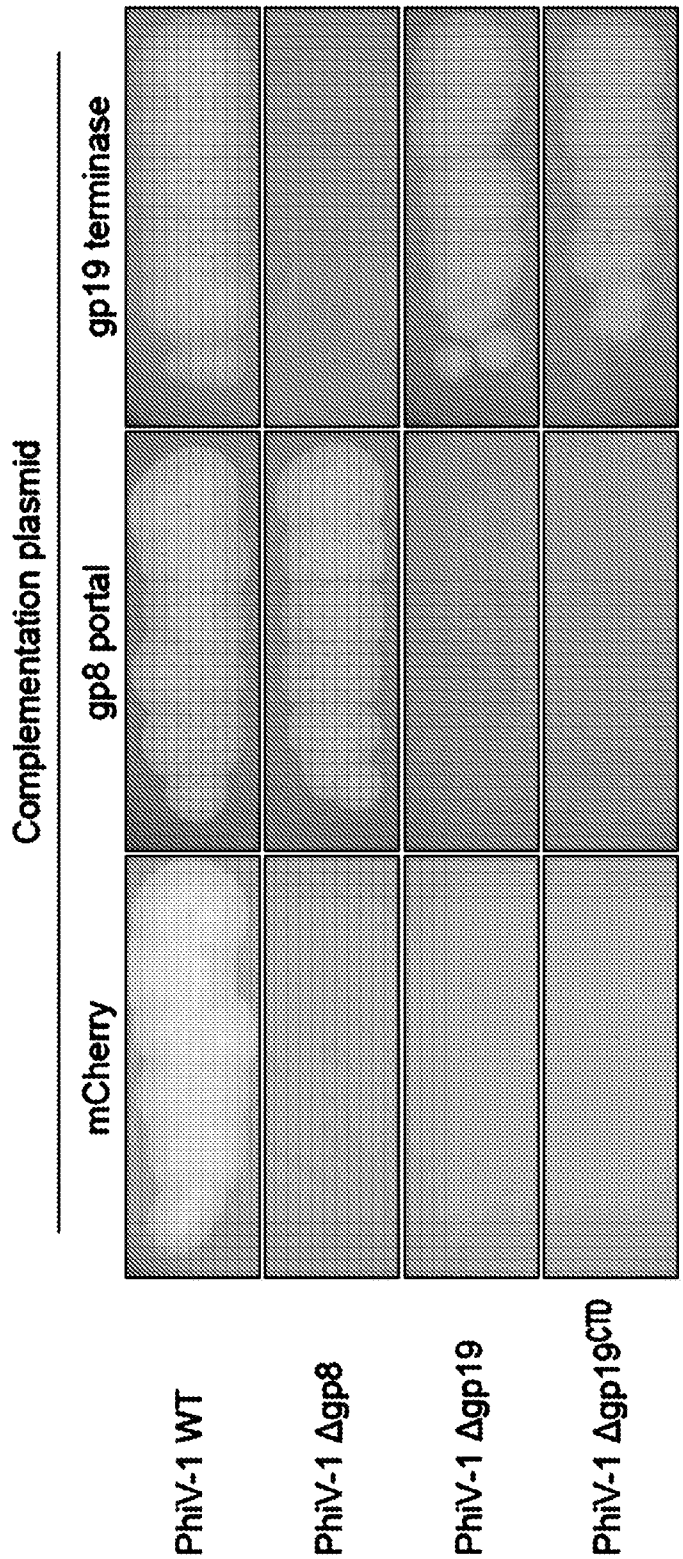

Complementation plasmid
mCherry
gp8 portal
gp19 terminase
PhiV-1 WT
PhiV-1 Δgp8
PhiV-1 Δgp19
PhiV-1 Δgp19CTD

FIG. 6B

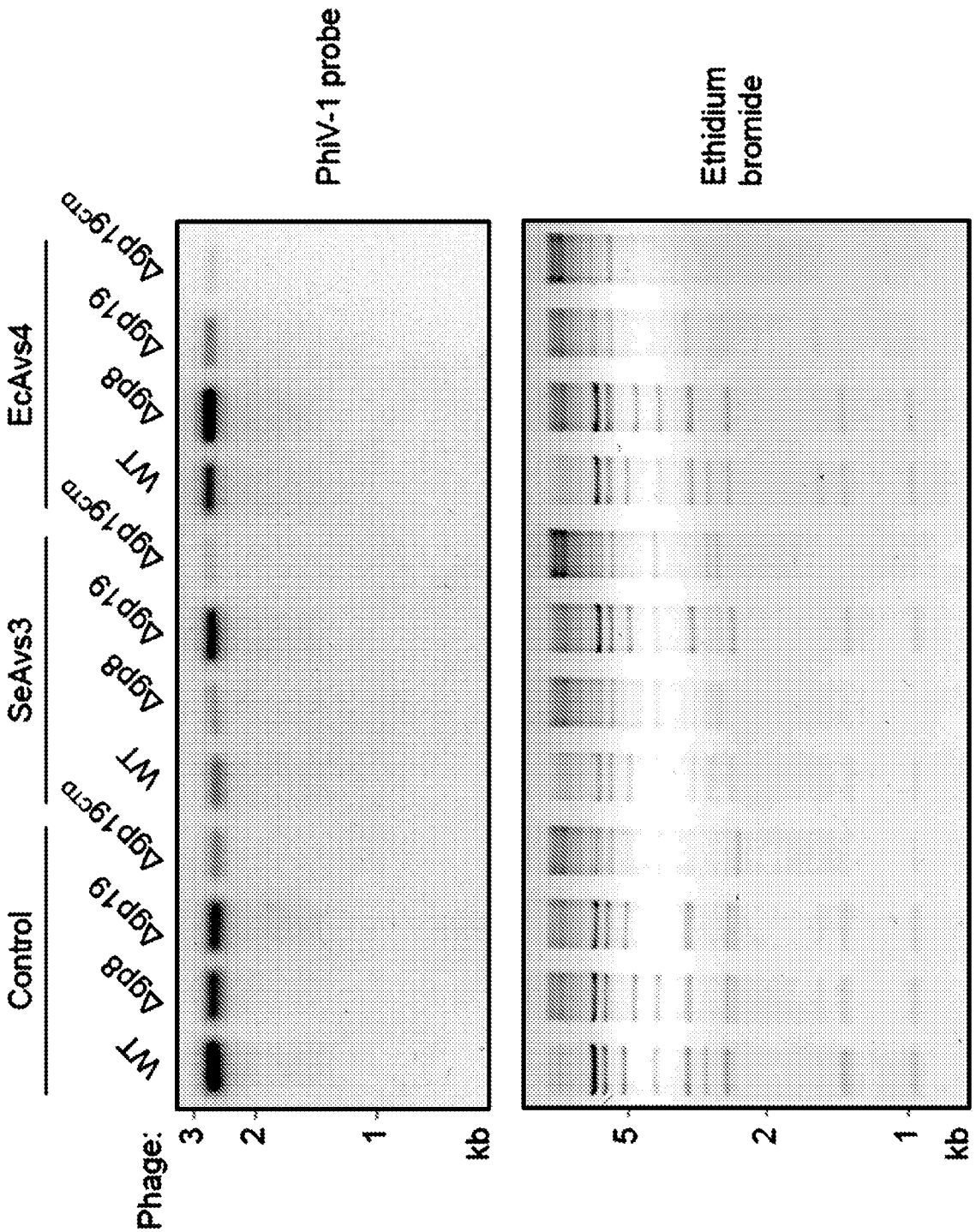

PhiV-1 probe
Ethidium bromide
Phage:
WT
Δgp8
Δgp19
Δgp19cm
WT
Δgp8
Δgp19
Δgp19cm
WT
Δgp8
Δgp19
Δgp19cm
Control
SeAvs3
EcAvs4
3
2
1
kb
5
2
1
kb

FIG. 6C

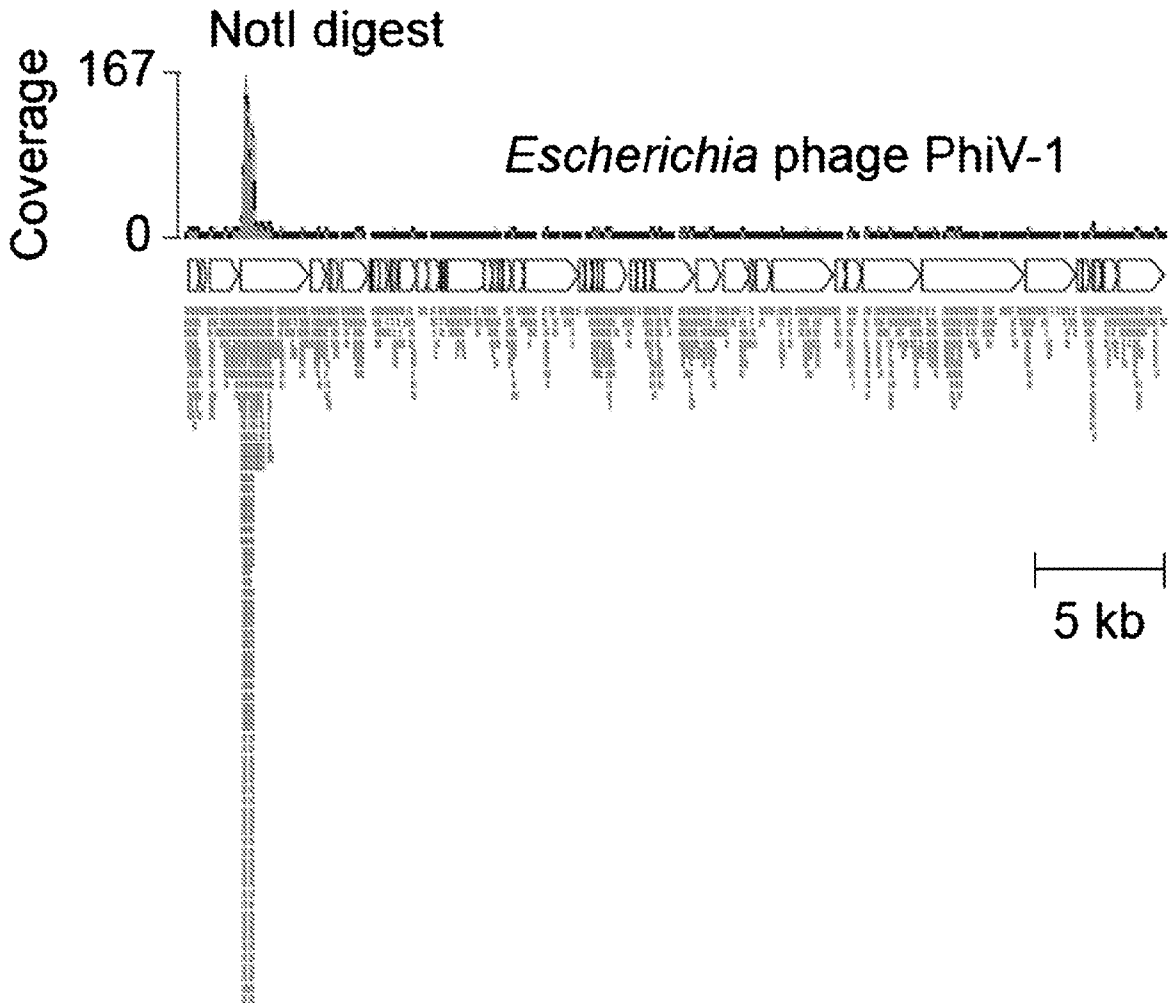


FIG. 6D

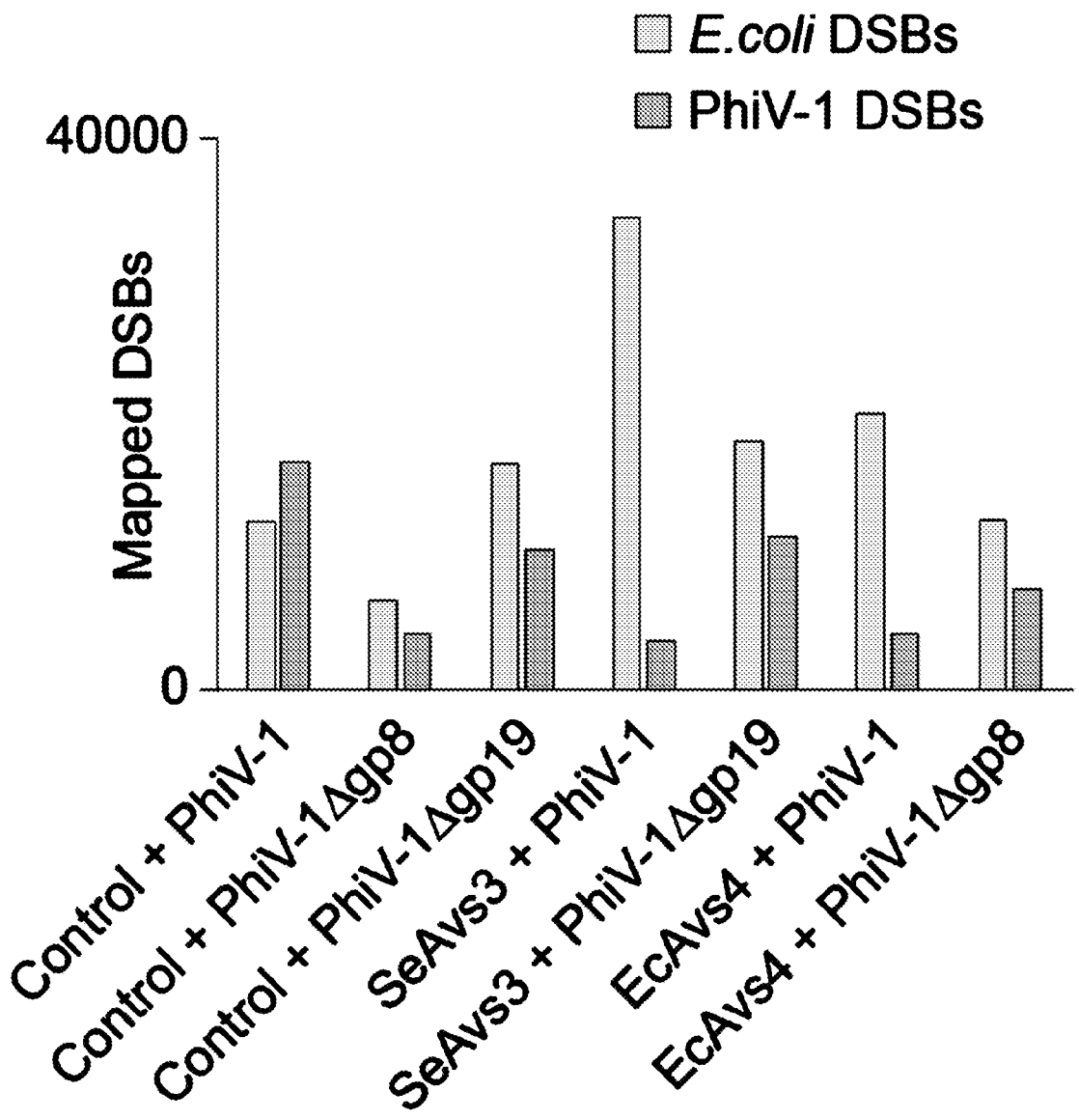

E.coli DSBs
PhiV-1 DSBs
40000
0
Mapped DSBs
Control + PhiV-1
Control + PhiV-1Δgp8
Control + PhiV-1Δgp19
SeAvs3 + PhiV-1
SeAvs3 + PhiV-1Δgp19
EcAvs4 + PhiV-1
EcAvs4 + PhiV-1Δgp8

FIG. 6E

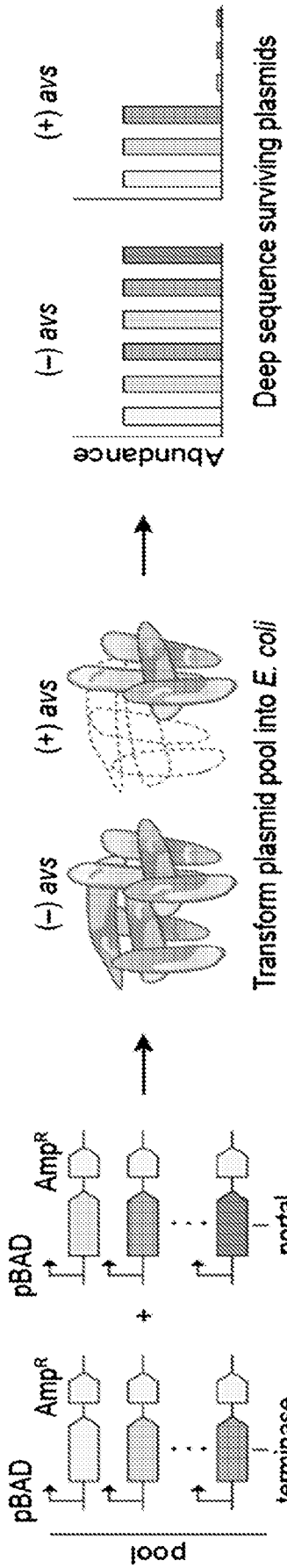

pool
pBAD
Amp^R
terminase
+
pBAD
Amp^R
portal
(–) avs
(+) avs
Transform plasmid pool into E. coli
(–) avs
(+) avs
Abundance
Deep sequence surviving plasmids

FIG. 7A

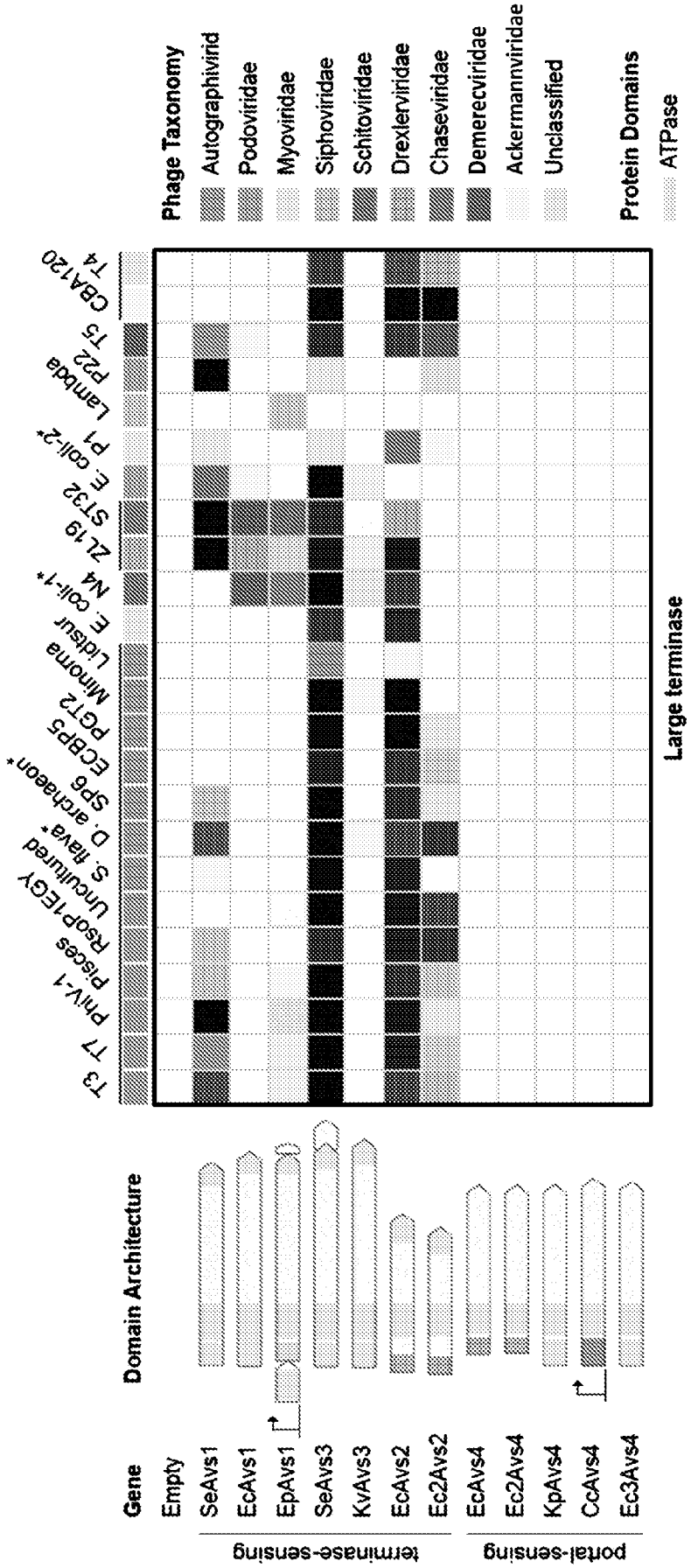

Phage Taxonomy
Autographiviridae
Podoviridae
Myoviridae
Siphoviridae
Schitoviridae
Drexlerviridae
Chaseviridae
Demerecviridae
Ackermannviridae
Unclassified
Protein Domains
ATPase
Large terminase
T3
T7
phiV-1
Pisces
RSoP1EGV
Uncultured
S. flava*
D. archaeon*
SP6
ECBP5
PGT2
Mimoma
Lidisur
E. coli-1*
N4
ZL-19
ST32
E. coli-2*
P1
Lambda
P22
T5
CBA120
T4
Gene
Domain Architecture
Empty
SeAvs1
EcAvs1
EpAvs1
SeAvs3
KvAvs3
EcAvs2
Ec2Avs2
EcAvs4
Ec2Avs4
KpAvs4
CcAvs4
Ec3Avs4
terminase-sensing
portal-sensing

FIG. 7B

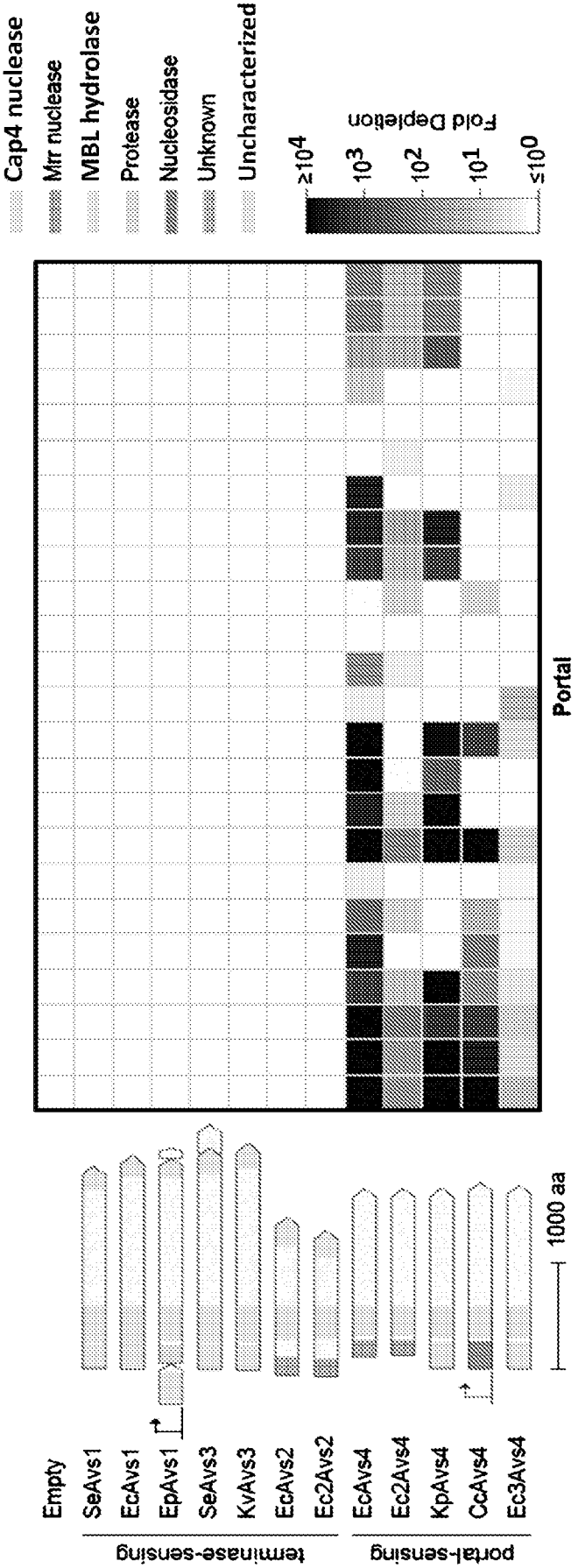


FIG. 7B Continued

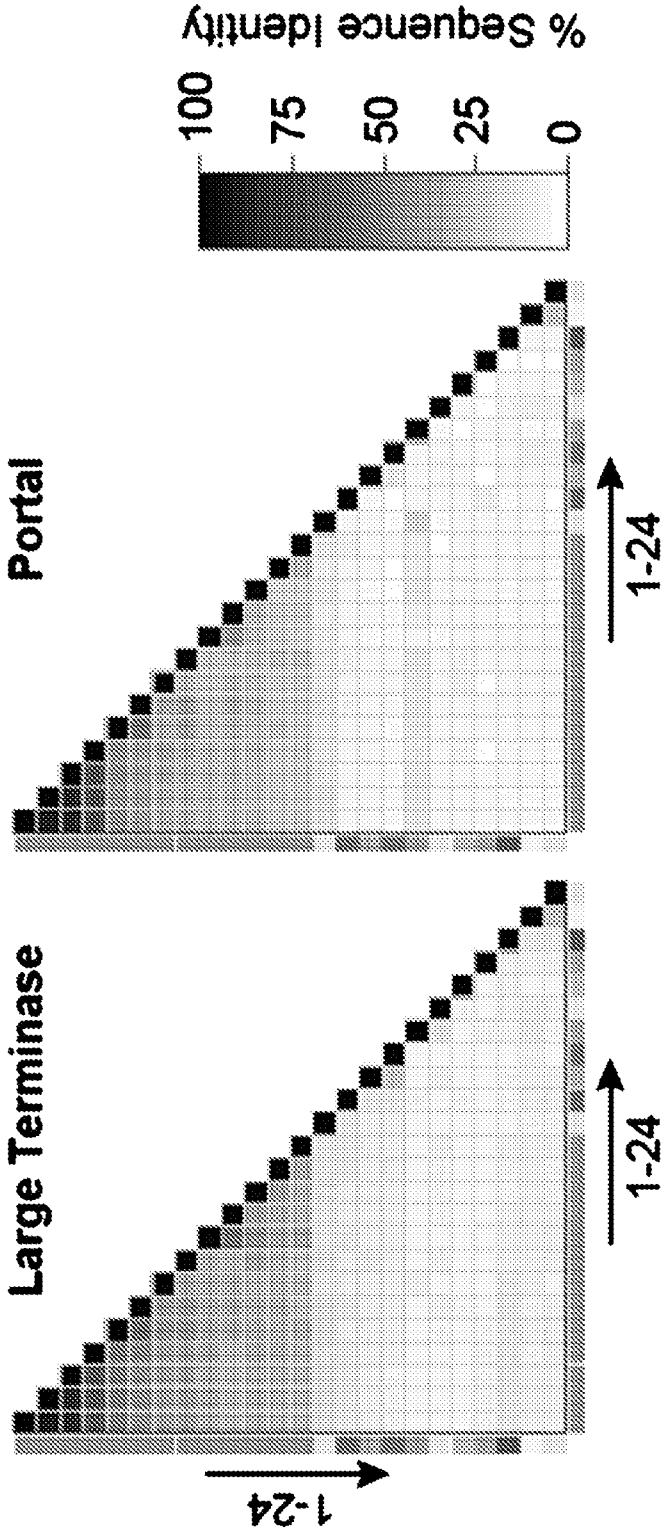

% Sequence Identity
100
75
50
25
0
Portal
1-24
Large Terminase
1-24
1-24

FIG. 7C

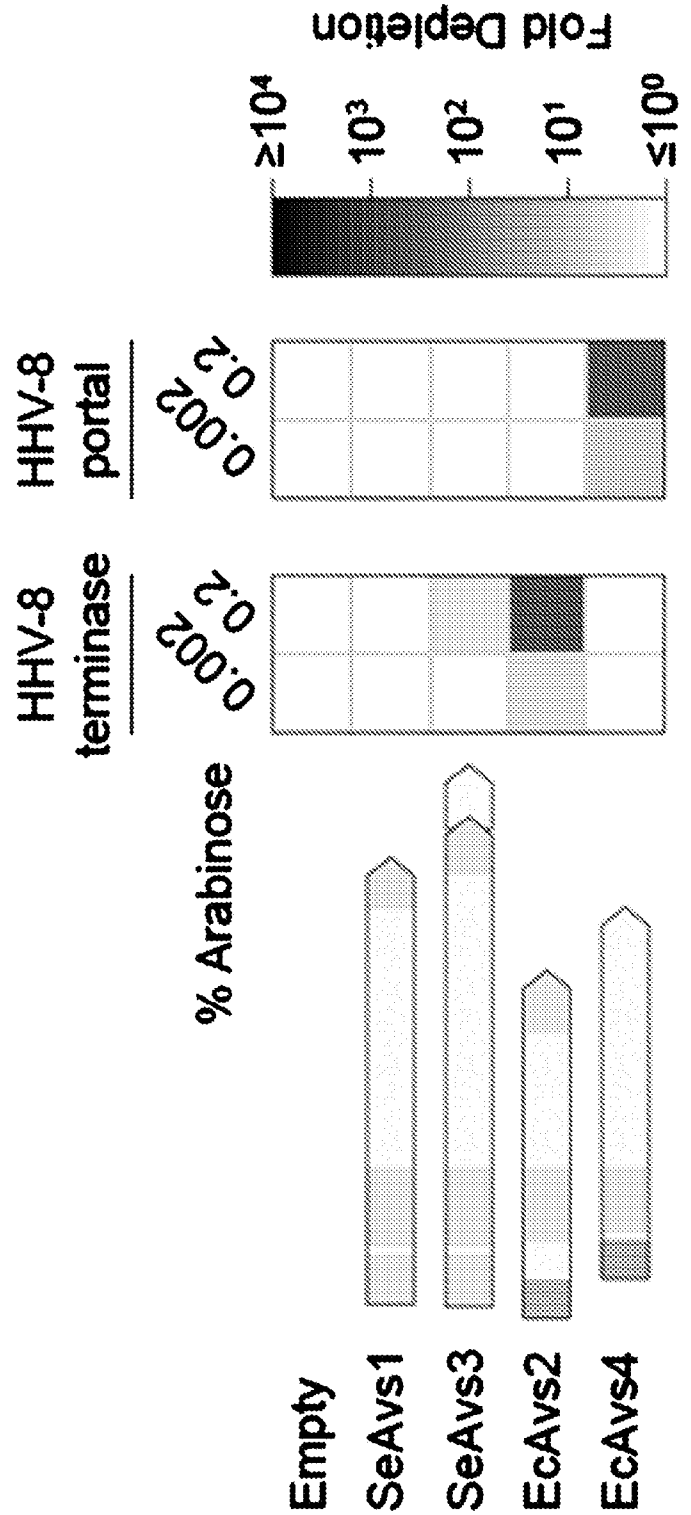

Fold Depletion
≥10^4
10^3
10^2
10^1
≤10^0
HHV-8 portal
HHV-8 terminase
0.2
0.02
0.002
% Arabinose
Empty
SeAvs1
SeAvs3
EcAvs2
EcAvs4

FIG. 7D

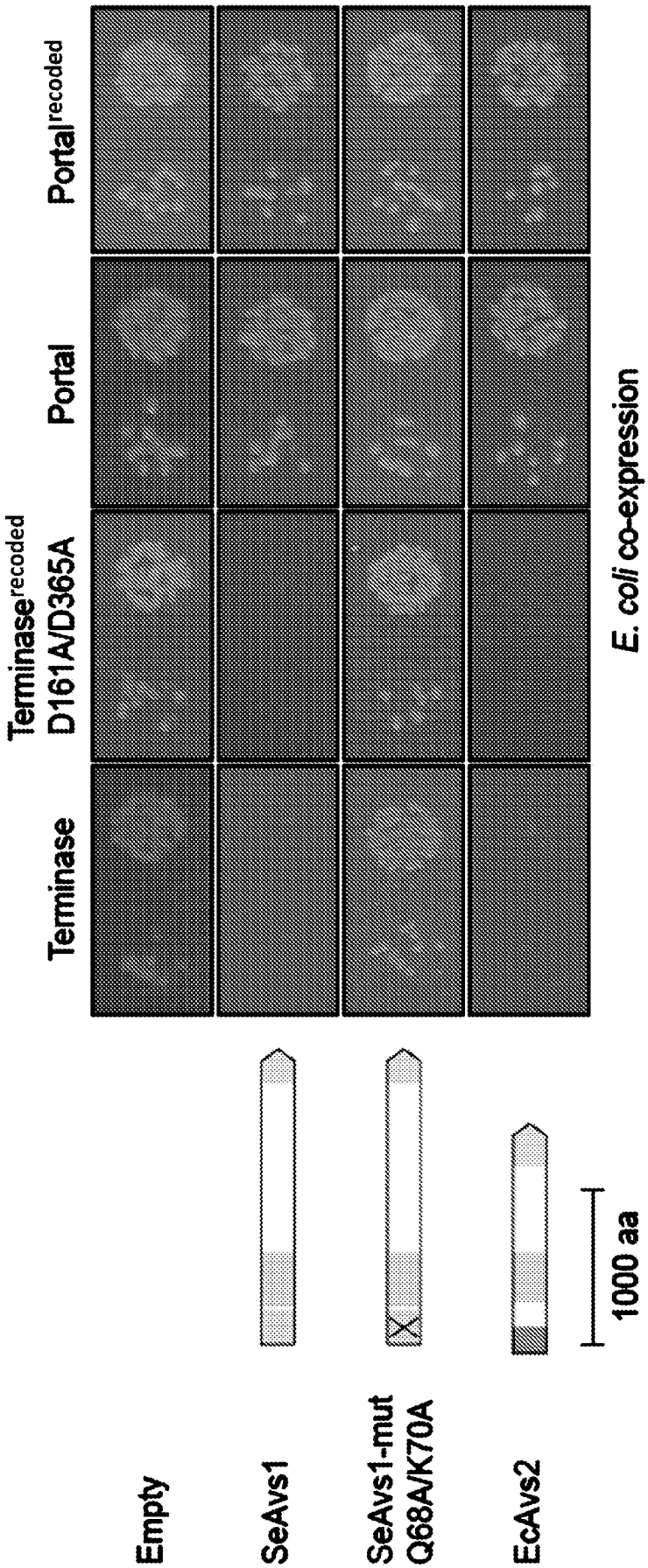

Terminase
Terminase recoded D161A/D365A
Portal
Portal recoded
E. coli co-expression
Empty
SeAvs1
SeAvs1-mut Q68A/K70A
EcAvs2
1000 aa

FIG. 8

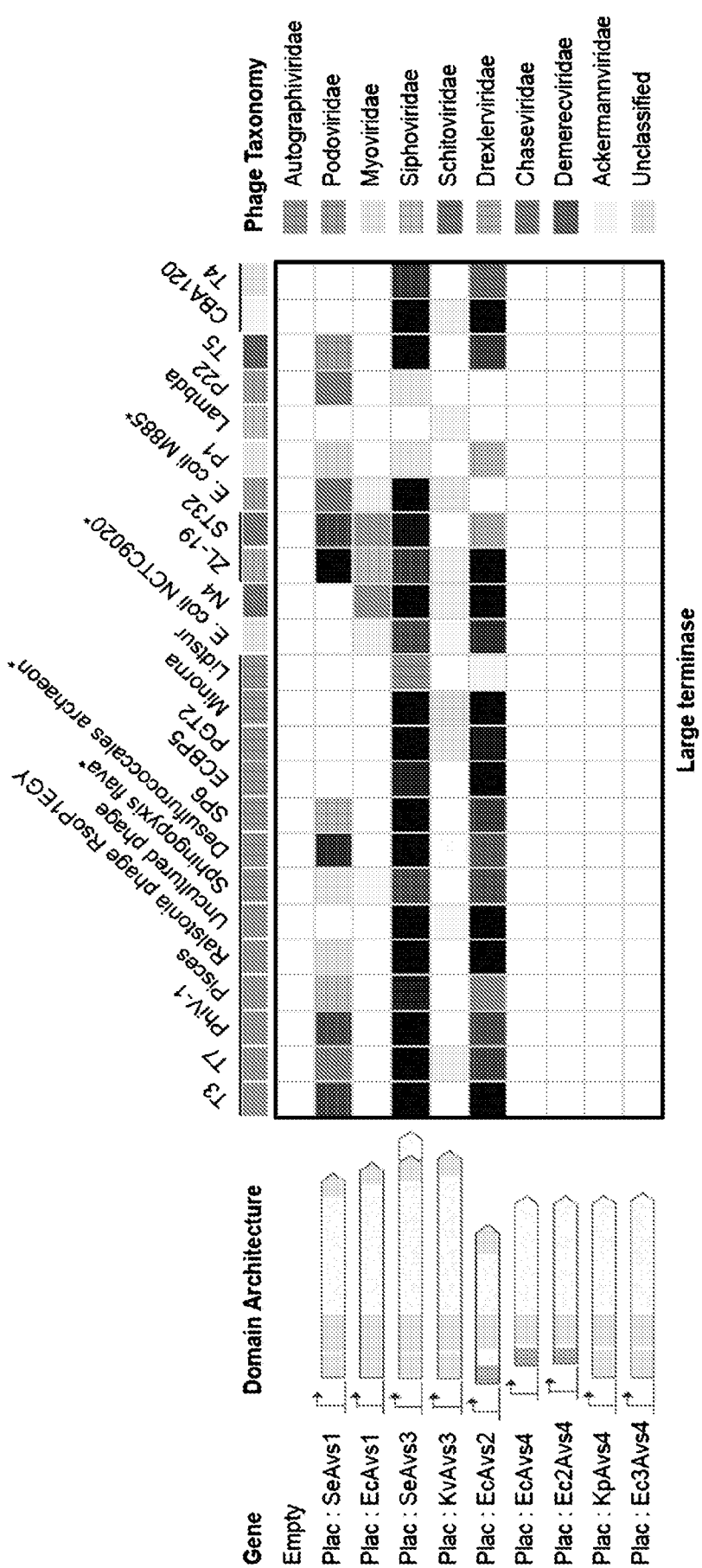

Gene
Domain Architecture
Empty
Plac : SeAvs1
Plac : EcAvs1
Plac : SeAvs3
Plac : KvAvs3
Plac : EcAvs2
Plac : EcAvs4
Plac : Ec2Avs4
Plac : KpAvs4
Plac : Ec3Avs4
T3
T7
phiV-1
Pisces
Ralstonia phage RsoP1EGY
Uncultured phage
Sphingopyxis flava*
Desulfurococcales archaeon*
SP6
ECBP5
PGT2
Minorma
Lidisur
E. coli NCTC9020*
N4
ZL-19
ST32
E. coli M885*
P1
Lambda
P22
T5
CBA120
T4
Phage Taxonomy
Autographiviridae
Podoviridae
Myoviridae
Siphoviridae
Schitoviridae
Drexlerviridae
Chaseviridae
Demerecviridae
Ackermannviridae
Unclassified
Large terminase

FIG. 9A

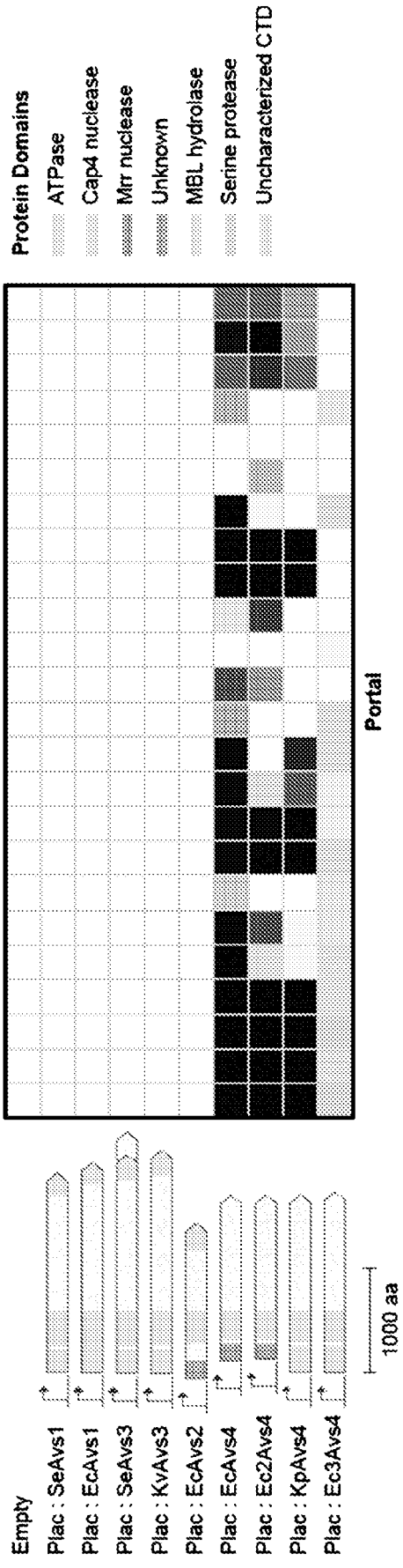


FIG. 9A Continued

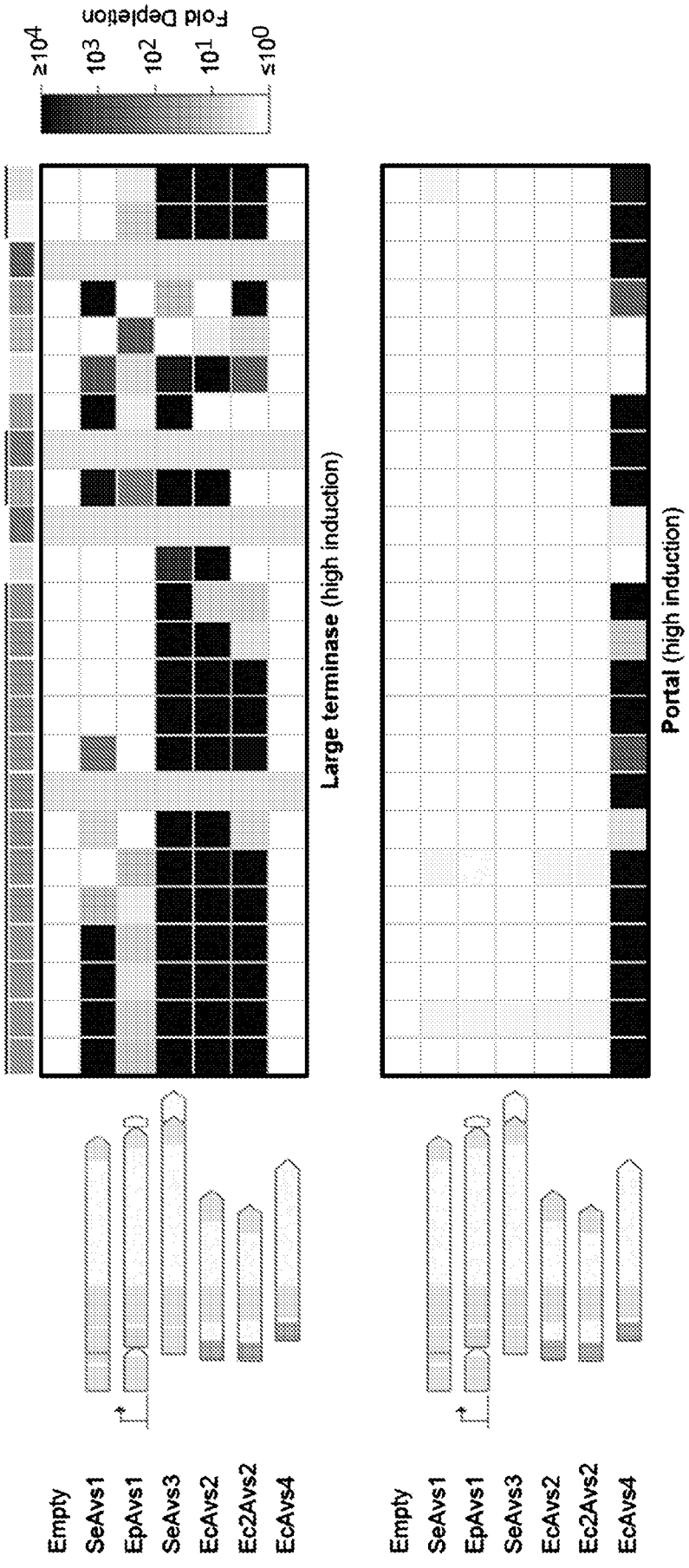

Fold Depletion
≥10^4
10^3
10^2
10^1
≤10^0
Large terminase (high induction)
Portal (high induction)
Empty
SeAvs1
EpAvs1
SeAvs3
EcAvs2
Ec2Avs2
EcAvs4

FIG. 9B

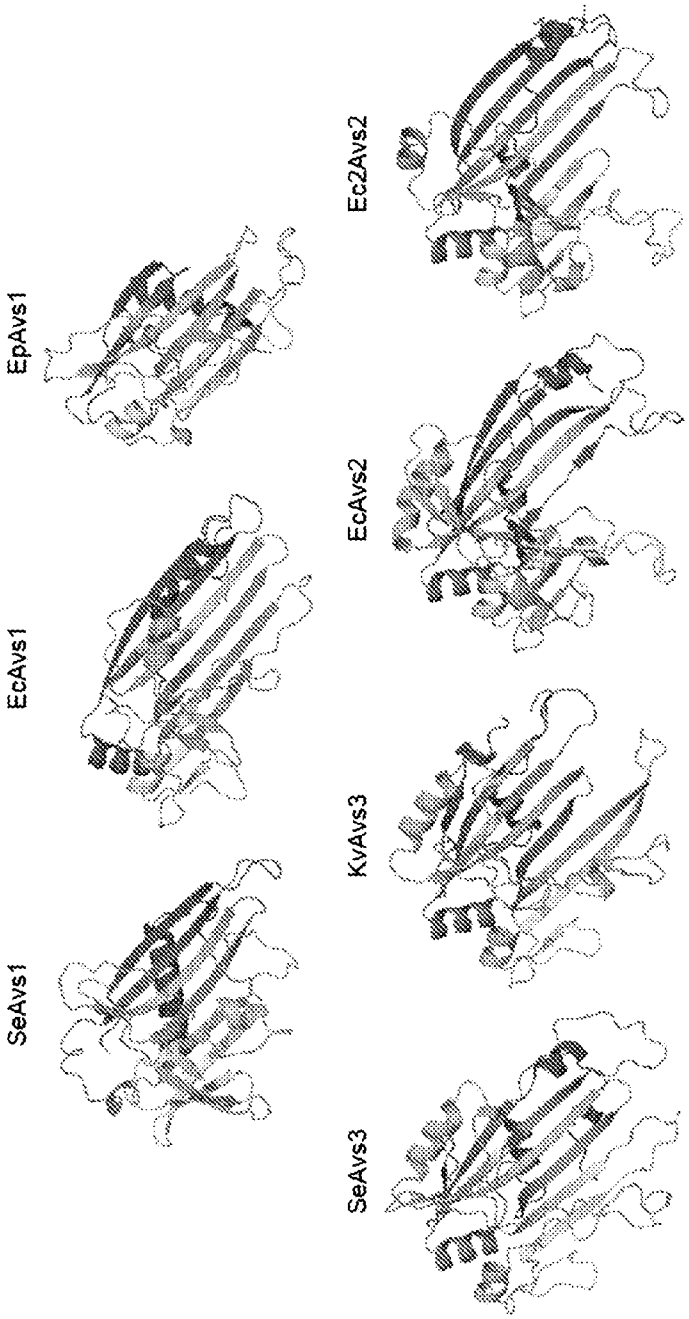


FIG. 10A

FIG. 10B

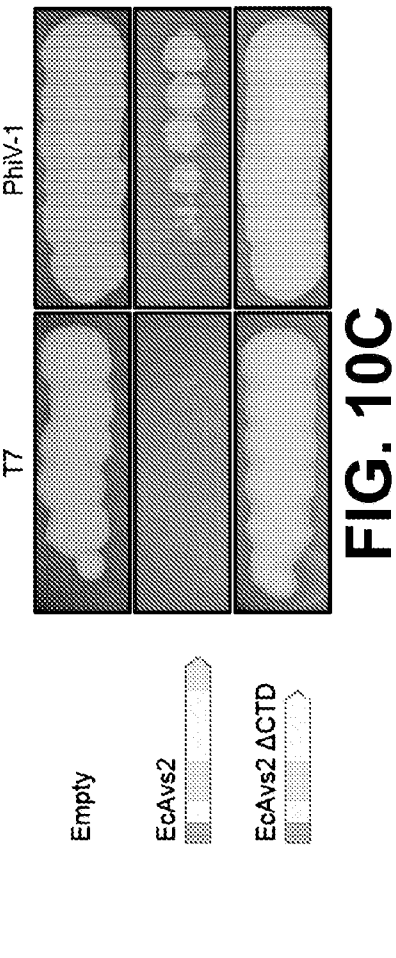


FIG. 10C

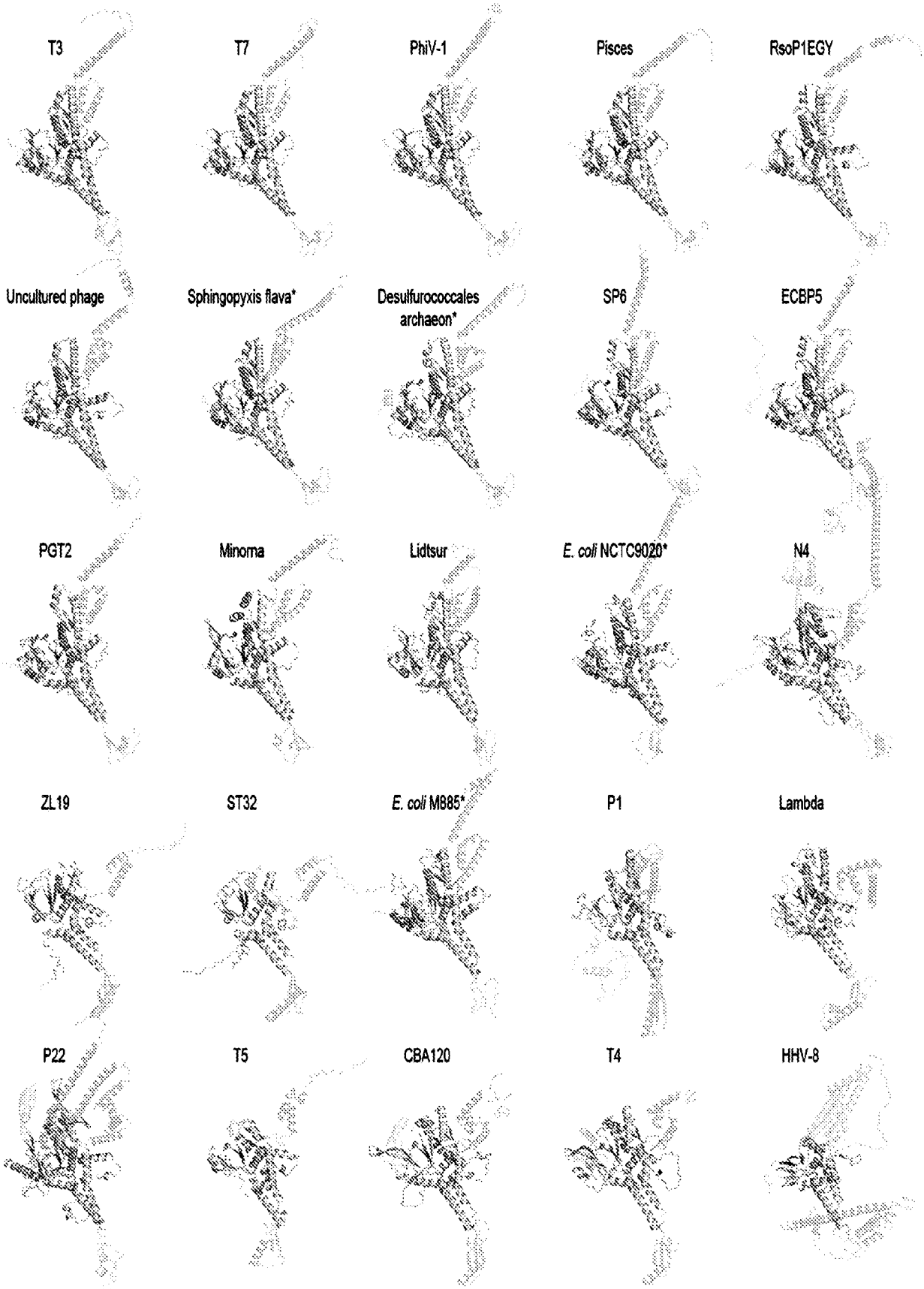

T3
T7
PhiV-1
Pisces
RsoP1EGY
Uncultured phage
Sphingopyxis flava*
Desulfurococcales archaeon*
SP6
ECBP5
PGT2
Minorna
Lidtsur
E. coli NCTC9020*
N4
ZL19
ST32
E. coli M885*
P1
Lambda
P22
T5
CBA120
T4
HHV-8

FIG. 11

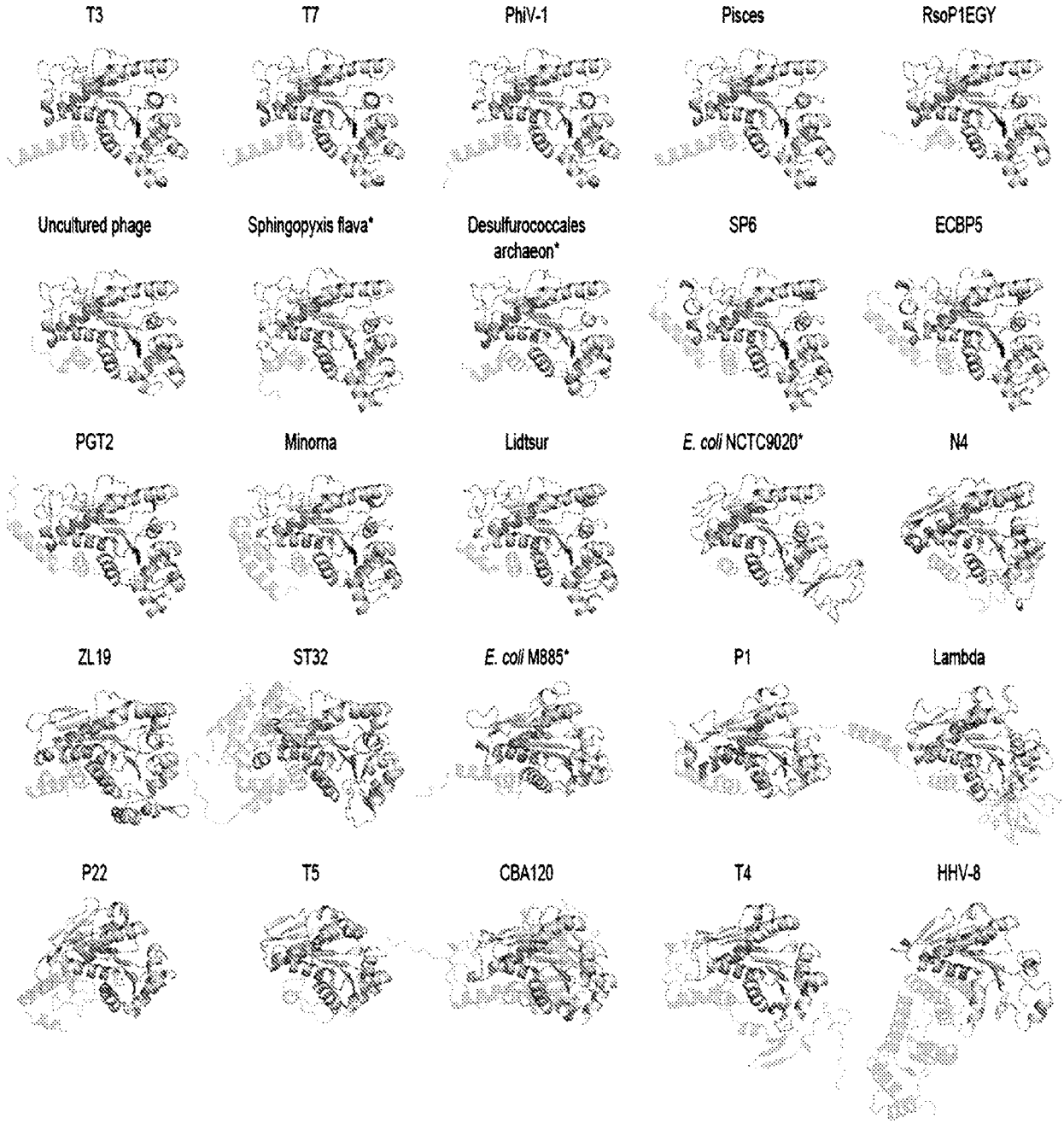

T3
T7
PhiV-1
Pisces
RsoP1EGY
Uncultured phage
Sphingopyxis flava*
Desulfurococcales archaeon*
SP6
ECBP5
PGT2
Minorna
Lidtsur
E. coli NCTC9020*
N4
ZL19
ST32
E. coli M885*
P1
Lambda
P22
T5
CBA120
T4
HHV-8

FIG. 12

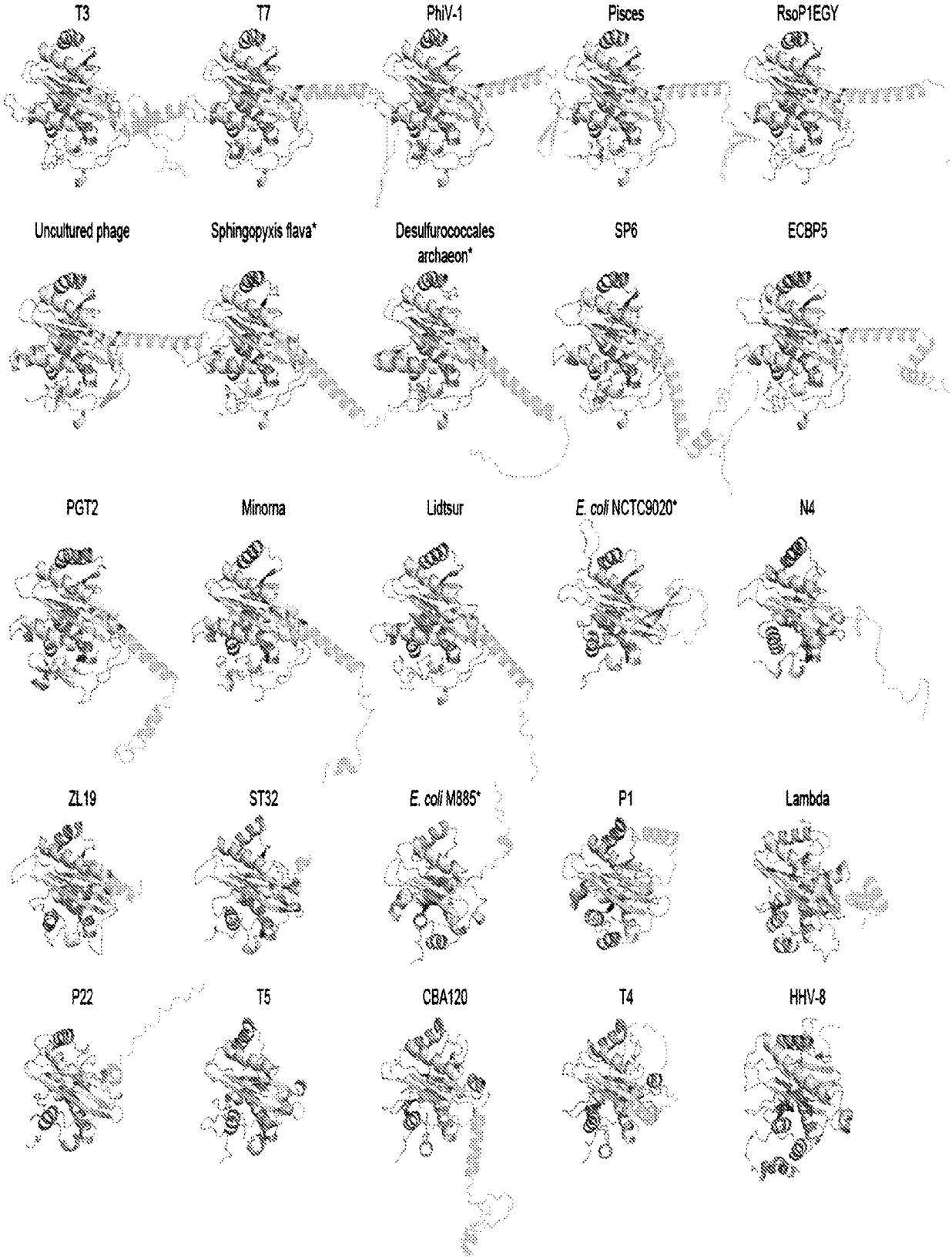

T3
T7
PhiV-1
Pisces
RsoP1EGY
Uncultured phage
Sphingopyxis flava*
Desulfurococcales archaeon*
SP6
ECBP5
PGT2
Minorna
Lidtsur
E. coli NCTC9020*
N4
ZL19
ST32
E. coli M885*
P1
Lambda
P22
T5
CBA120
T4
HHV-8

FIG. 13

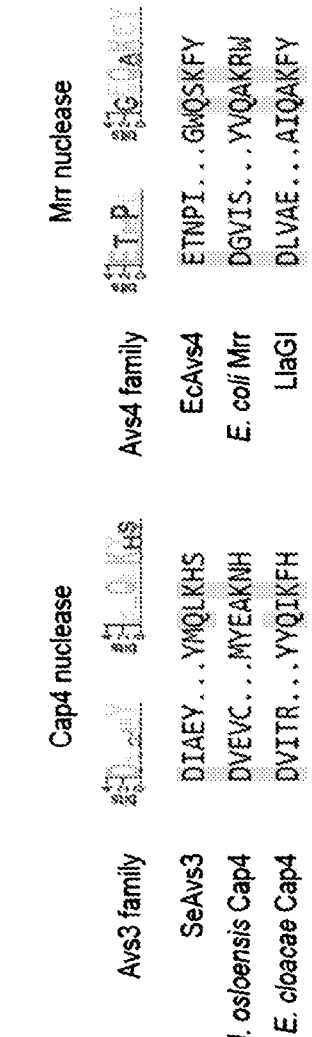


FIG. 14B

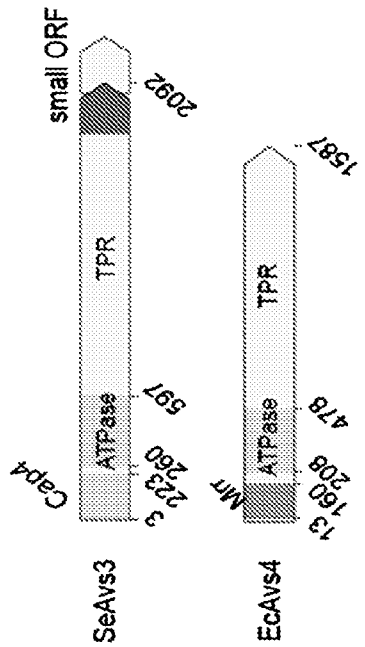


FIG. 14A

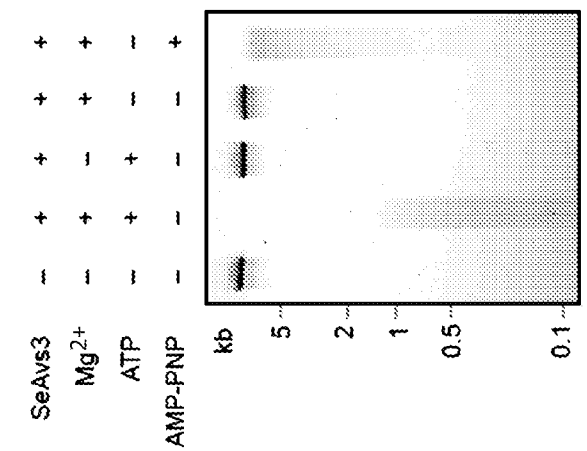


FIG. 14E

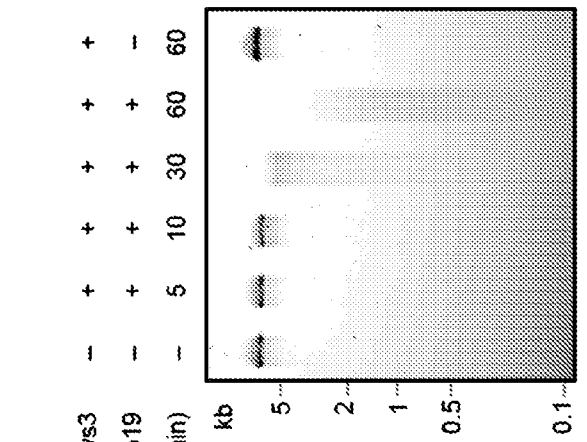


FIG. 14D

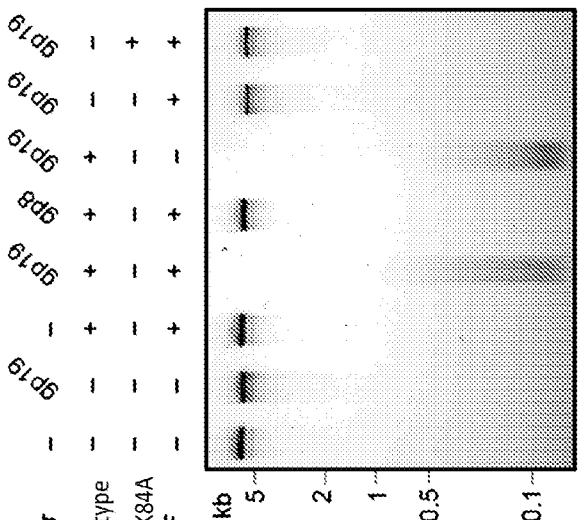


FIG. 14C

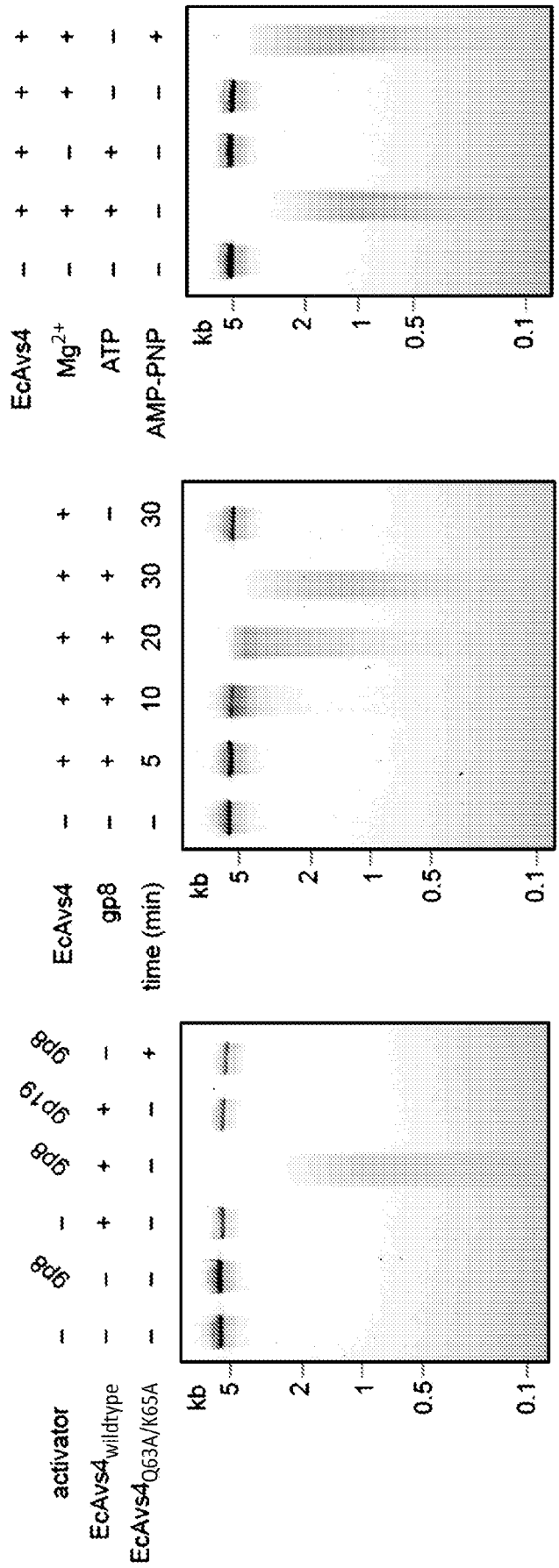


FIG. 14F

FIG. 14G

FIG. 14H

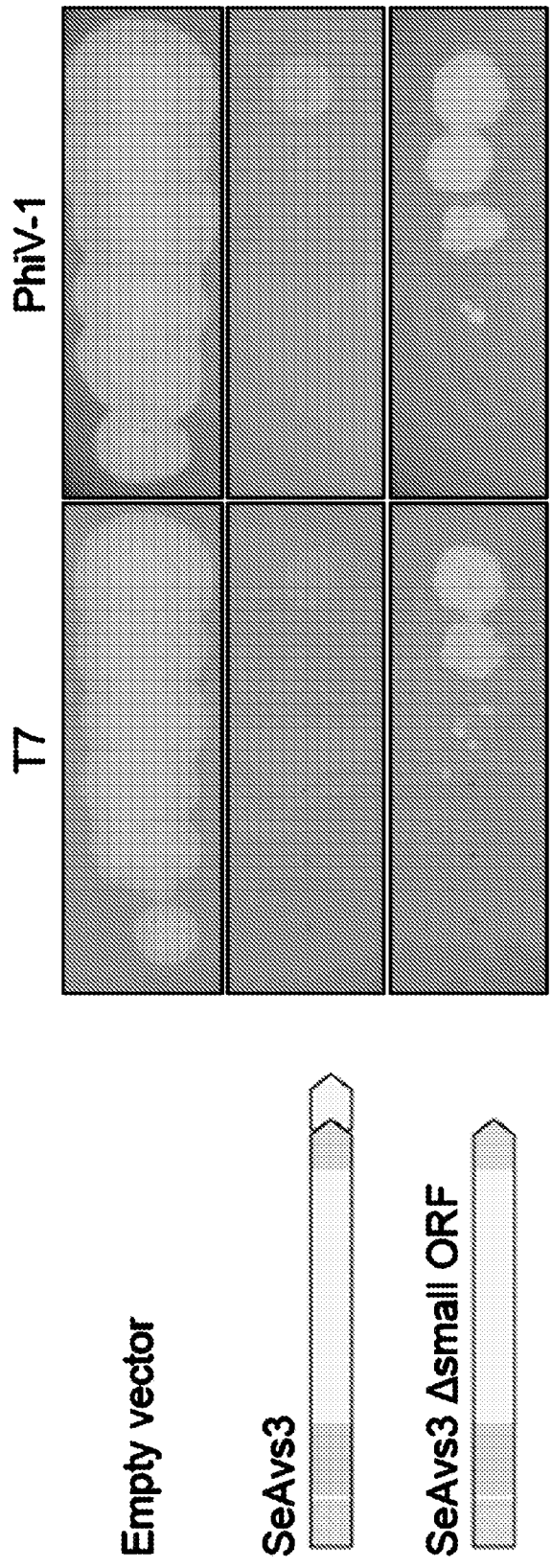


FIG. 15A

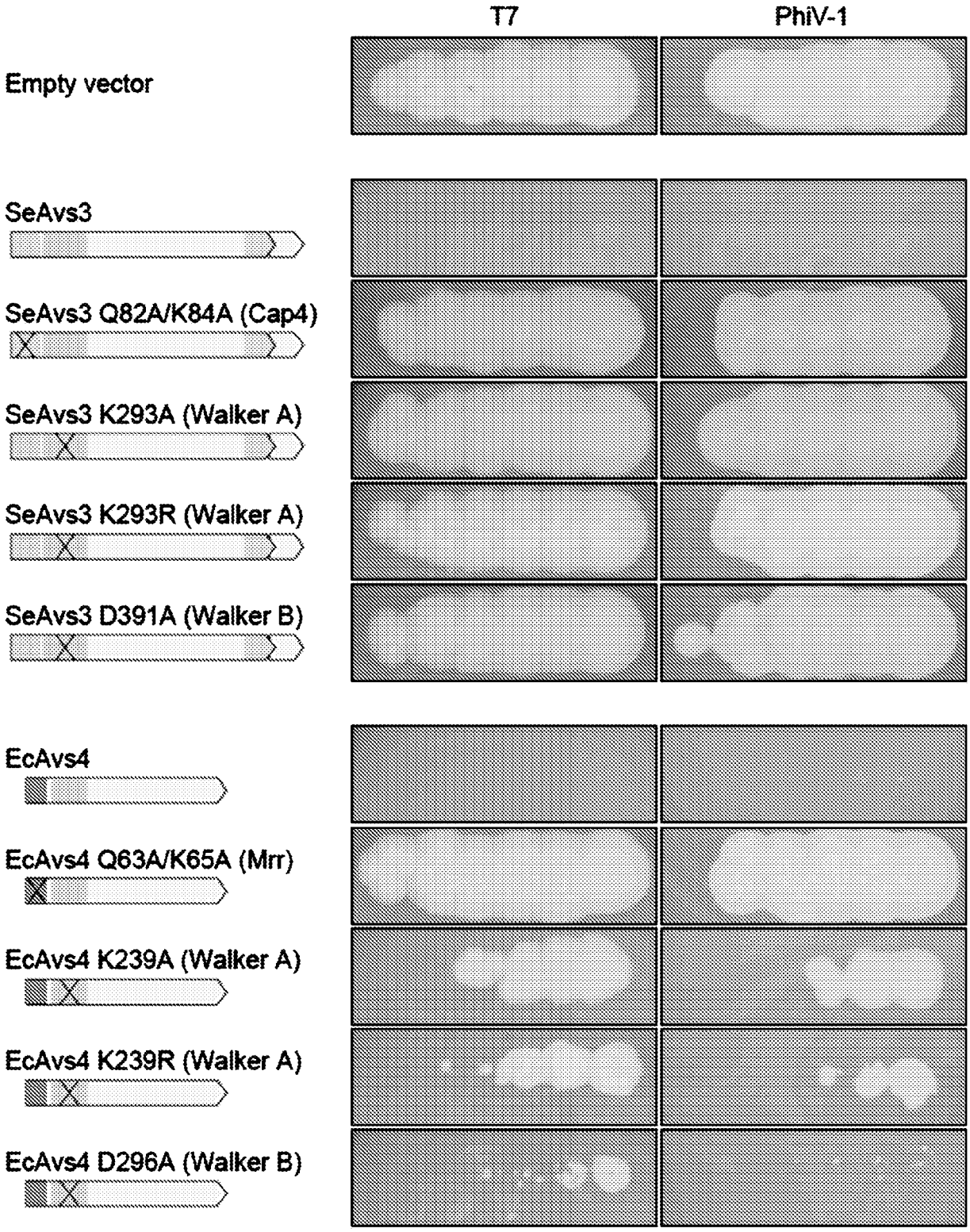

T7
PhiV-1
Empty vector
SeAvs3
SeAvs3 Q82A/K84A (Cap4)
SeAvs3 K293A (Walker A)
SeAvs3 K293R (Walker A)
SeAvs3 D391A (Walker B)
EcAvs4
EcAvs4 Q63A/K65A (Mrr)
EcAvs4 K239A (Walker A)
EcAvs4 K239R (Walker A)
EcAvs4 D296A (Walker B)

FIG. 15B

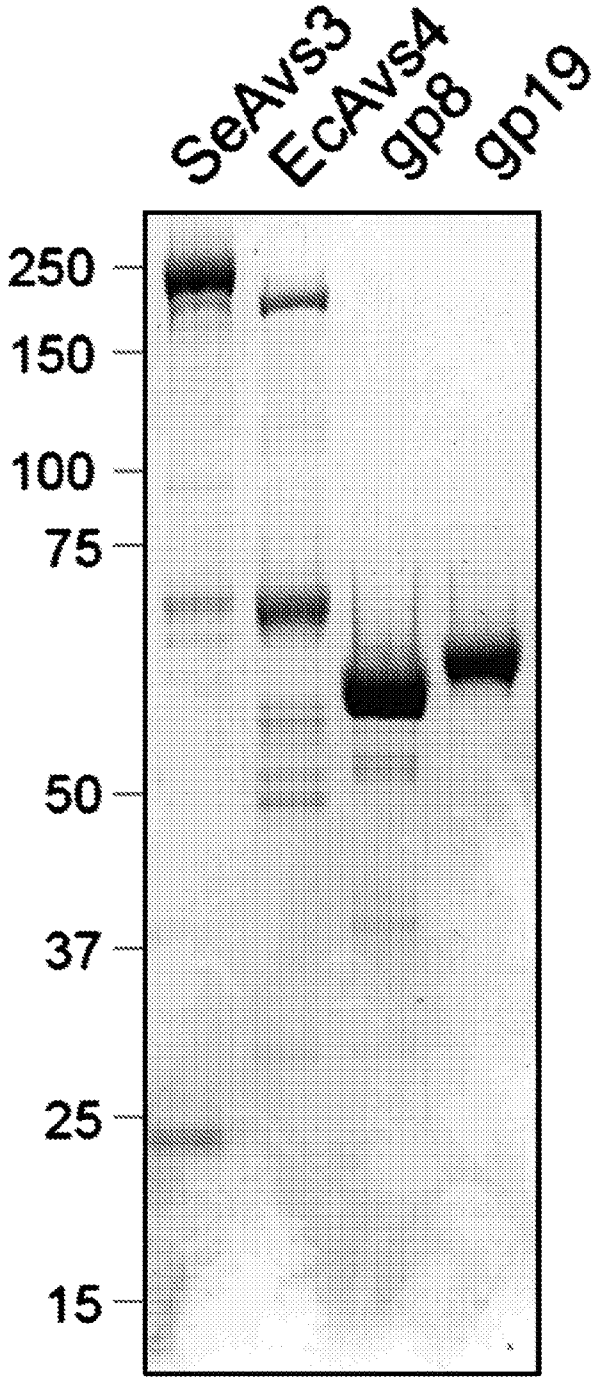


FIG. 16A

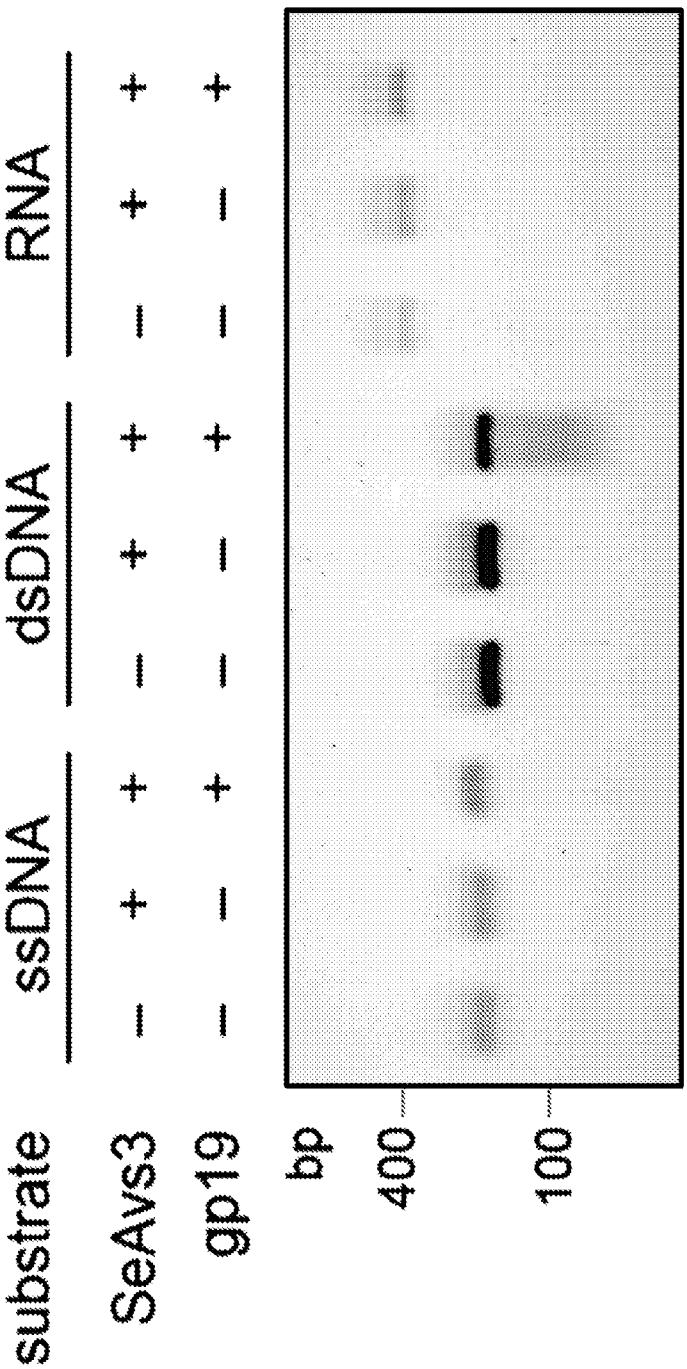


FIG. 16B

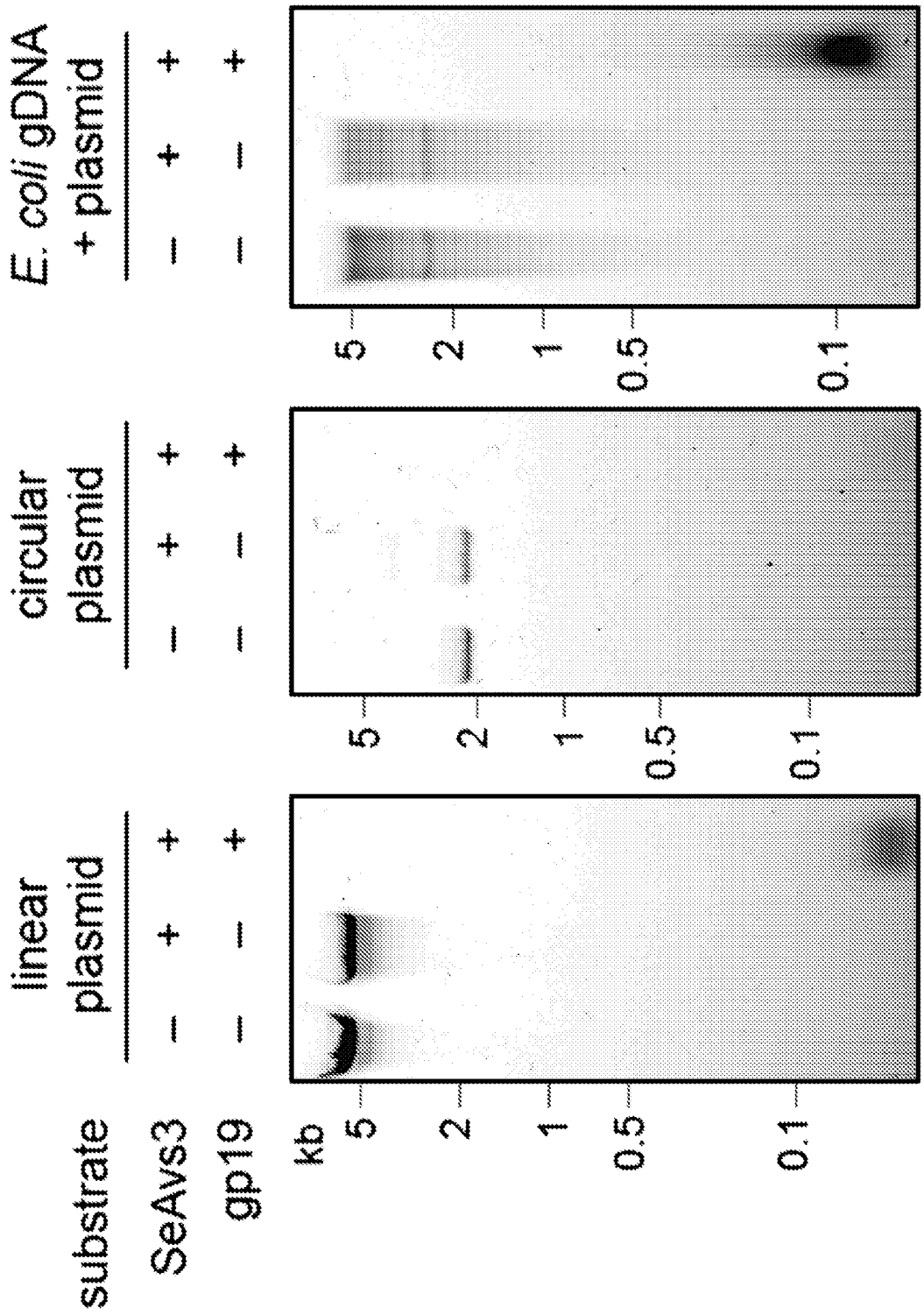

substrate
linear plasmid
circular plasmid
E. coli gDNA + plasmid
SeAvs3
gp19
kb
5
2
1
0.5
0.1
+
—

FIG. 16C

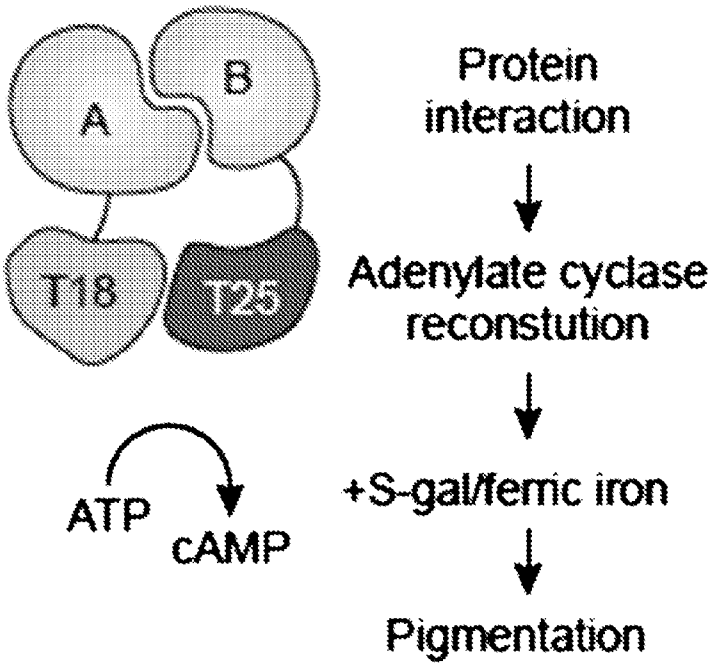


FIG. 17A

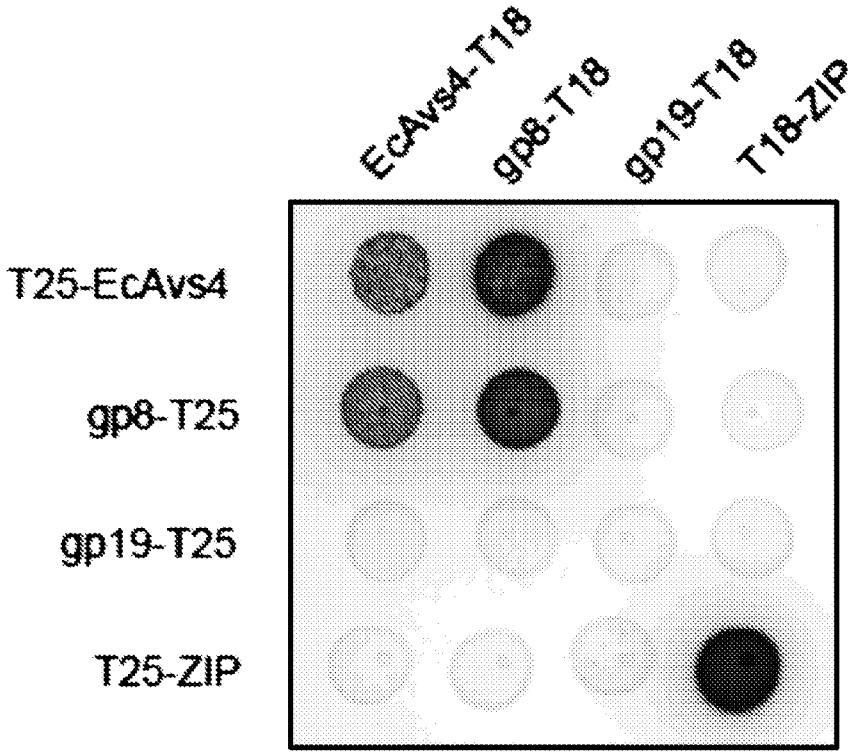


FIG. 17B

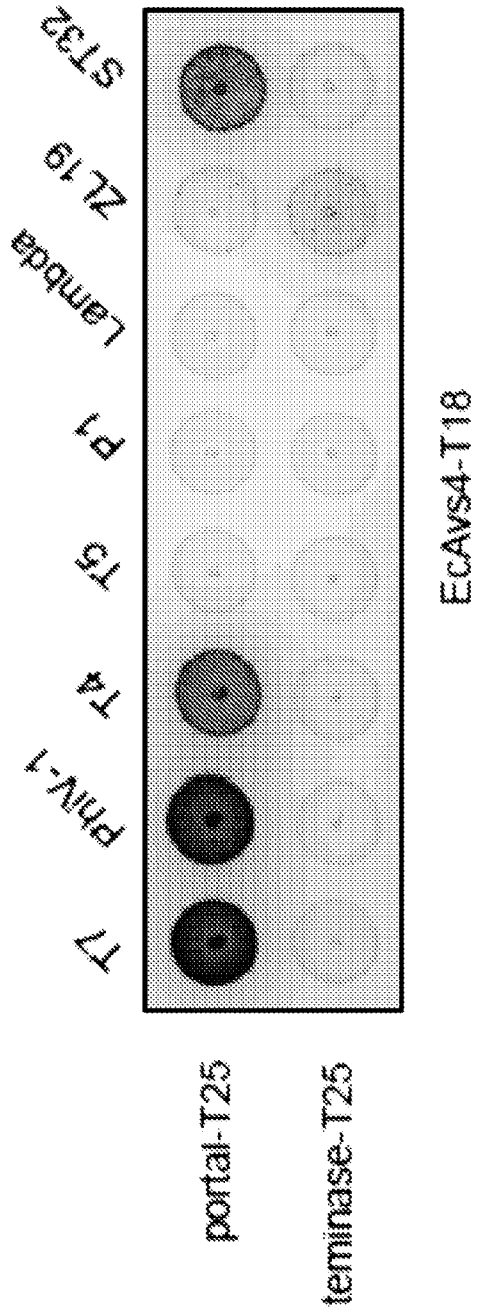

T7
pHK-1
T4
T5
P1
Lambda
ZL 19
ST32
portal-T25
teminase-T25
EcAvs4-T18

FIG. 17C

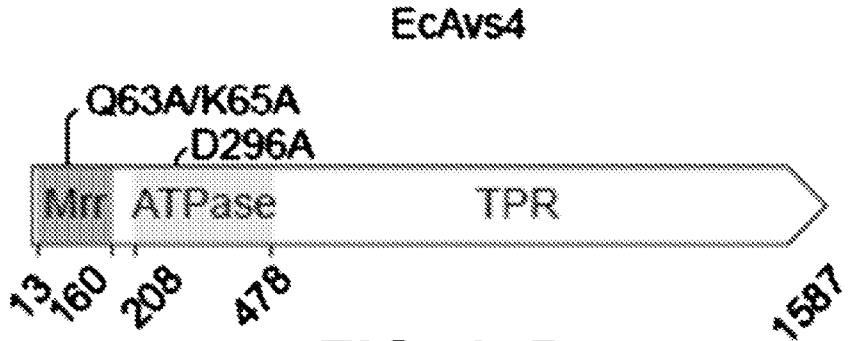


FIG. 17D

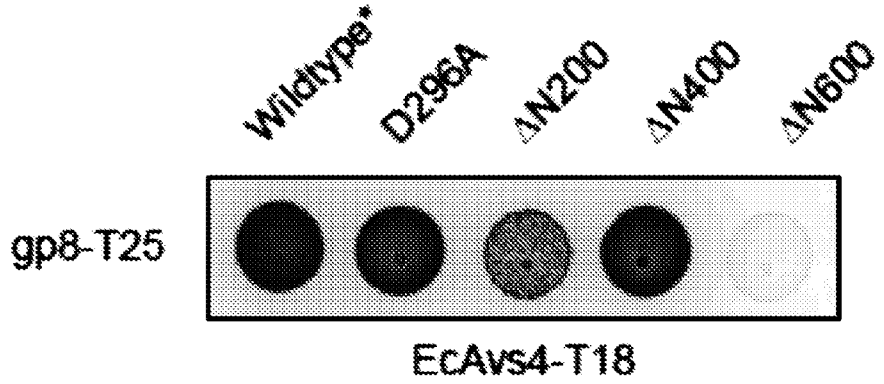


FIG. 17E

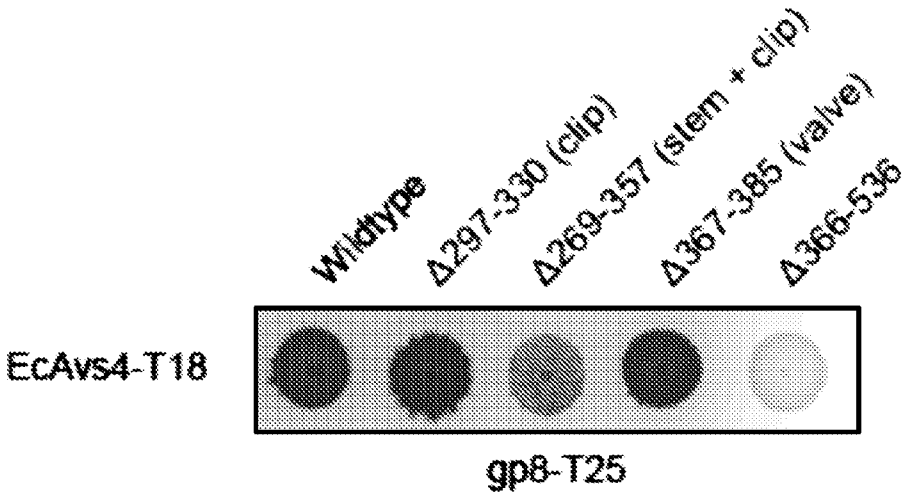


FIG. 17F

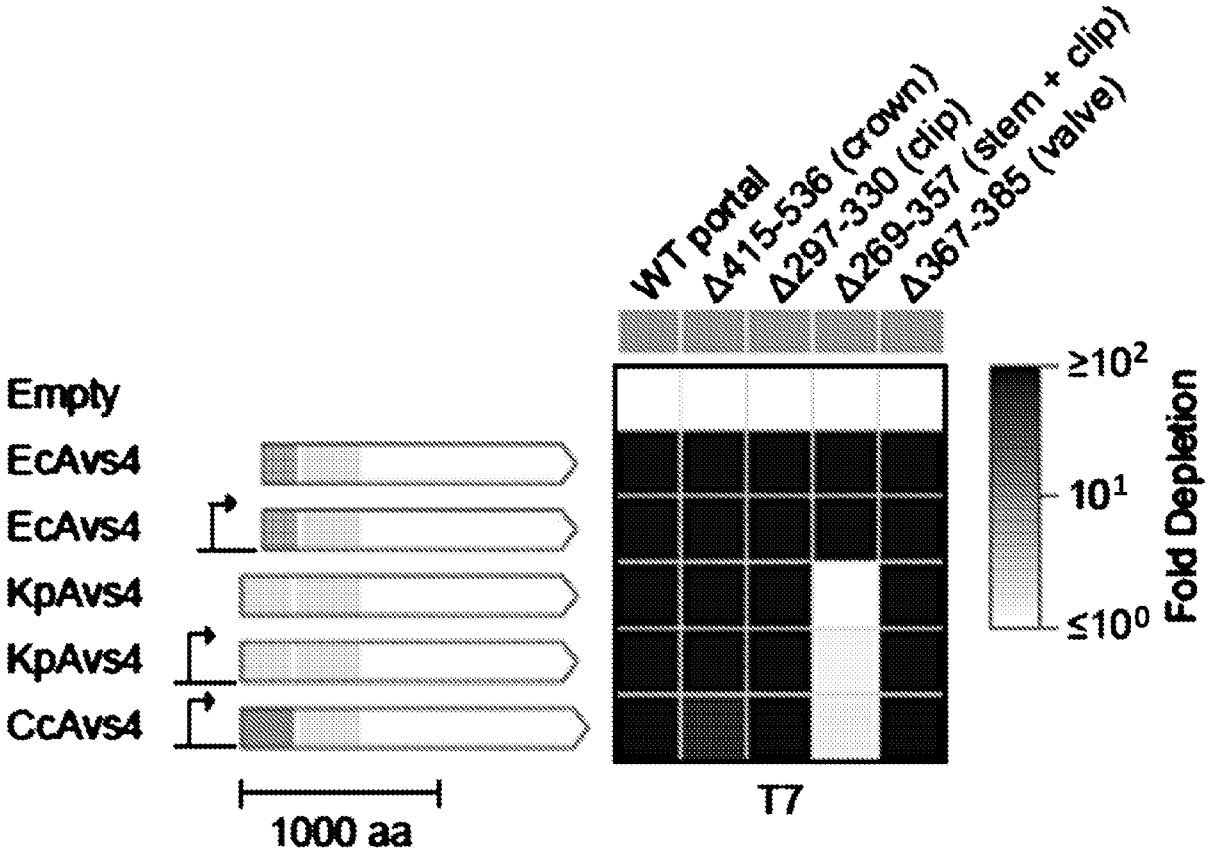

WT portal
Δ415-536 (crown)
Δ297-330 (clip)
Δ269-357 (stem + clip)
Δ367-385 (valve)
Empty
EcAvs4
EcAvs4
KpAvs4
KpAvs4
CcAvs4
1000 aa
T7
≥10²
10¹
≤10⁰
Fold Depletion

FIG. 17G

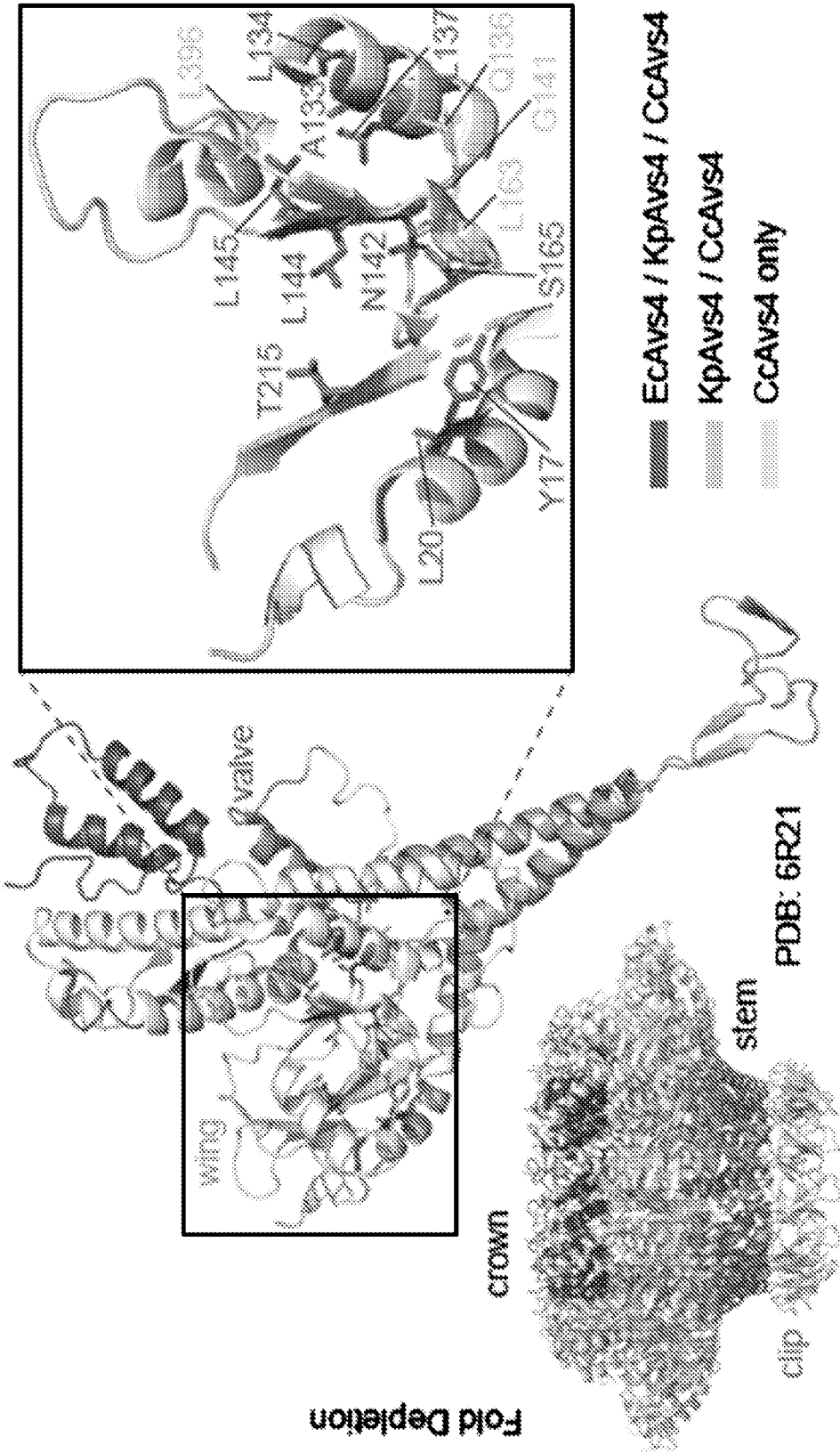

EcAvs4 / KpAvs4 / CcAvs4
KpAvs4 / CcAvs4
CcAvs4 only
L136
L134
A133
I137
L145
L144
N142
D138
Q136
Q141
T215
S165
Y17
L20
wing
naive
crown
stem
clip
PDB: 6R21
Fold Depletion

FIG. 17H

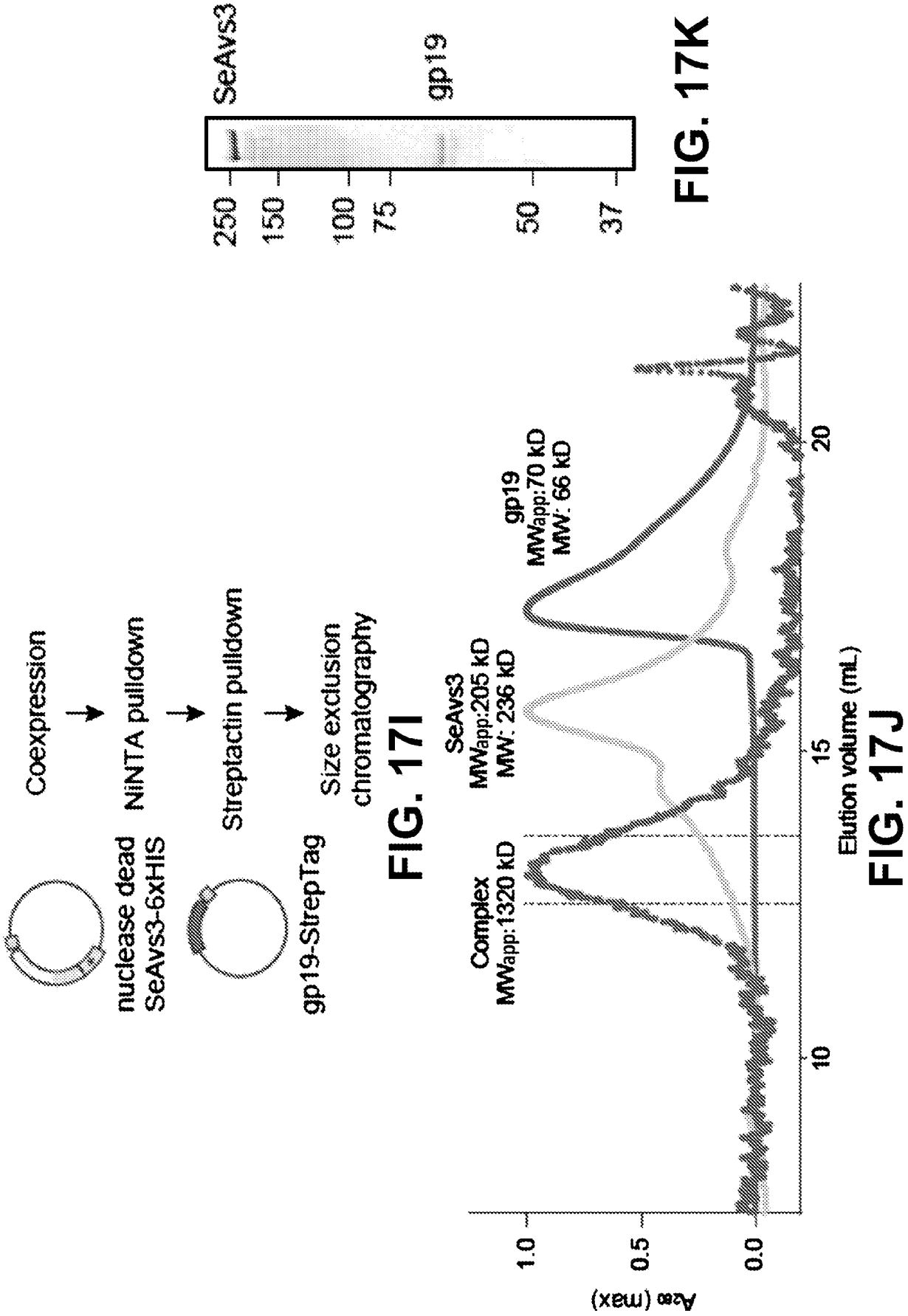


FIG. 17I

FIG. 17J

FIG. 17K

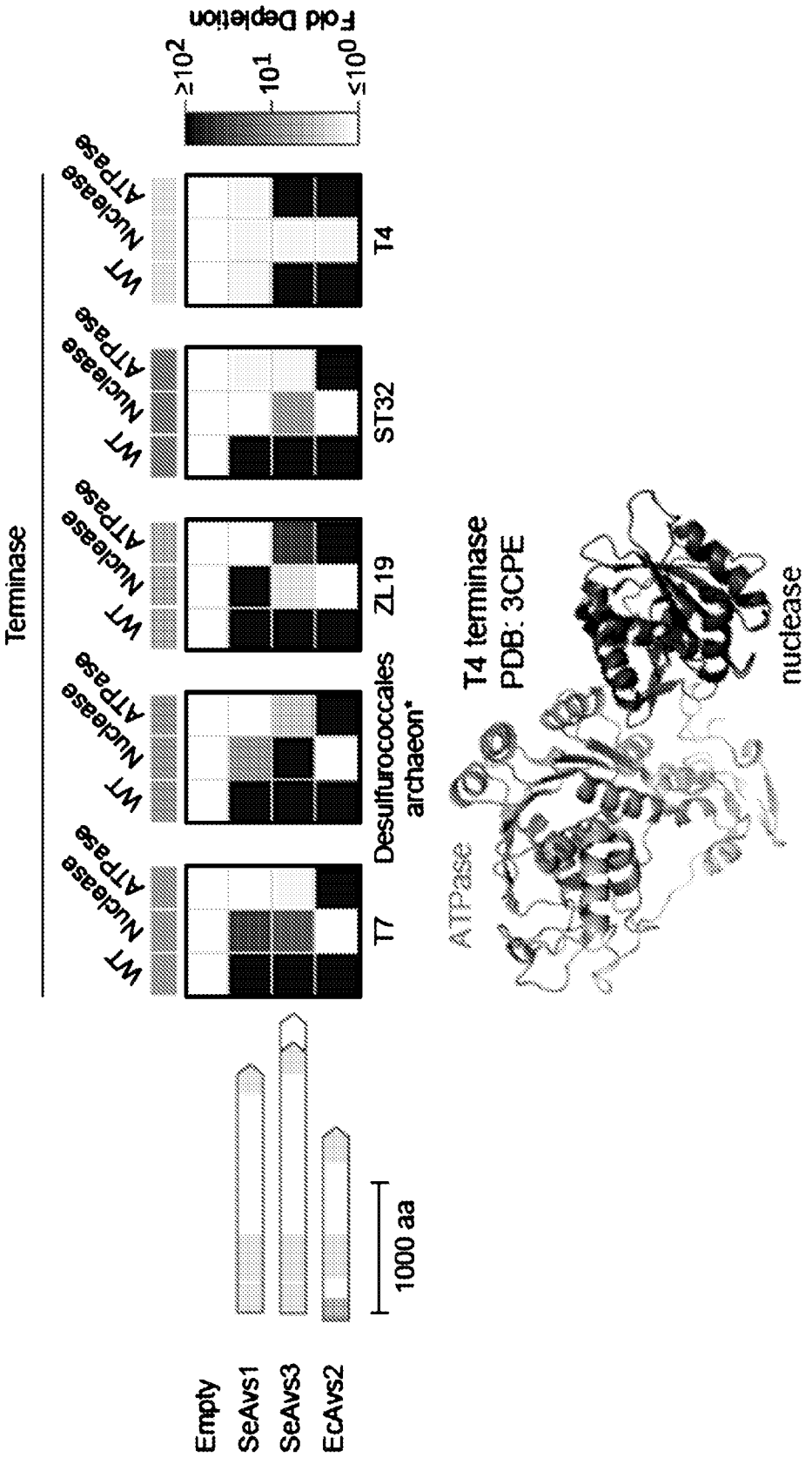

Fold Depletion
≥10^2
10^1
≤10^0
Terminase
WT
Nuclease
ATPase
T7
Desulfurococcales archaeon*
ZL19
ST32
T4
T4 terminase
PDB: 3CPE
ATPase
nuclease
1000 aa
Empty
SeAvs1
SeAvs3
EcAvs2

FIG. 17L

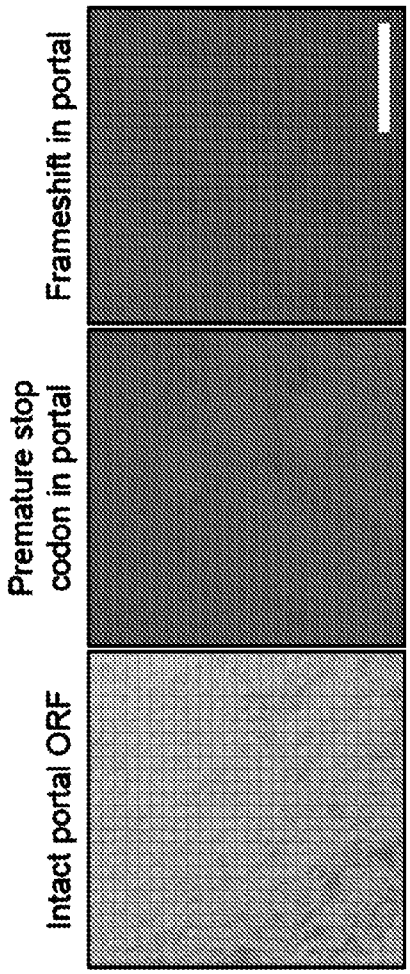


FIG. 18B

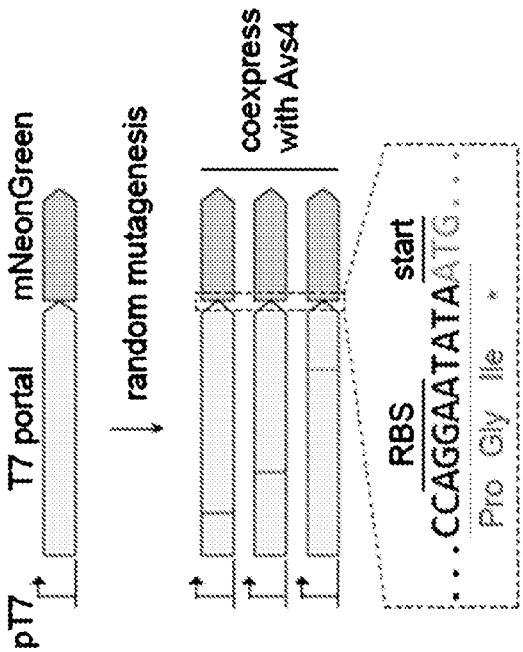


FIG. 18A

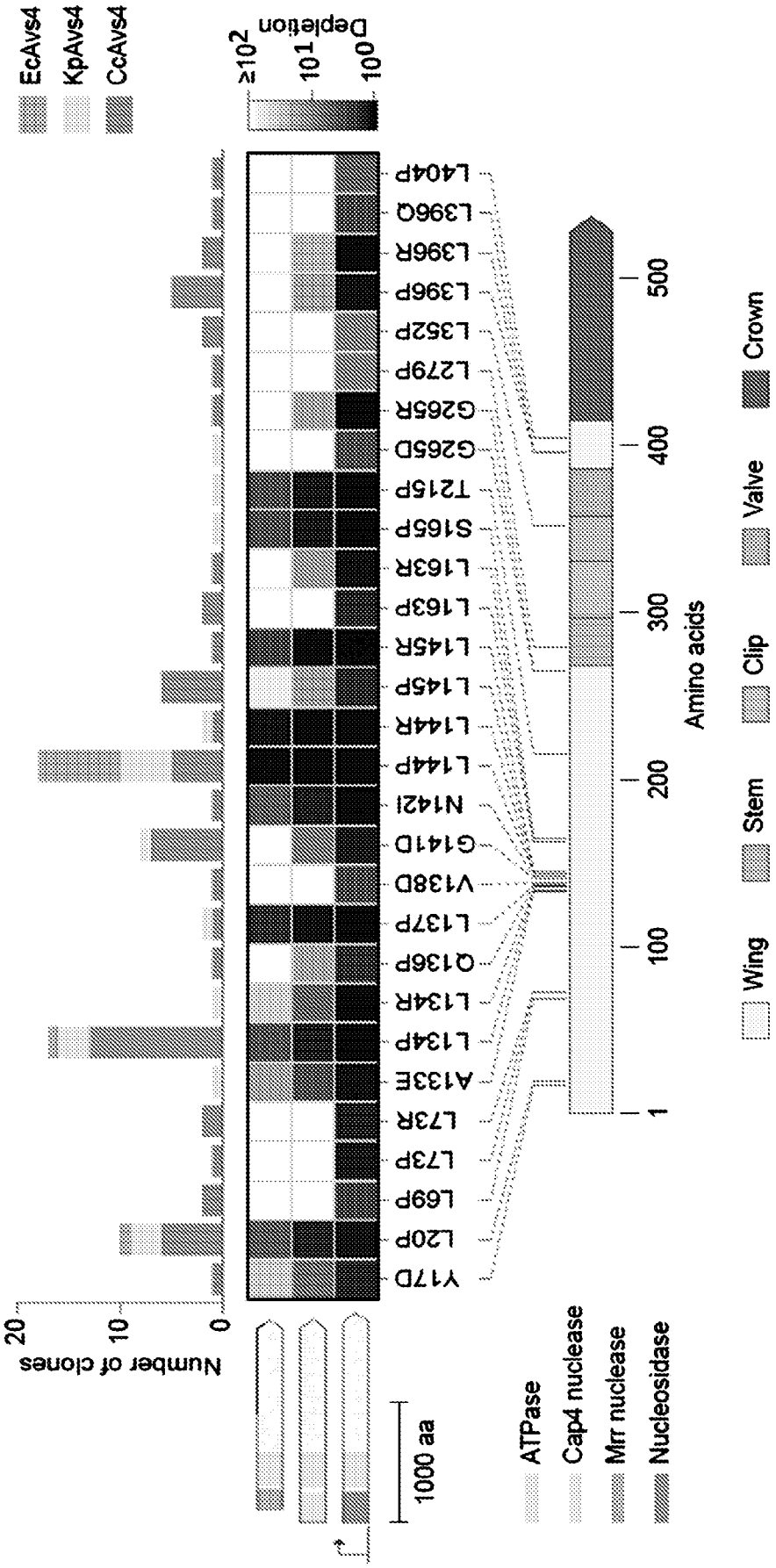

EcAvs4
KpAvs4
CcAvs4
Depletion
≥10^2
10^1
10^0
Y17D
L20P
L69P
L73P
L73R
A133E
L134P
L134R
Q136P
L137P
V138D
G141D
N142I
L144P
L144R
L145P
L145R
L163P
L163R
S165P
T215P
G265D
G265R
L279P
L352P
L396P
L396R
L396Q
L404P
Number of clones
20
10
0
1000 aa
ATPase
Cap4 nuclease
Mrr nuclease
Nucleosidase
Amino acids
1
100
200
300
400
500
Wing
Stem
Clip
Valve
Crown

FIG. 18C

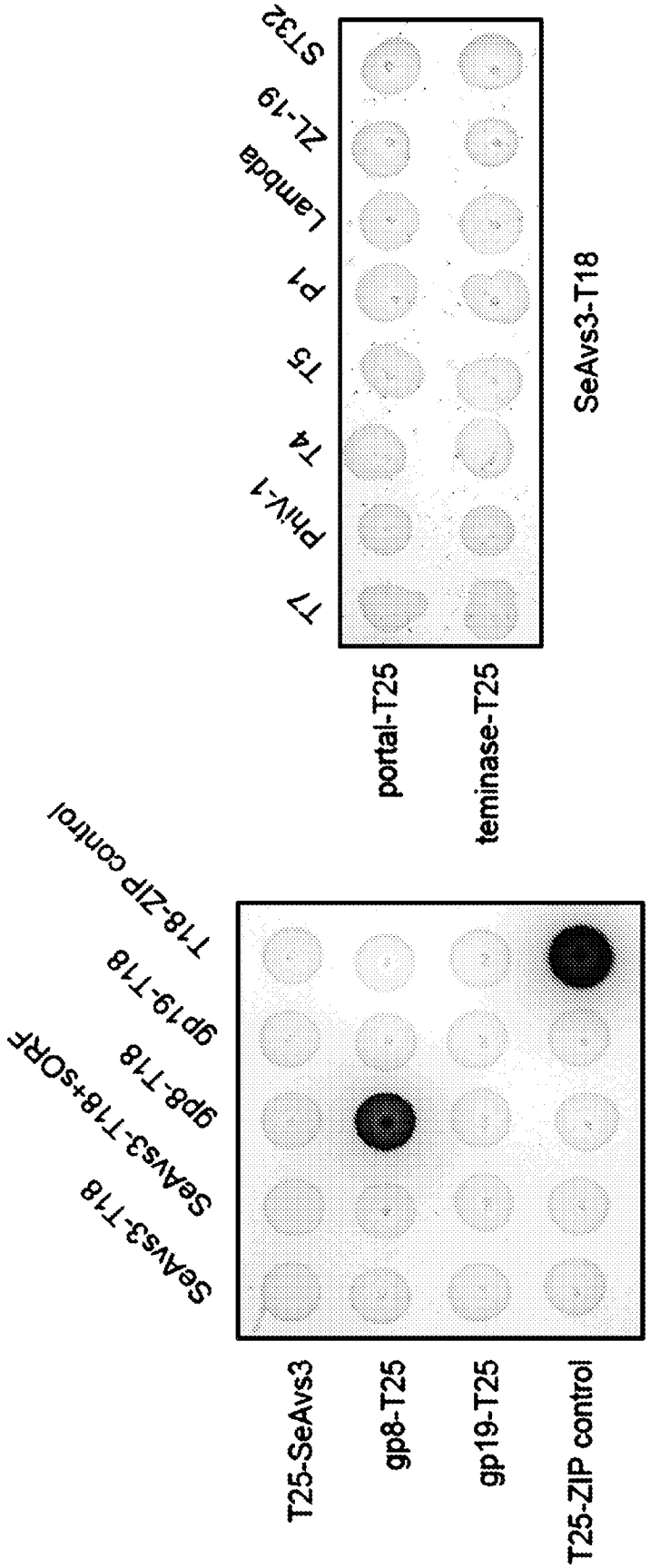
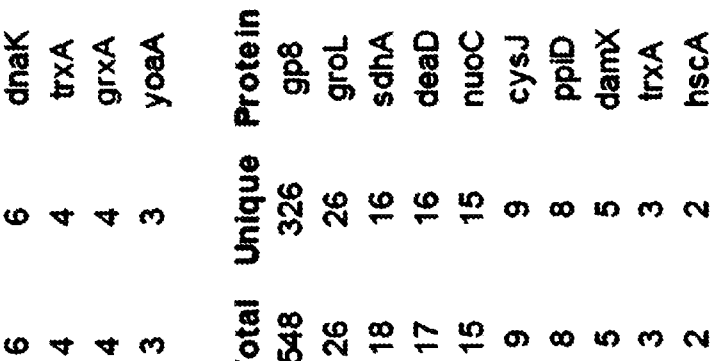


FIG. 19A

FIG. 19B



Total	Unique	Protein
71	42	gp19
25	24	rpsA
19	19	ppiD
18	18	deaD
6	6	groL
6	6	dnaK
4	4	trxA
4	4	grxA
3	3	yoaA

FIG. 19E

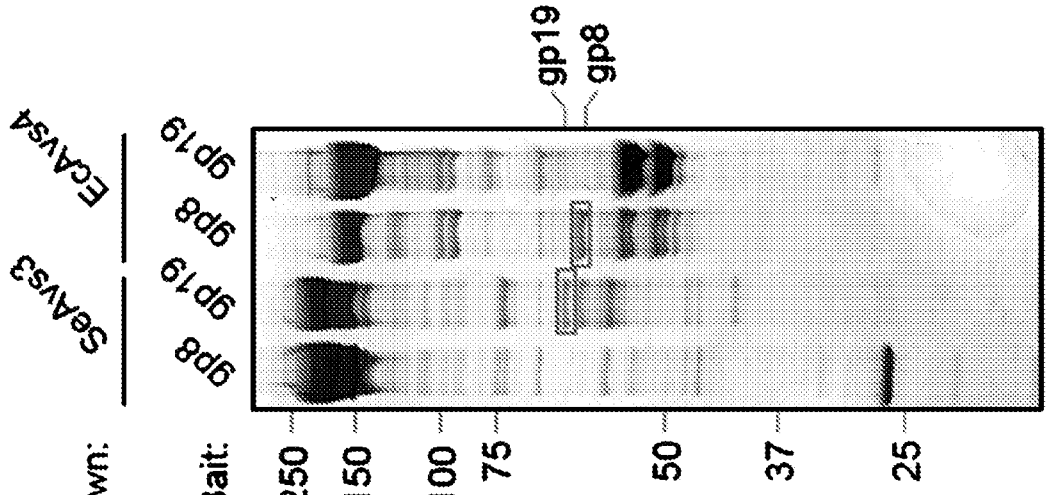


FIG. 19D

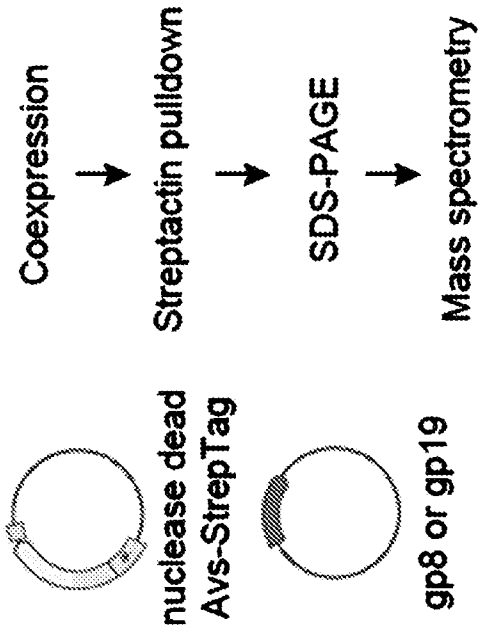


FIG. 19C

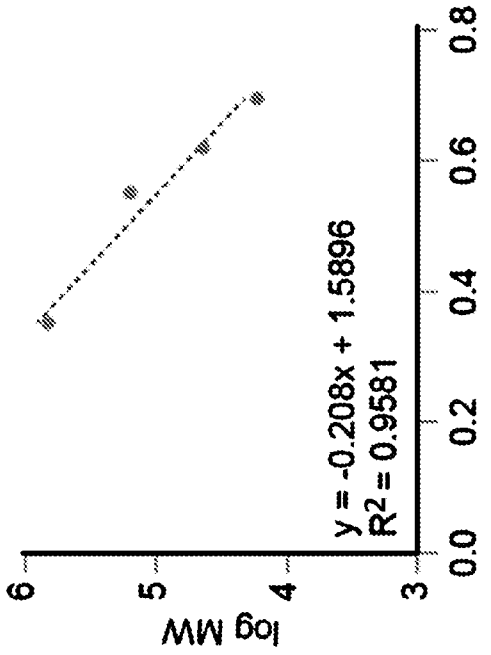


FIG. 19G

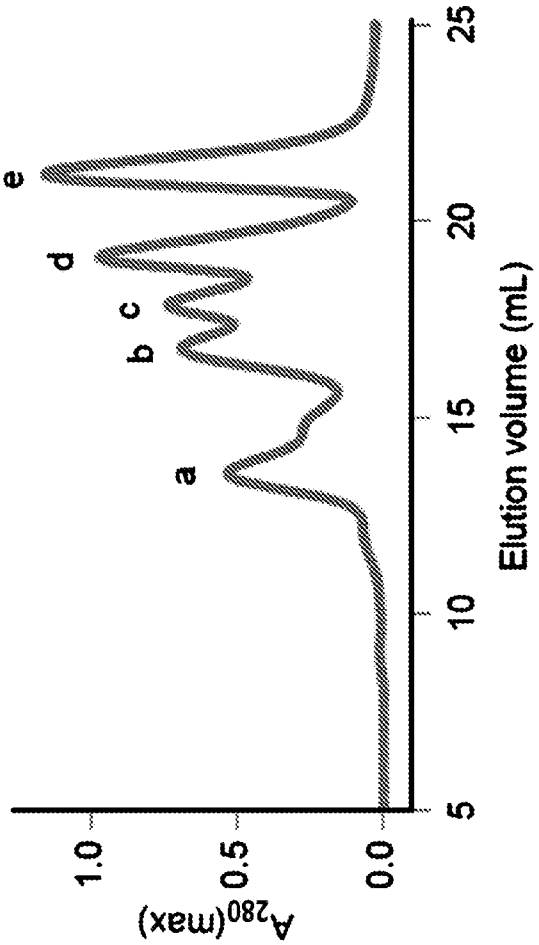


FIG. 19F

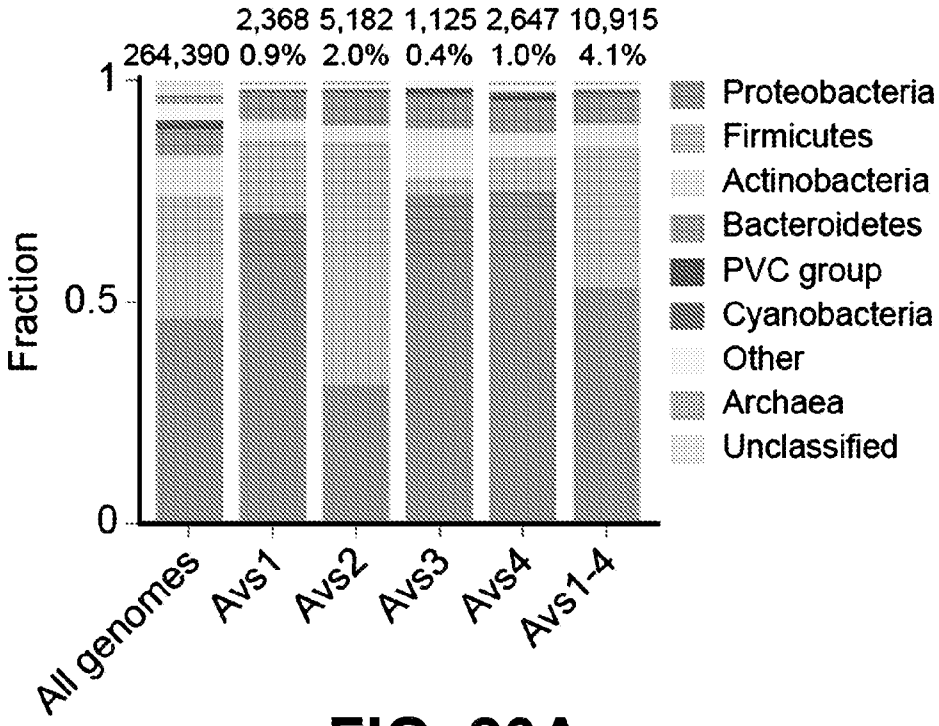


FIG. 20A

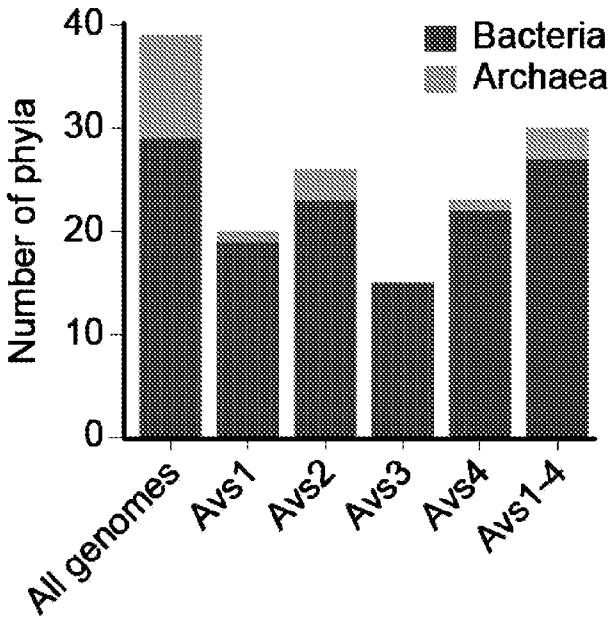


FIG. 20B

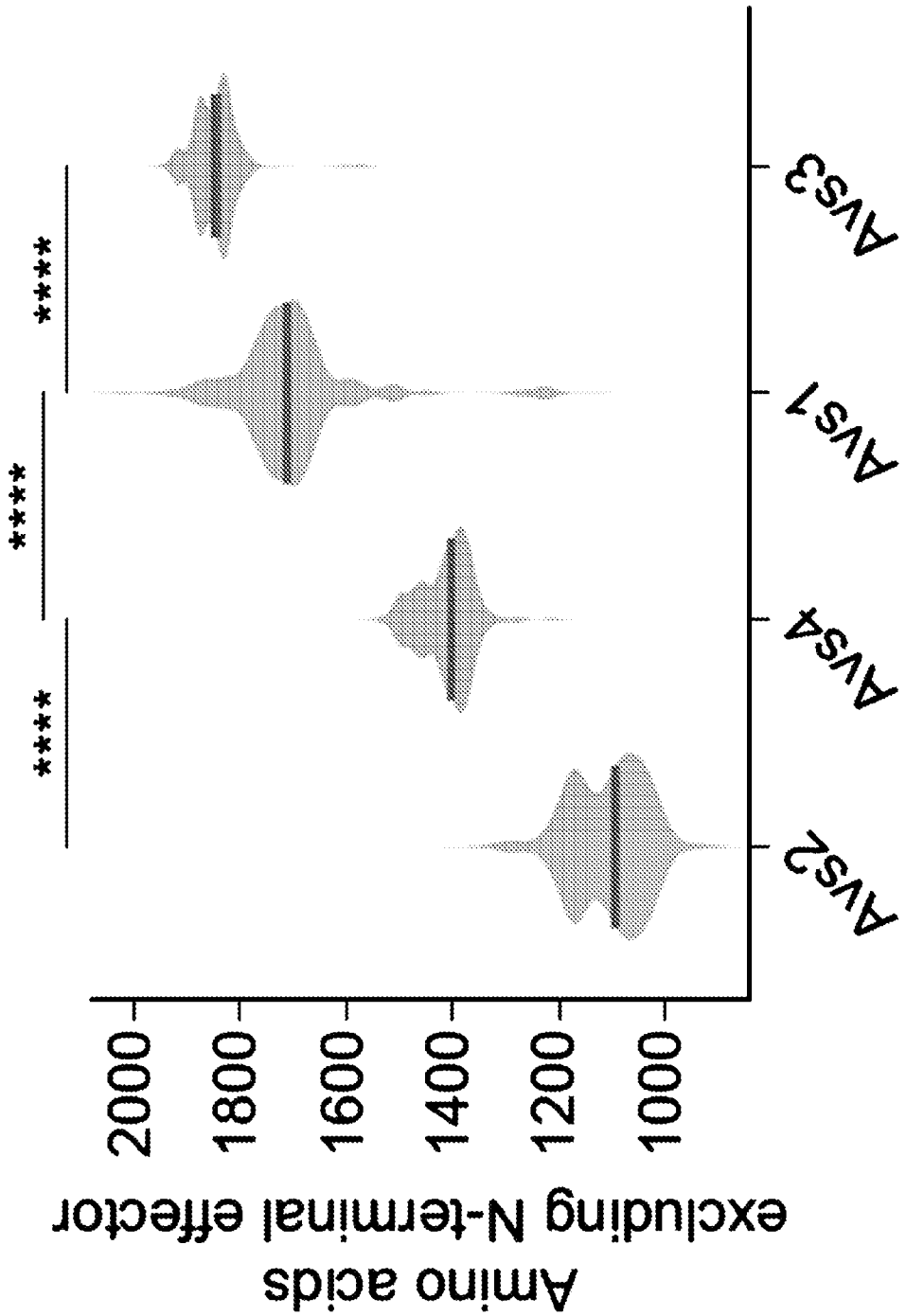

Amino acids
excluding N-terminal effector
2000
1800
1600
1400
1200
1000
AVS2
AVS4
AVS1
AVS3
****
****
****
****

FIG. 20C

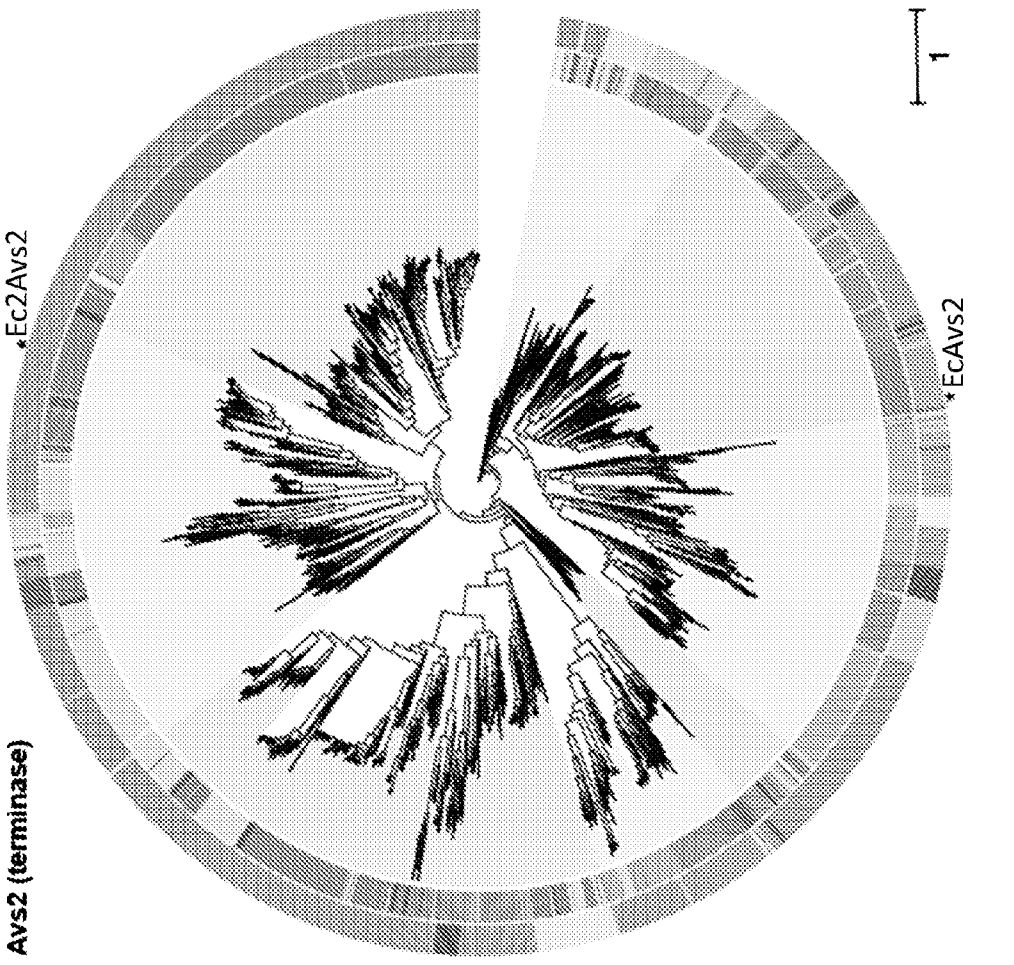

Avs2 (terminase)
Ec2Avs2
EcAvs2
1
Taxonomy (inner ring)
Proteobacteria
Actinobacteria
Bacteroidetes
Firmicutes
Cyanobacteria
Archaea
PVC group
Other
Unclassified
N-terminal domains (outer ring)
NTD ATPase sensor
Cap4
Mrr
HNH
REase
Putative REase
MBL + serine protease
MBL + serine protease
Trypco2 + serine protease
Caspase + EAD10
Nucleosidase
CMP hydrolase
TIR
SIR2
Transmembrane
Transmembrane
Unknown-1
Unknown-2
Unclassified
Short
0 1 2 3
log10 count
Nucleases
Proteases
Other

FIG. 20D

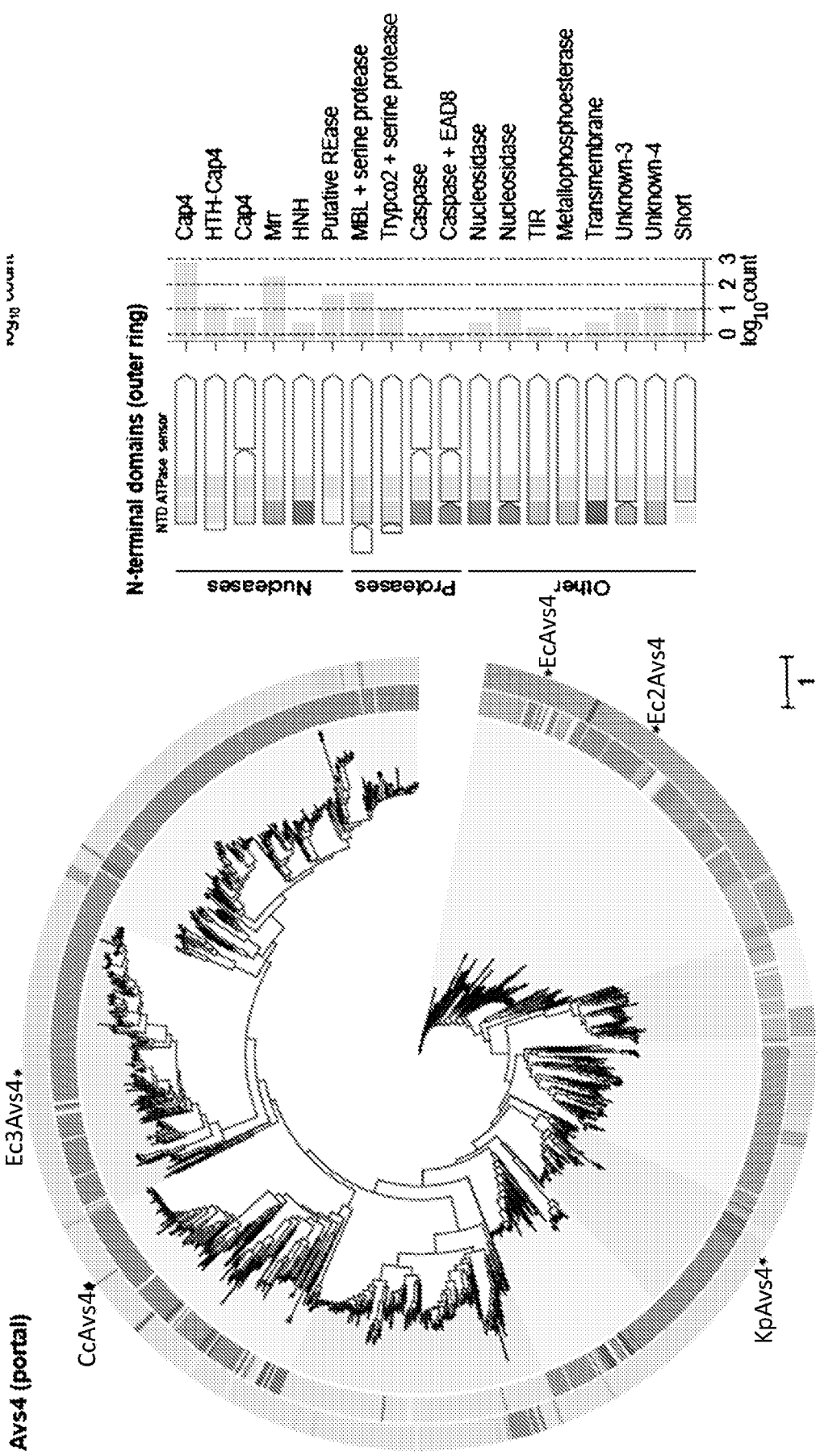

Avs4 (portal)
Ec3Avs4
CcAvs4
EcAvs4
Ec2Avs4
KpAvs4
1
N-terminal domains (outer ring)
NTD ATPase sensor
Nucleases
Proteases
Other
Cap4
HTH-Cap4
Cap4
Mrr
HNH
Putative REase
MBL + serine protease
Trypco2 + serine protease
Caspase
Caspase + EAD8
Nucleosidase
Nucleosidase
TIR
Metallophosphoesterase
Transmembrane
Unknown-3
Unknown-4
Short
0 1 2 3
log10count

FIG. 20E

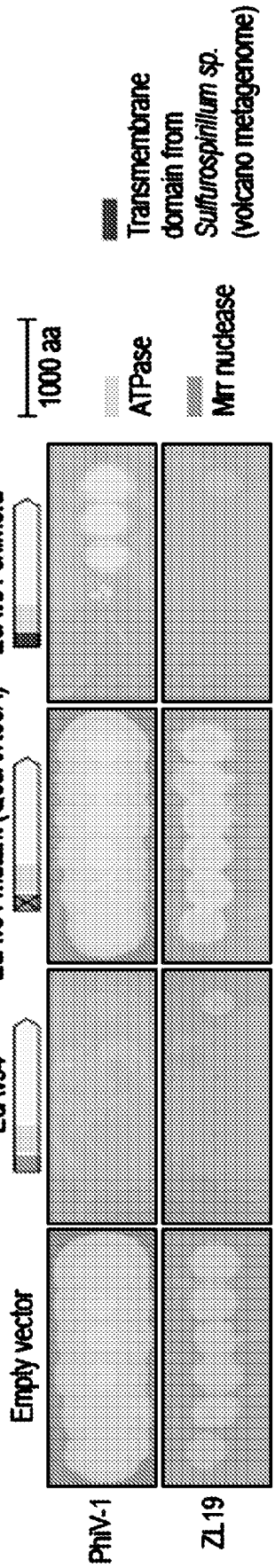

Transmembrane domain from Sulfurospirillum sp. (volcano metagenome)
1000 aa
ATPase
Mrr nuclease
EcAvs4 chimera
EcAvs4 mutant (Q63A/K65A)
EcAvs4
Empty vector
PhiV-1
Z1.19

FIG. 20F

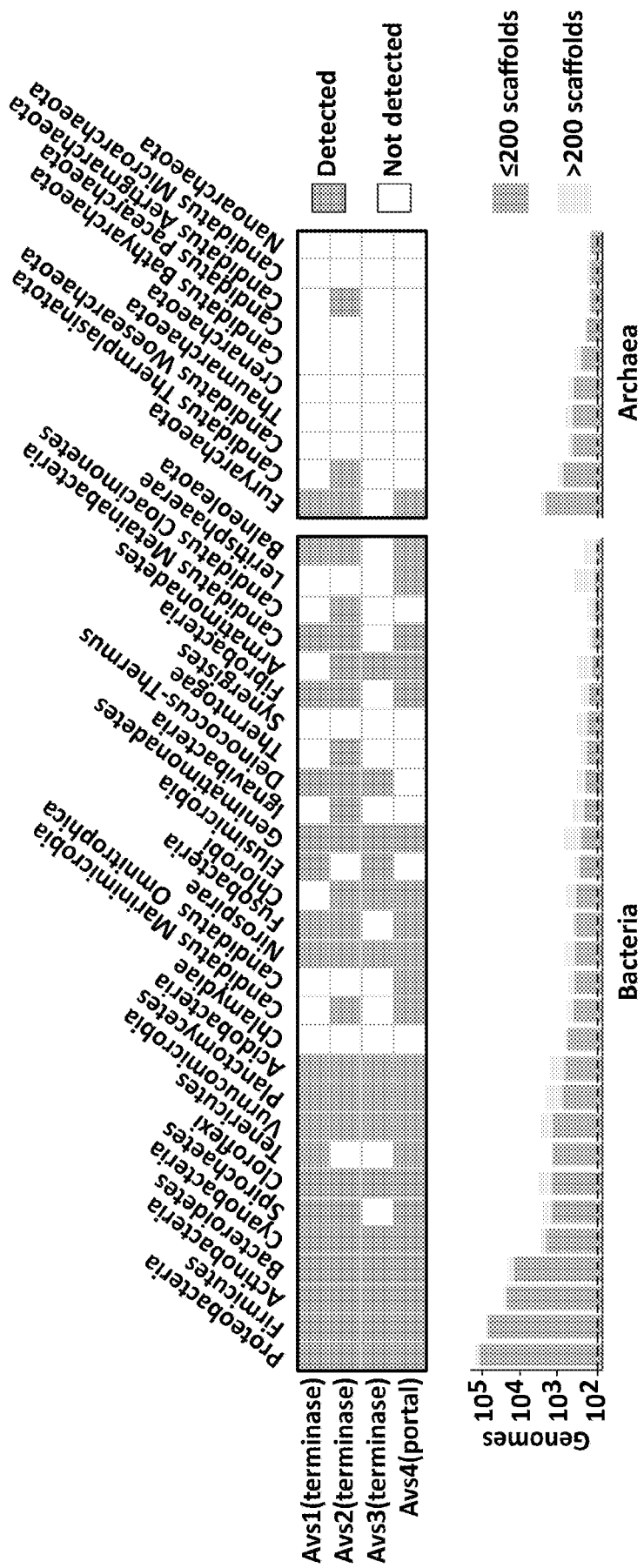

Detected
Not detected
≤200 scaffolds
>200 scaffolds
Archaea
Bacteria
Genomes
10^5
10^4
10^3
10^2
Avs1(terminase)
Avs2(terminase)
Avs3(terminase)
Avs4(portal)
Proteobacteria
Firmicutes
Actinobacteria
Bacteroidetes
Cyanobacteria
Spirochaetes
Chloroflexi
Tenericutes
Verrucomicrobia
Planctomycetes
Acidobacteria
Chlamydiae
Candidatus Marinimicrobia
Nitrospirae
Fusobacteria
Chlorobi
Elusimicrobia
Gemmatimonadetes
Ignavibacteria
Deinococcus-Thermus
Thermotogae
Synergistes
Fibrobacteres
Armatimonadetes
Candidatus Metainabacteria
Candidatus Cloacimonetes
Lentisphaerae
Balneolaeota
Euryarchaeota
Candidatus Thermoplasmatota
Thaumarchaeota
Crenarchaeota
Candidatus Woesearchaeota
Candidatus Bathyarchaeota
Candidatus Pacearchaeota
Candidatus Aerophobetes
Candidatus Micrarchaeota
Nanoarchaeota

FIG. 21A

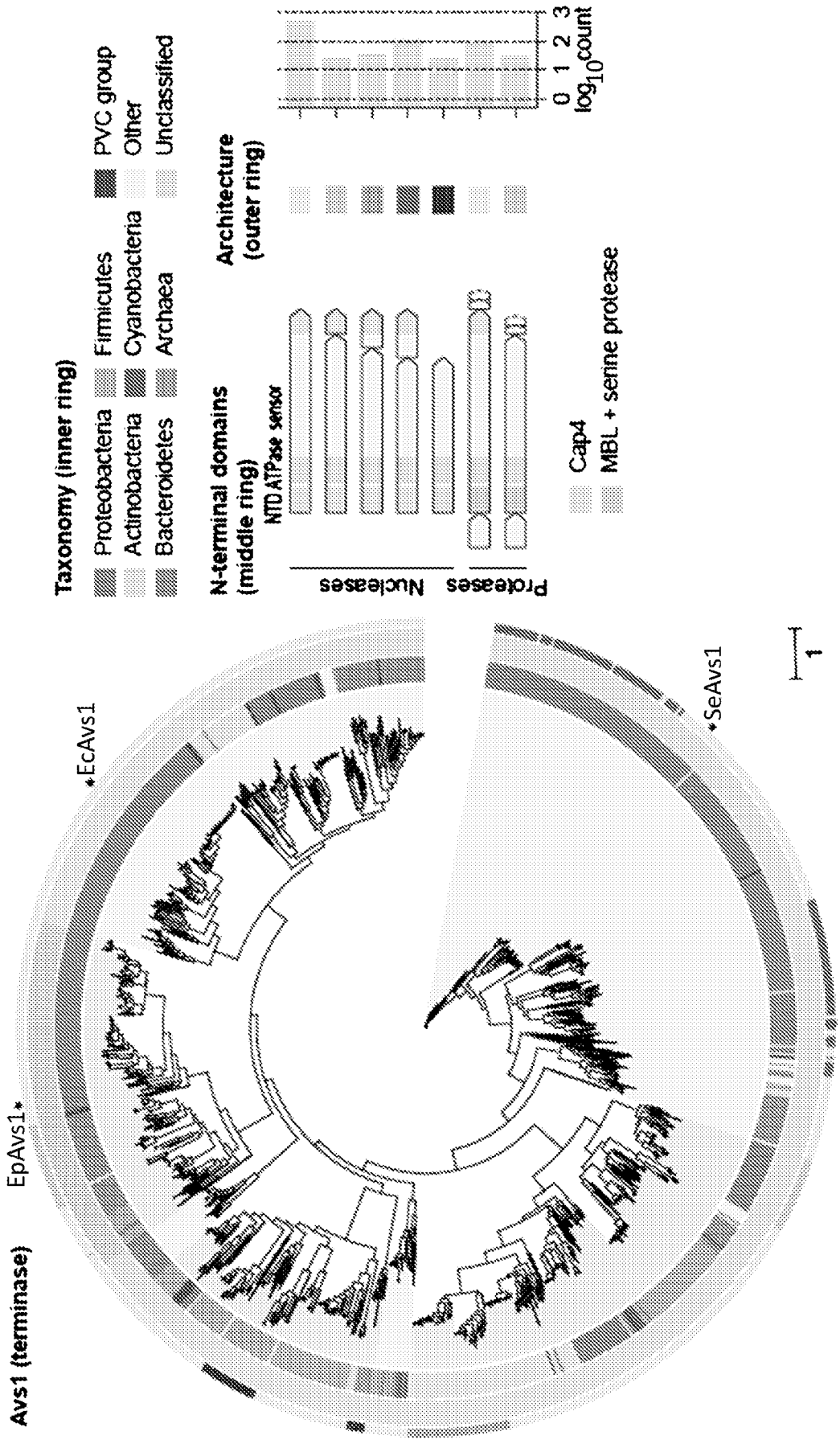

Avs1 (terminase)
EpAvs1
EcAvs1
SeAvs1
1
Taxonomy (inner ring)
Proteobacteria
Actinobacteria
Bacteroidetes
Firmicutes
Cyanobacteria
Archaea
PVC group
Other
Unclassified
N-terminal domains (middle ring)
NTD
ATPase sensor
Nucleases
Proteases
Architecture (outer ring)
Cap4
MBL + serine protease
0 1 2 3
log10count

FIG. 21B

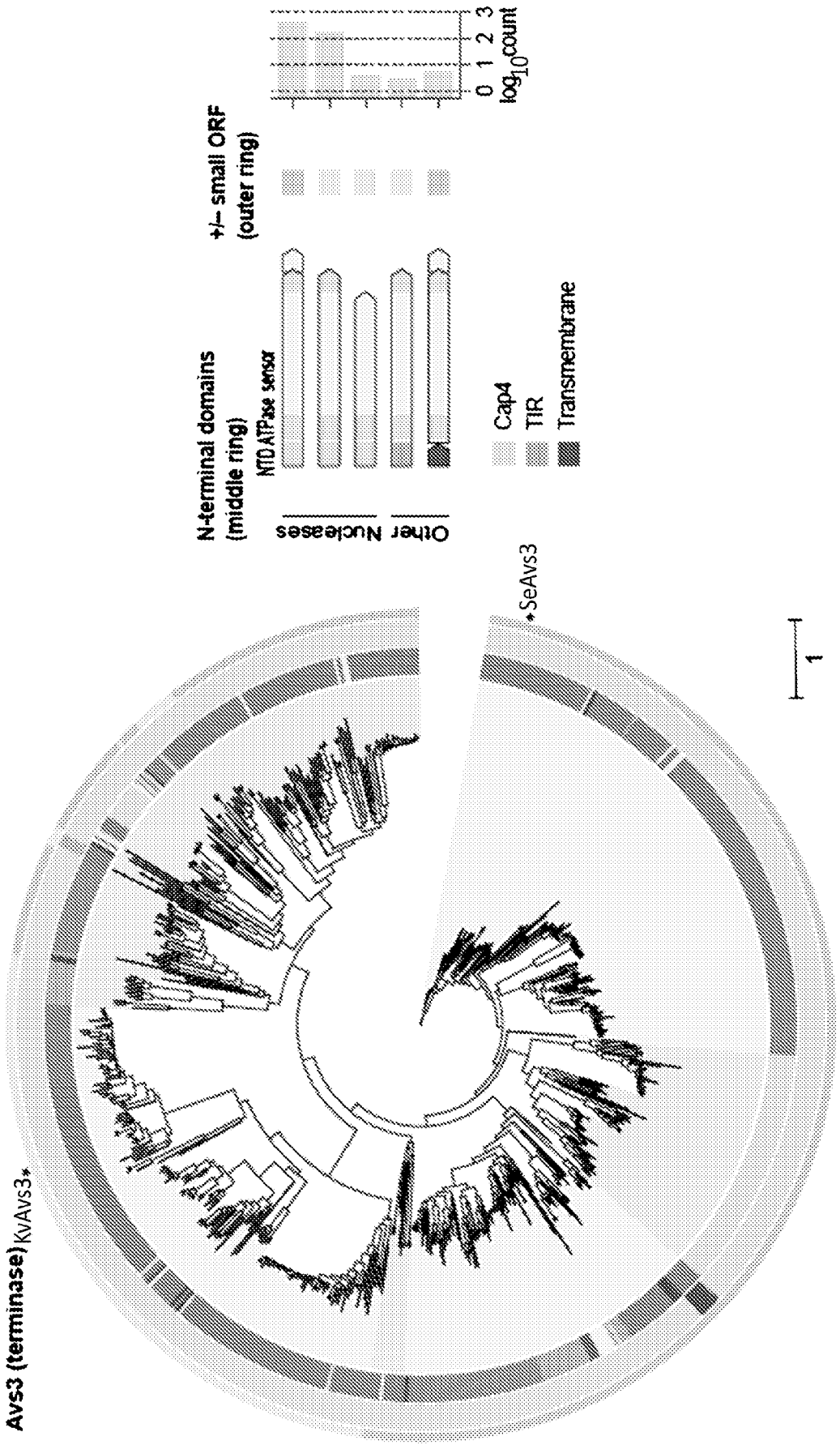

Avs3 (terminase)KvAvs3*
*SeAvs3
1
N-terminal domains
(middle ring)
NTD ATPase sensor
Other Nucleases
+/- small ORF
(outer ring)
log10count
0 1 2 3
Cap4
TIR
Transmembrane

FIG. 21C

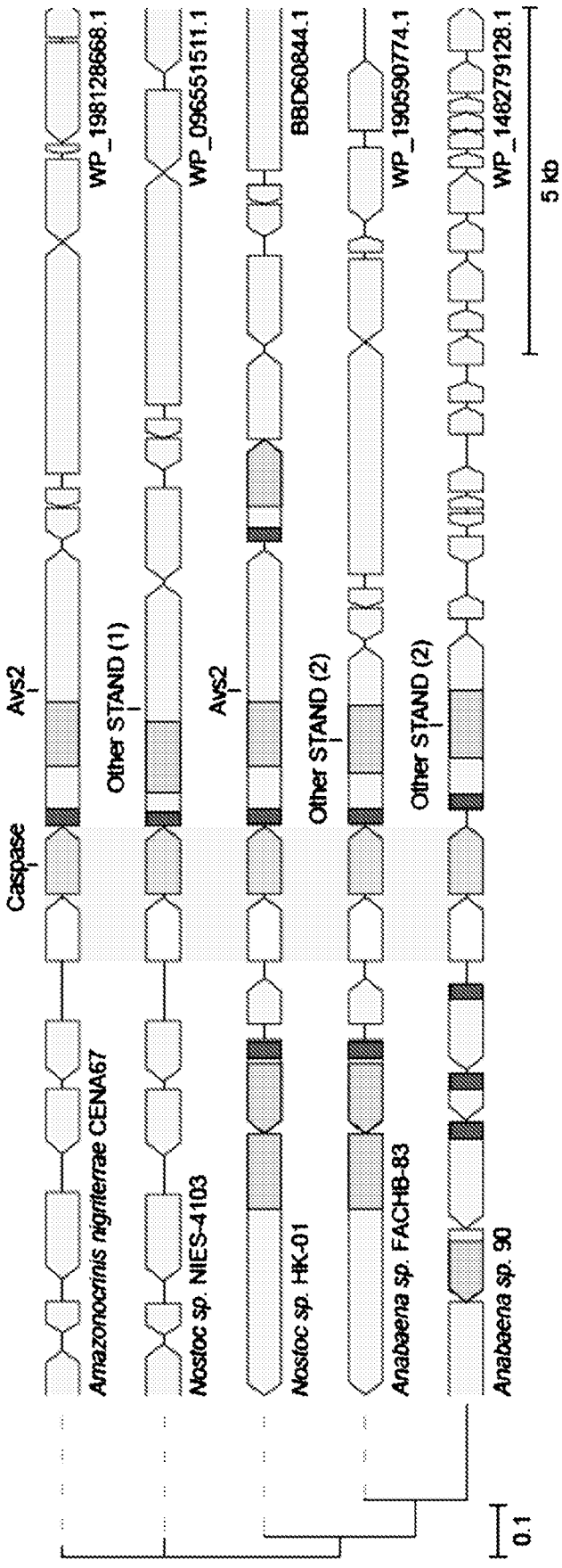


FIG. 22A

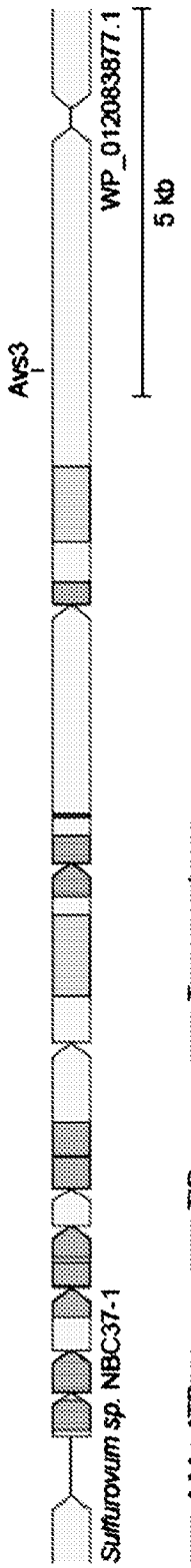


FIG. 22B

FIG. 23A

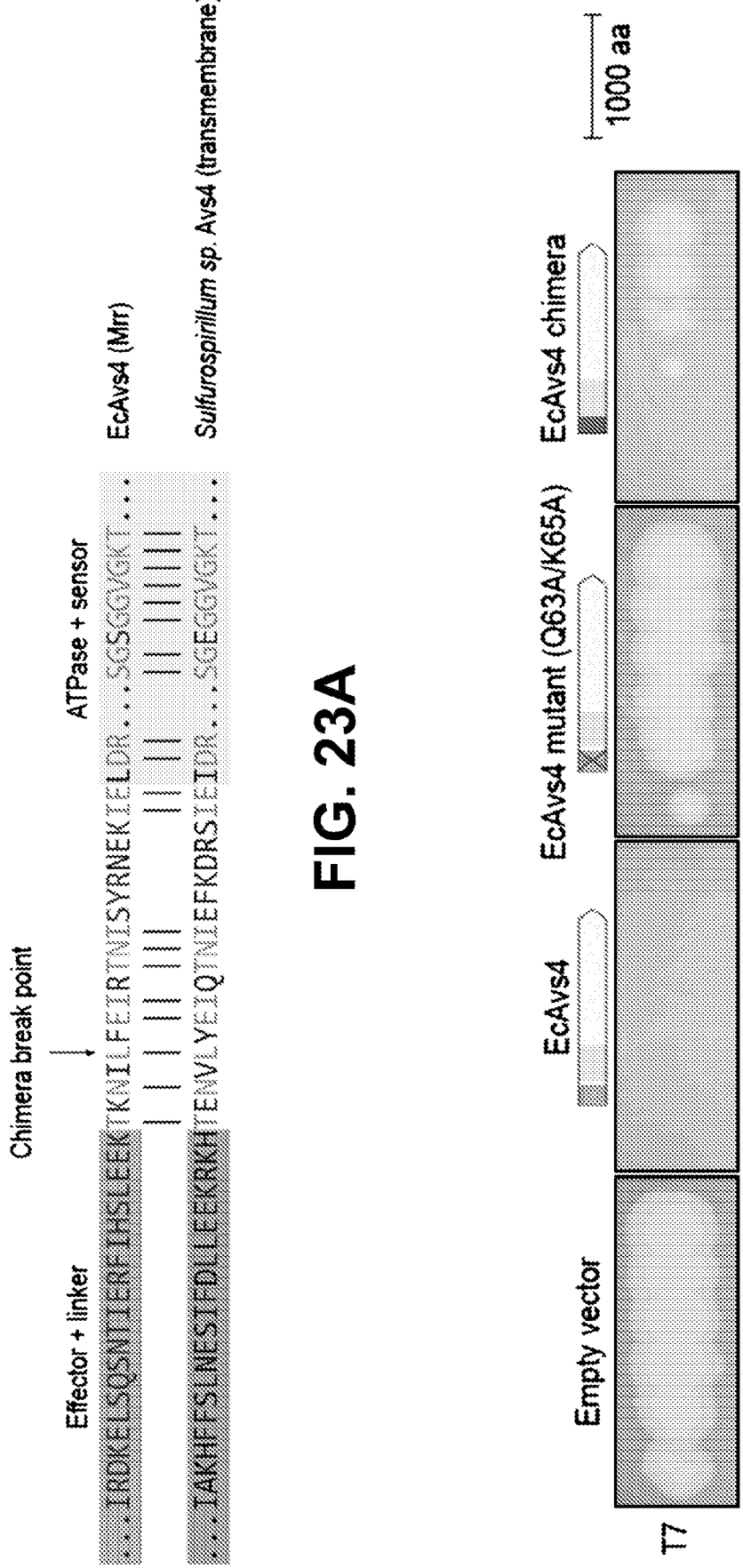


FIG. 23B

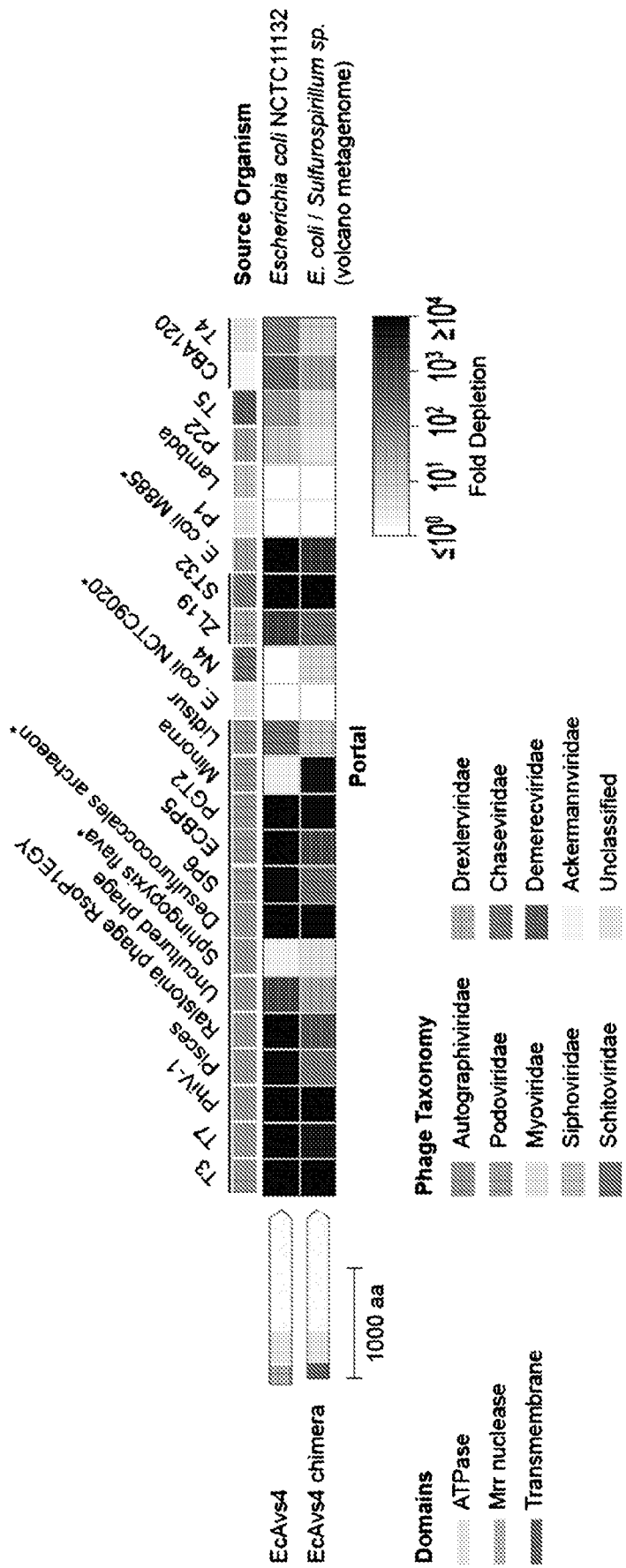

Source Organism
Escherichia coli NCTC11132
E. coli / Sulfurospirillum sp. (volcano metagenome)
Fold Depletion
≤10⁰ 10¹ 10² 10³ ≥10⁴
Portal
T3
T7
phiV-1
Pseas
Ralstonia phage RsoP1EGY
Uncultured phage
Sphingopyxis flava*
Desulfurococcales archaeon*
SP6
ECBP5
PGT2
Mimoma
Lidisur
E. coli NCTC9020*
N4
ZL-19
ST32
E. coli M885*
P1
Lambda
P22
T5
CBA120
T4
Phage Taxonomy
Autographiviridae
Podoviridae
Myoviridae
Siphoviridae
Schitoviridae
Drexlerviridae
Chaseviridae
Demerecviridae
Ackermannviridae
Unclassified
Domains
ATPase
Mrr nuclease
Transmembrane
EcAvs4
EcAvs4 chimera
1000 aa

FIG. 23C

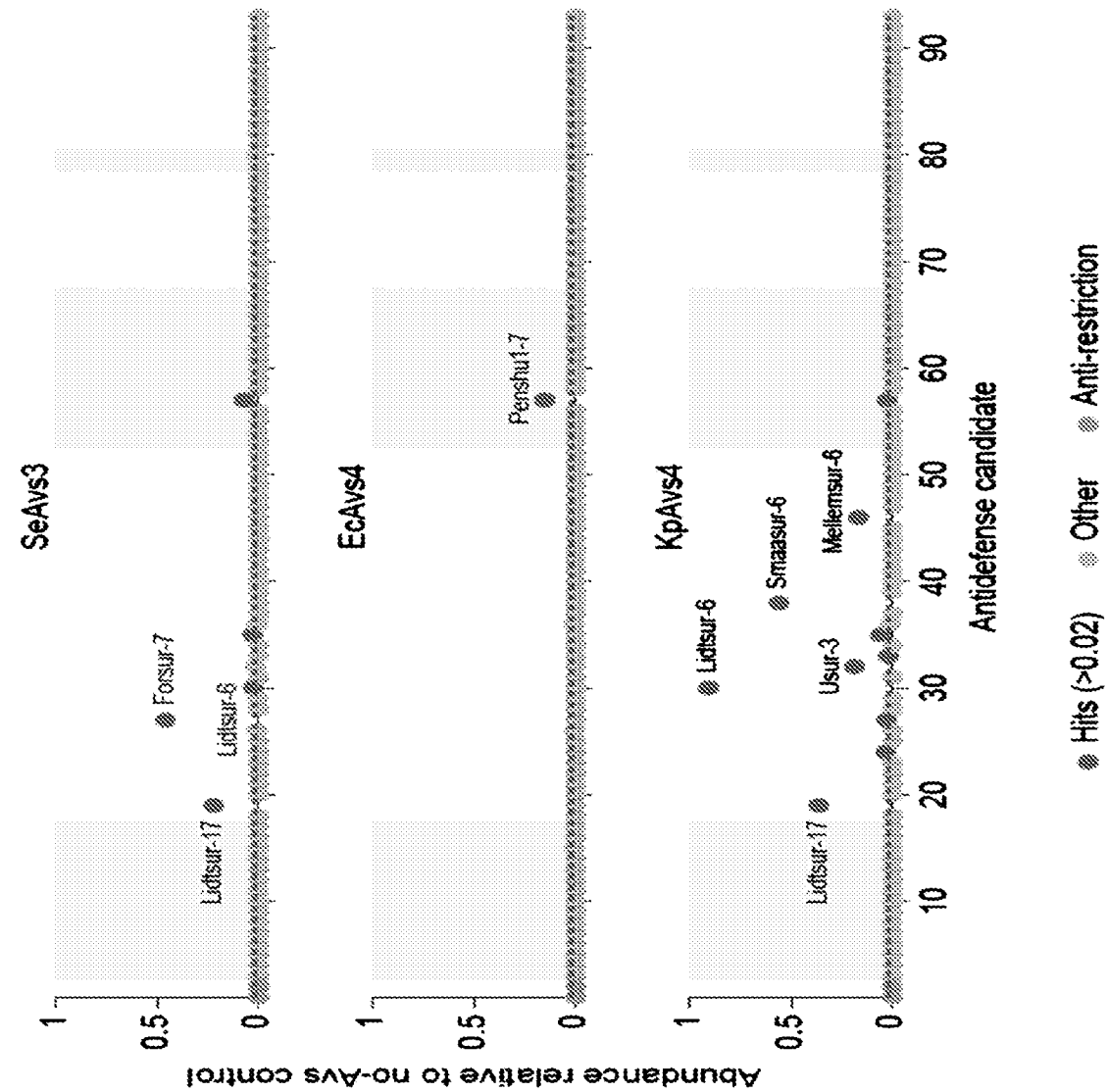


FIG. 24B

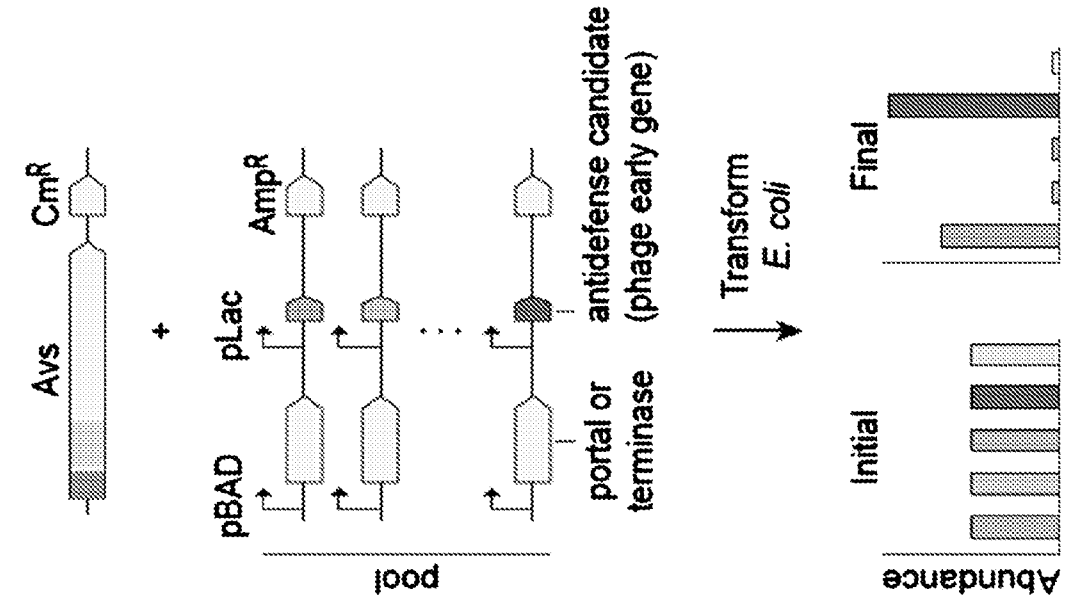


FIG. 24A

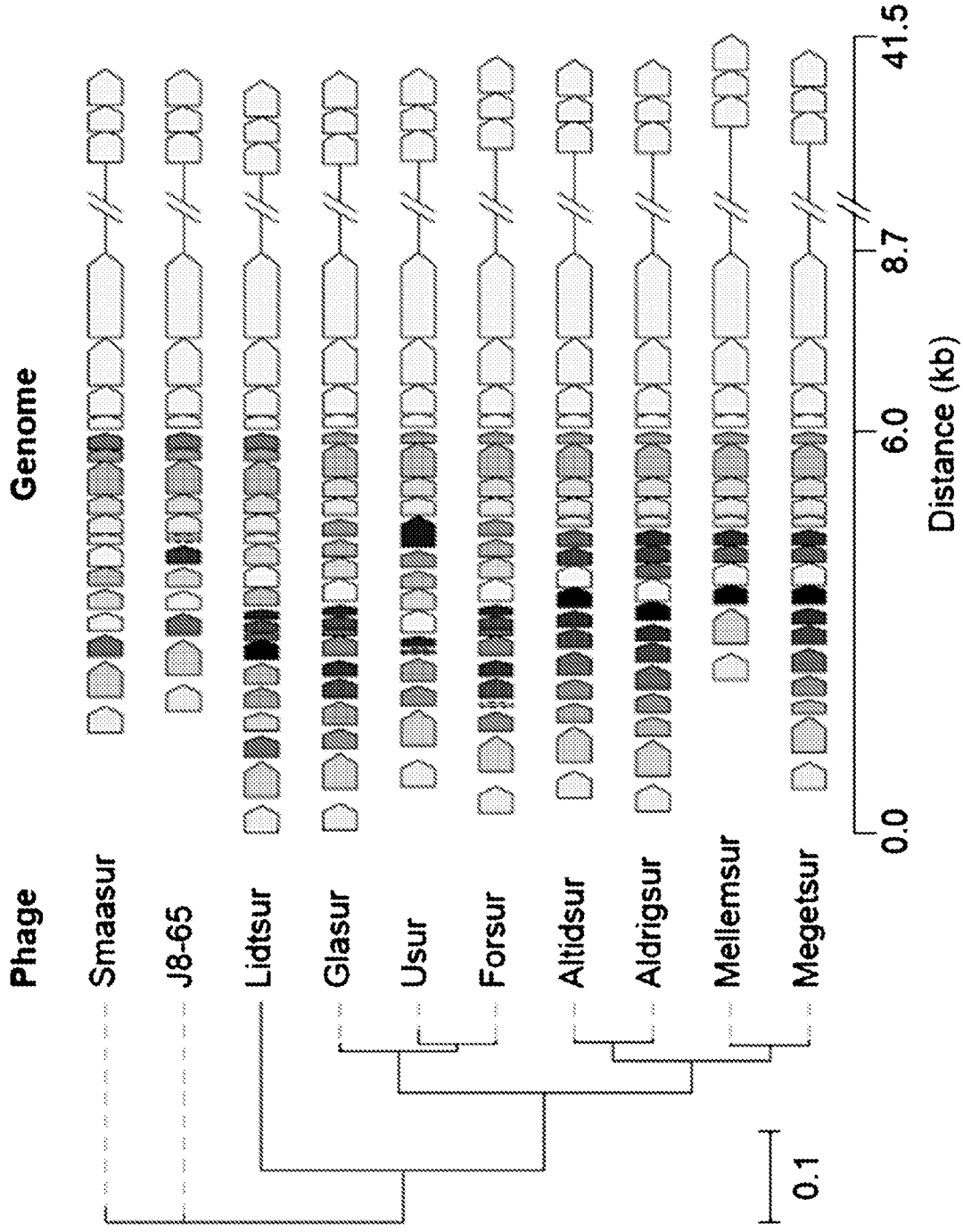

Phage
Genome
Smaasur
J8-65
Lidtsur
Glasur
Usur
Forsur
Altidsur
Aldrigsur
Mellemsur
Megetsur
0.1
0.0
6.0
8.7
41.5
Distance (kb)

FIG. 24C

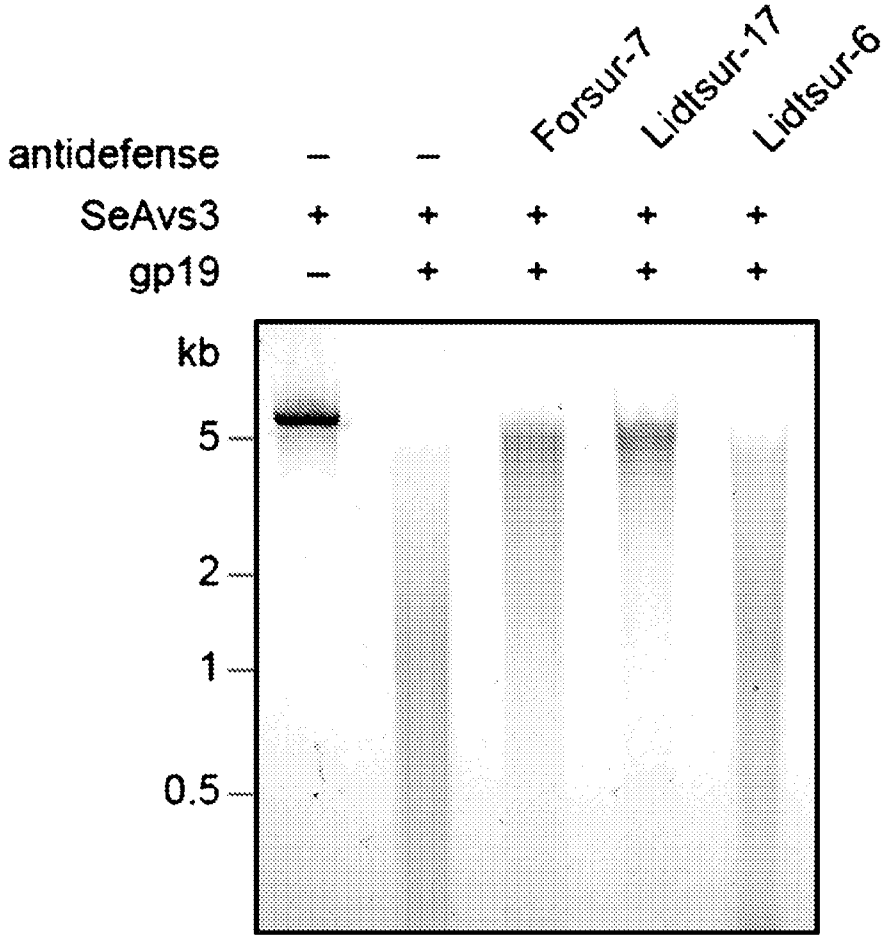

Forsur-7
Lidtsur-17
Lidtsur-6
antidefense – – + + +
SeAvs3 + + + + +
gp19 – + + + +
kb
5
2
1
0.5

FIG. 24D

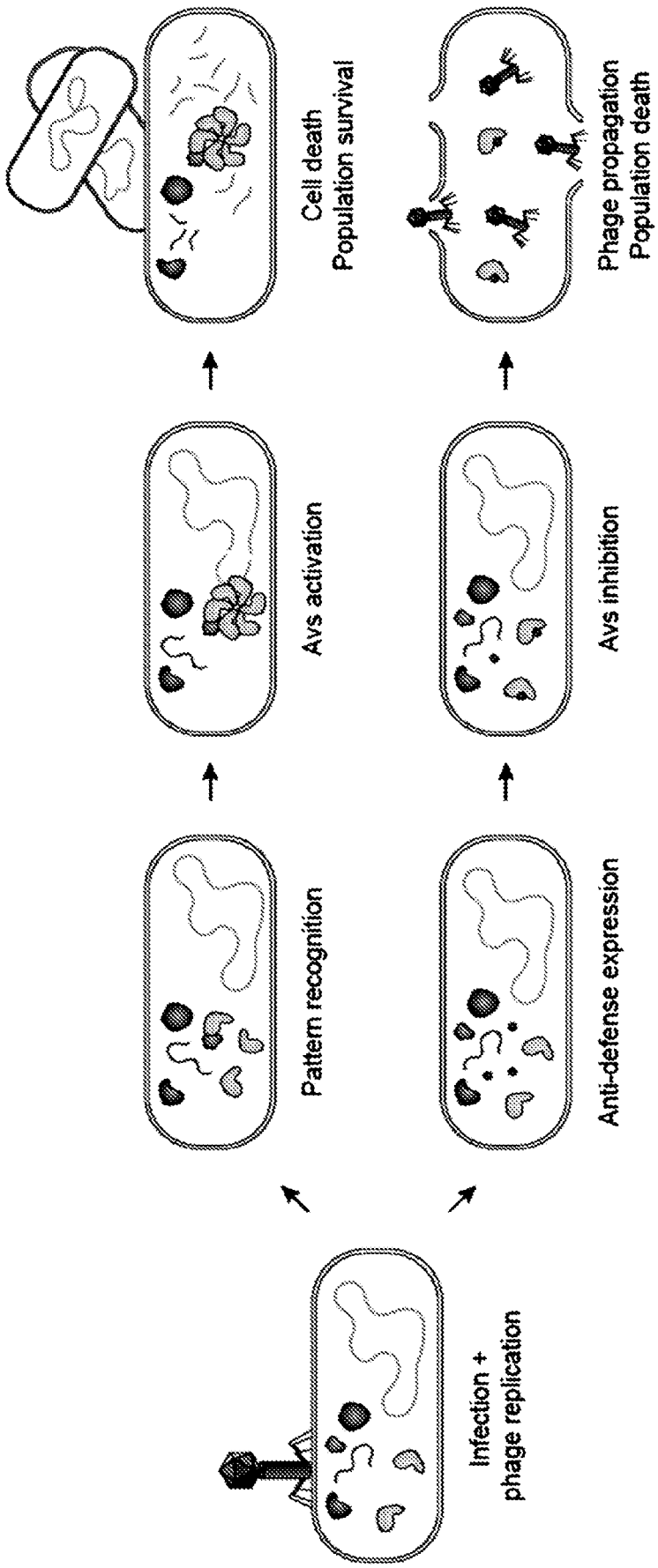

Infection +
phage replication
Pattern recognition
Anti-defense expression
Avs activation
Avs inhibition
Cell death
Population survival
Phage propagation
Population death

FIG. 24E

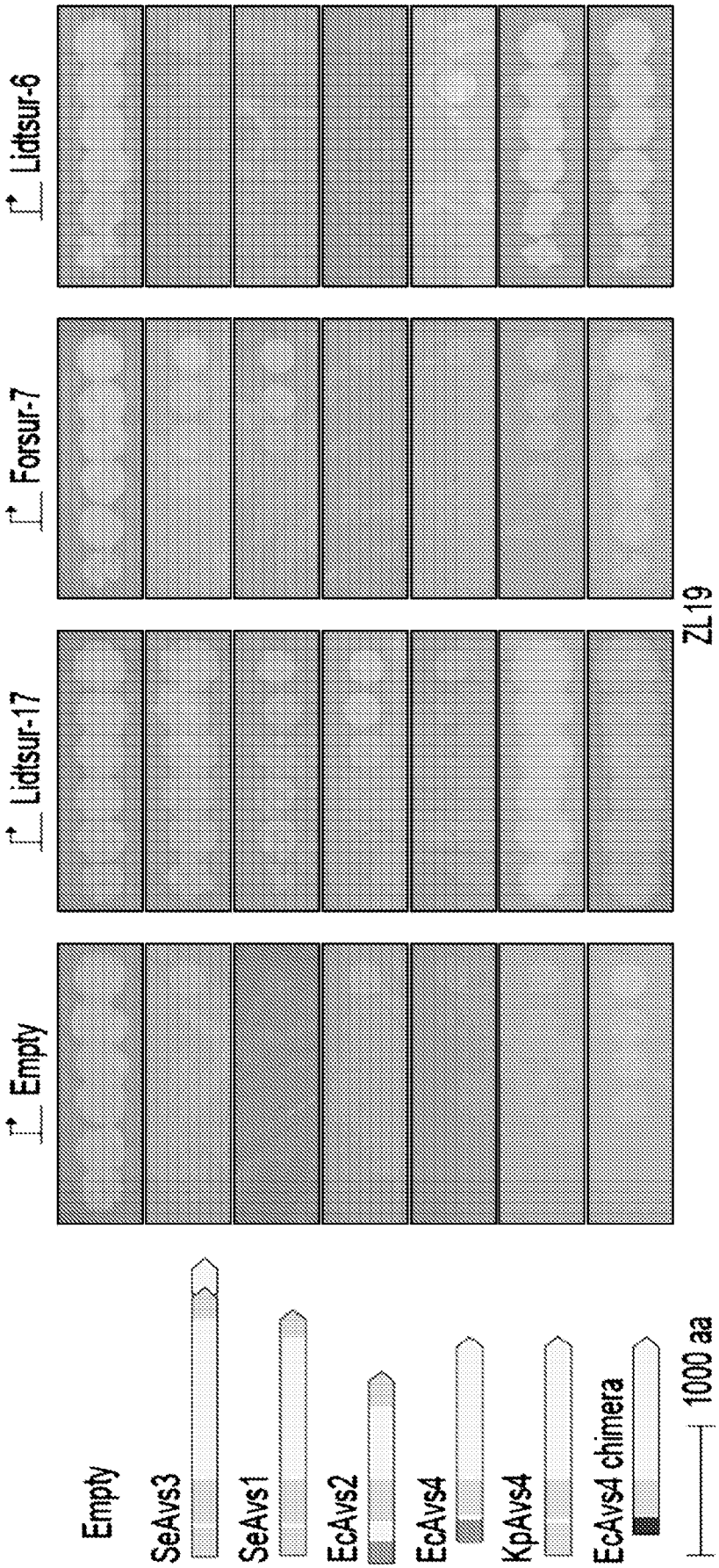

Empty
Lidtsur-17
Forsur-7
Lidtsur-6
ZL19
Empty
SeAvs3
SeAvs1
EcAvs2
EcAvs4
KpAvs4
EcAvs4 chimera
1000 aa

FIG. 25A

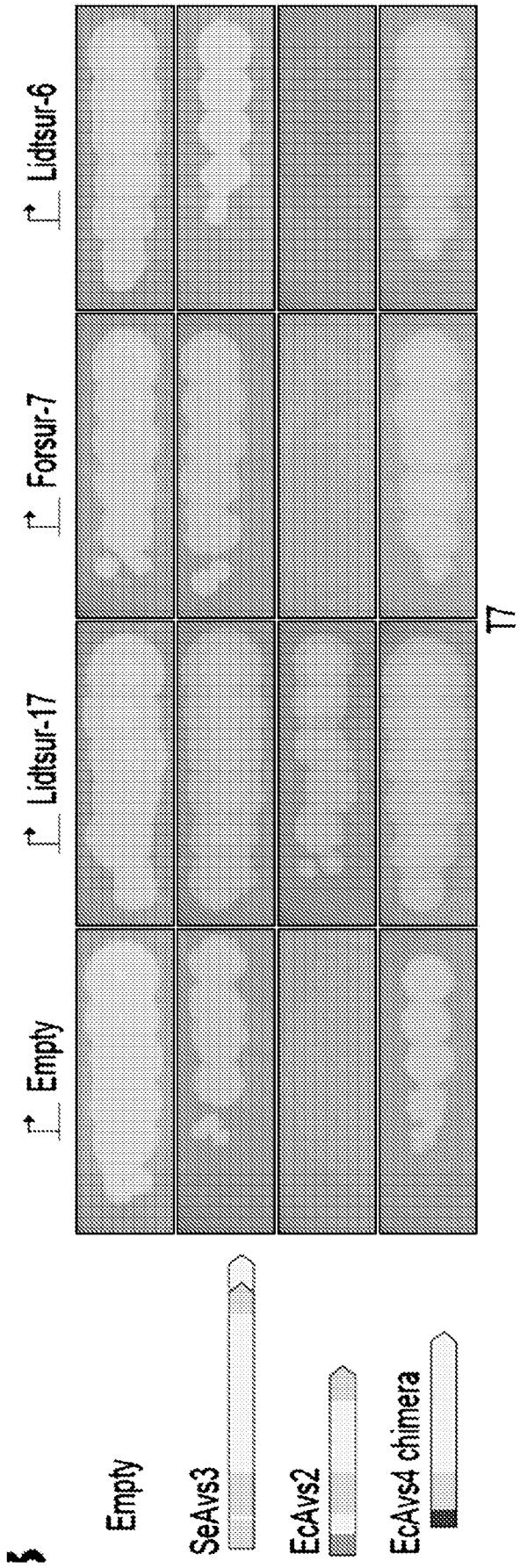

Empty
Lidtsur-17
Forsur-7
Lidtsur-6
T7
Empty
SeAvs3
EcAvs2
EcAvs4 chimera

FIG. 25B

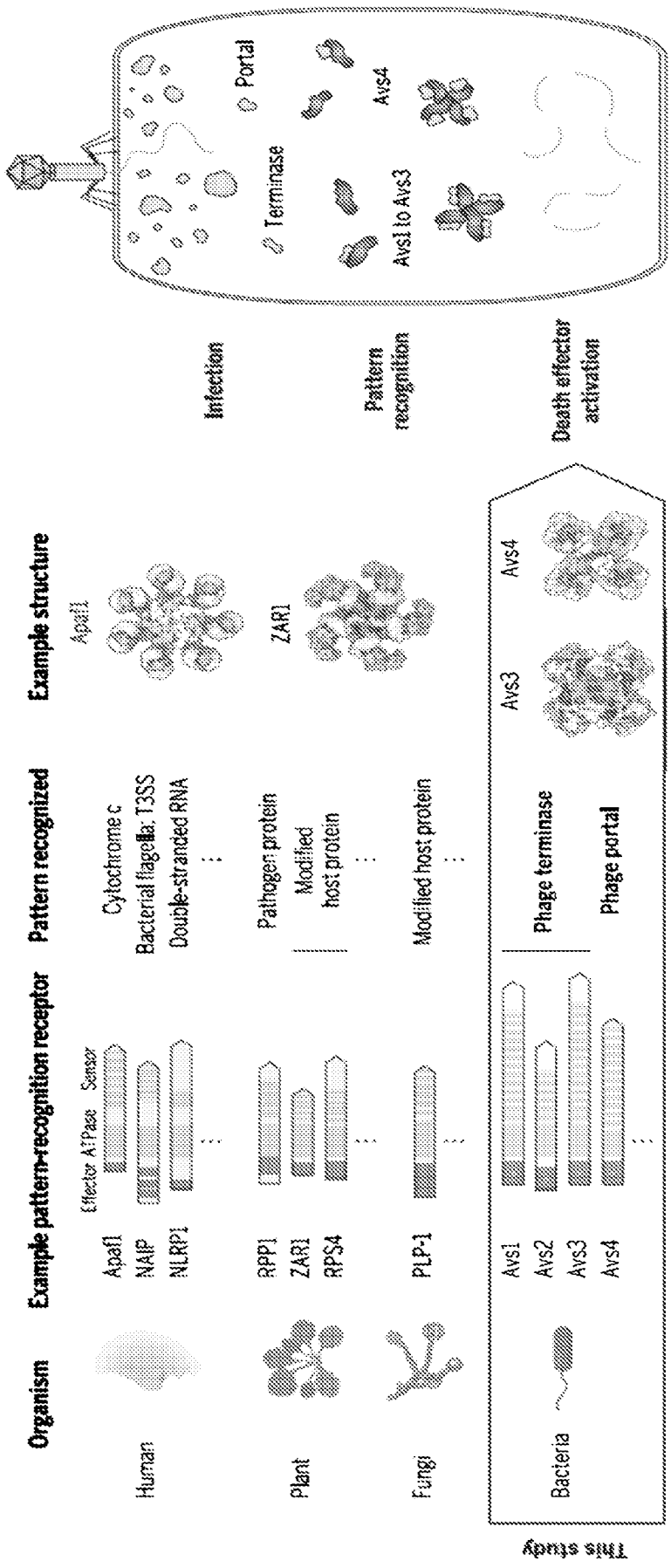

Organism
Example pattern-recognition receptor
Effector ATPase Sensor
Pattern recognized
Example structure
Human
Apaf1
NAIP
NLRP1
Cytochrome c
Bacterial flagella; T3SS
Double-stranded RNA
Apaf1
Plant
RPP1
ZAR1
RPS4
Pathogen protein
Modified host protein
ZAR1
Fungi
PLP-1
Modified host protein
Bacteria
Avs1
Avs2
Avs3
Avs4
Phage terminase
Phage portal
Avs3
Avs4
Infection
Pattern recognition
Death effector activation
Portal
Terminase
Avs1 to Avs3
Avs4
This study

FIG. 26

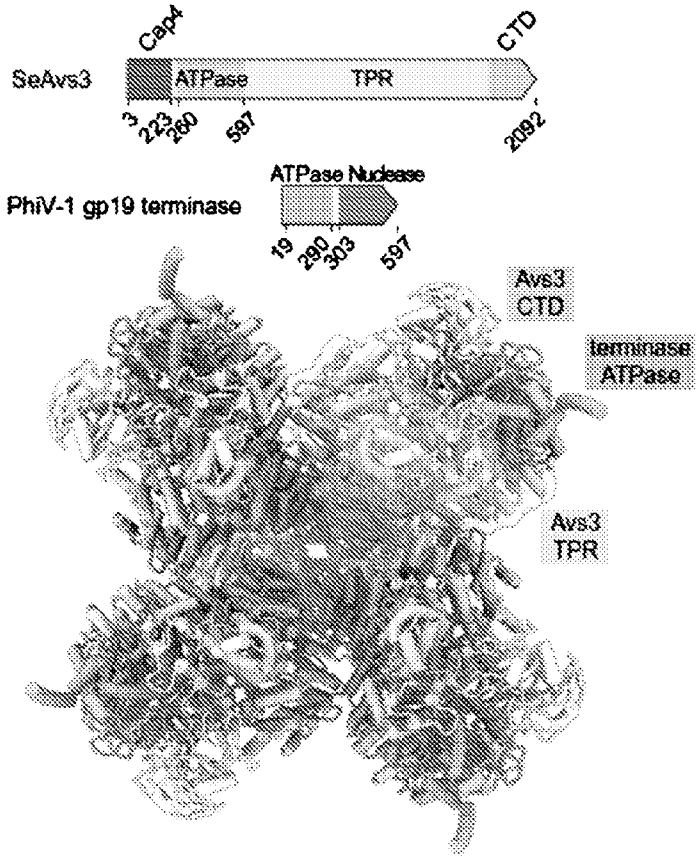


FIG. 27A

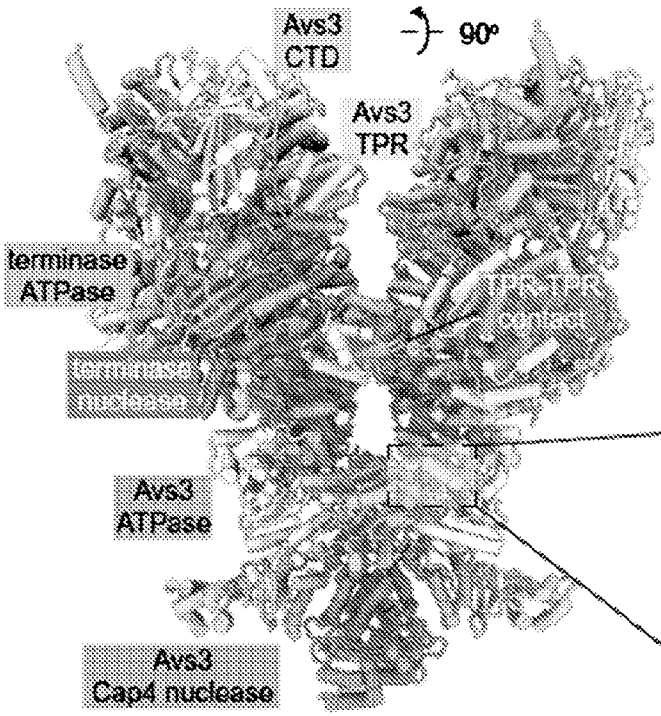


FIG. 27B

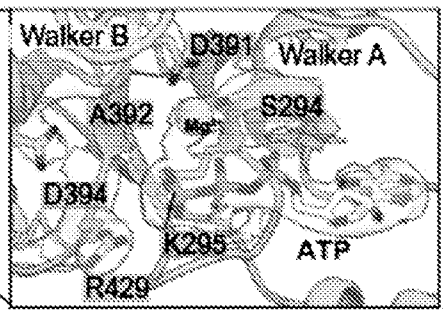


FIG. 27F

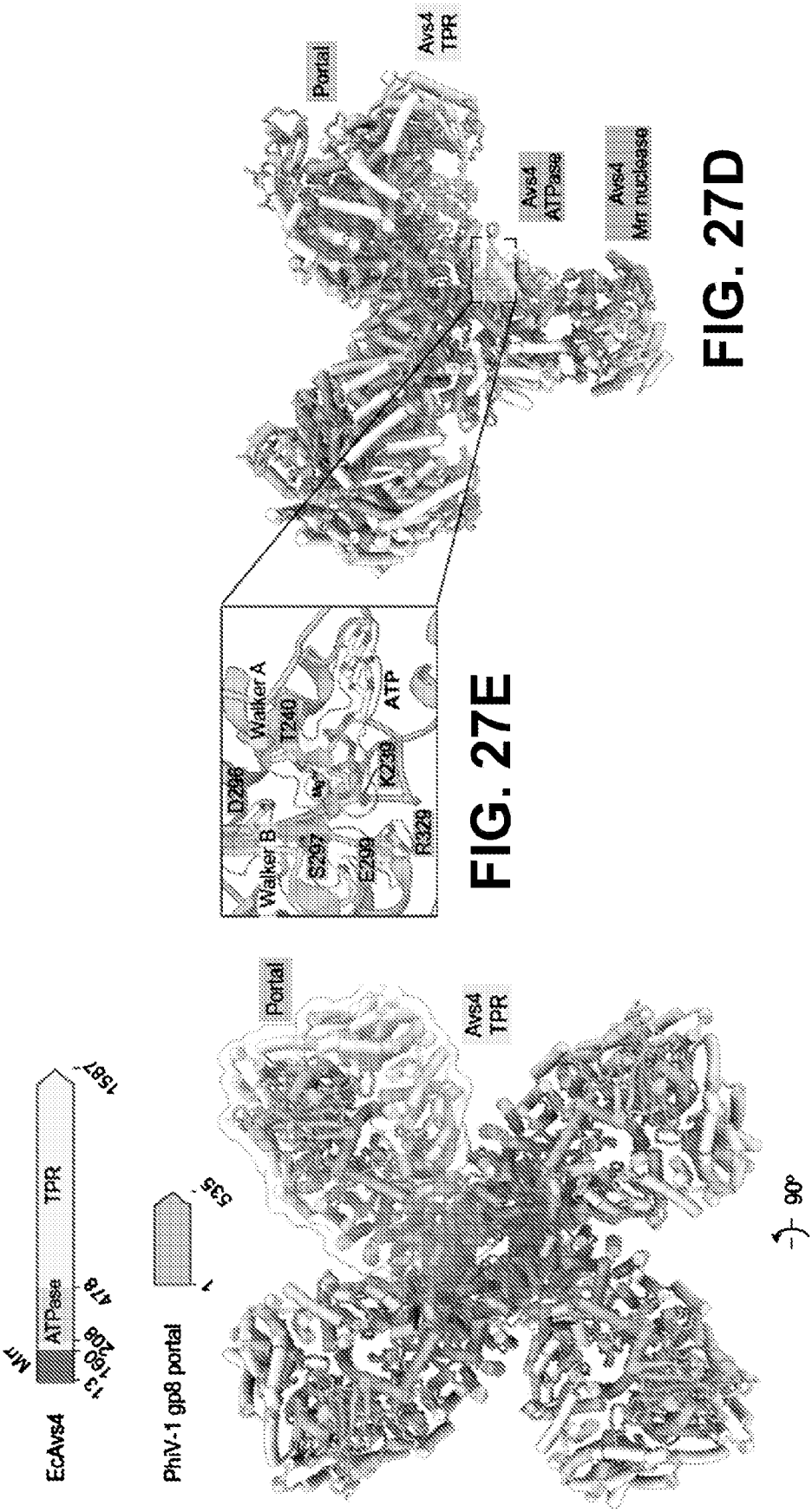
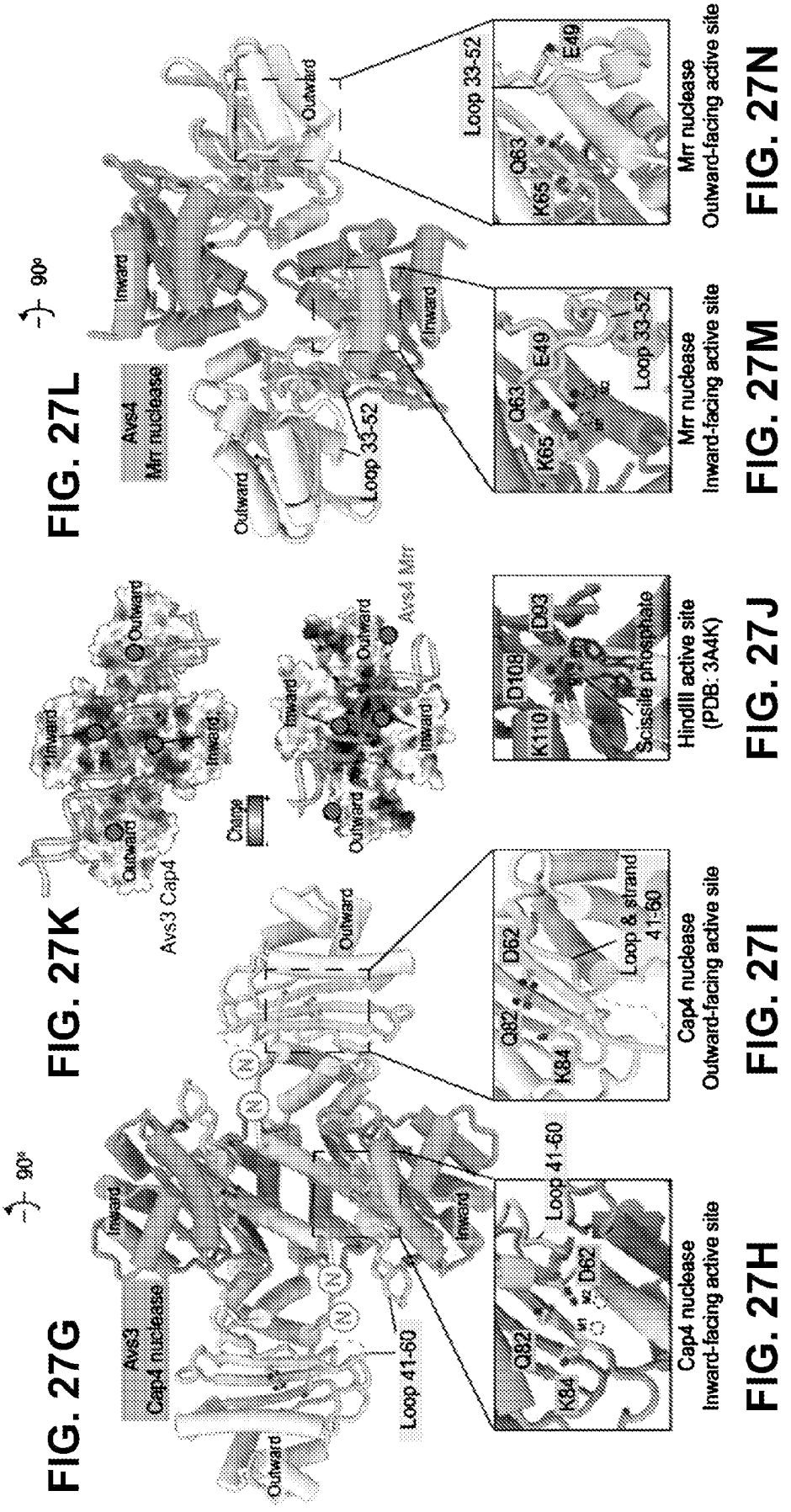


FIG. 27C

FIG. 27D

FIG. 27E


FIG. 27G
90°
Avs3
Cap4 nuclease
Inward
Outward
Loop 41-60
Q82
K84
D62
Cap4 nuclease
Inward-facing active site
FIG. 27H
Avs3 Cap4
Outward
Inward
Charge
Avs4 Mrr
FIG. 27K
Loop & strand
41-60
Cap4 nuclease
Outward-facing active site
FIG. 27I
D108
D93
K110
Scissile phosphate
HindIII active site
(PDB: 3A4K)
FIG. 27J
90°
Avs4
Mrr nuclease
Loop 33-52
Q63
K65
E49
Mrr nuclease
Inward-facing active site
FIG. 27L
Mrr nuclease
Outward-facing active site
FIG. 27M
FIG. 27N

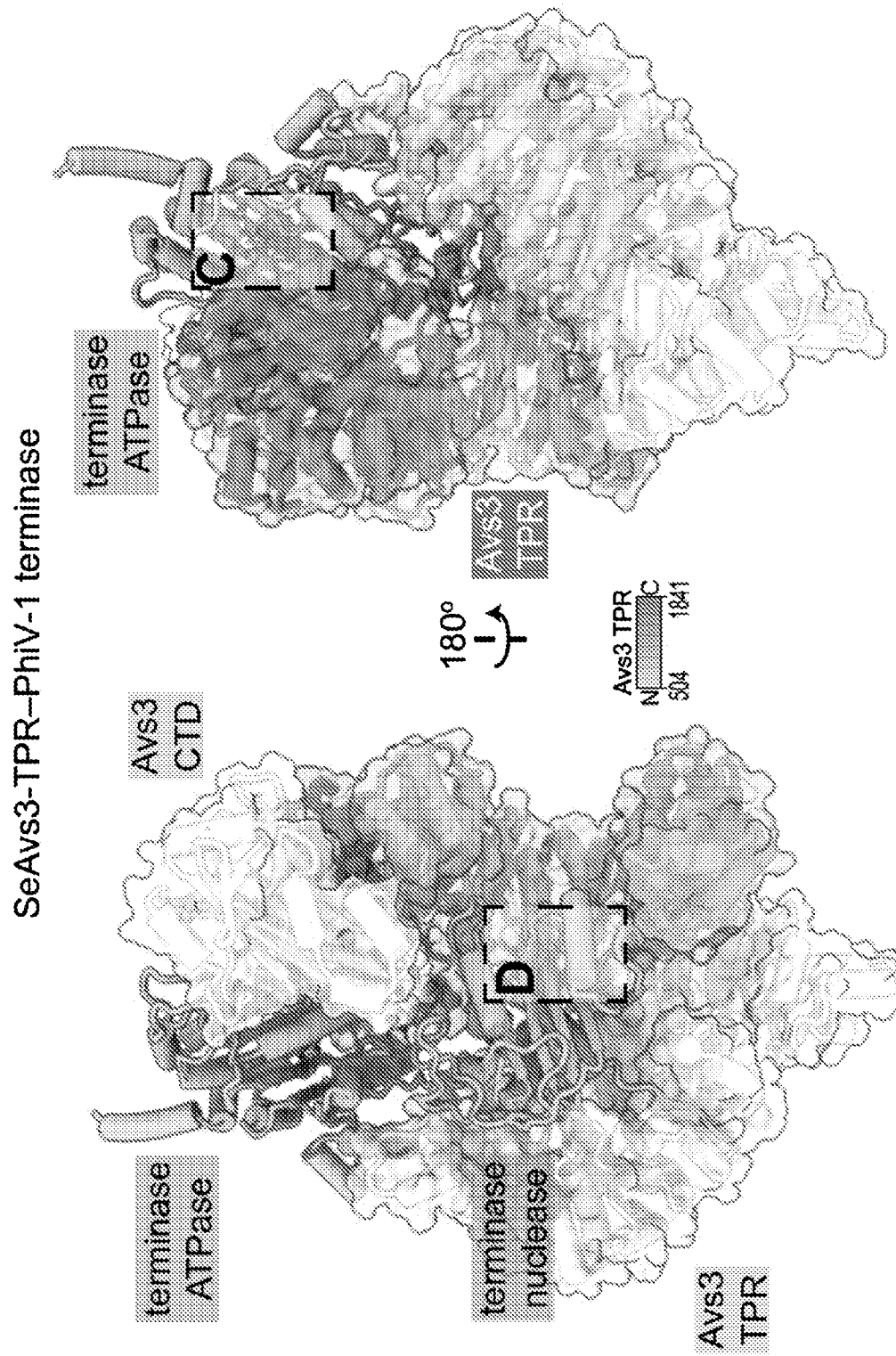

SeAvs3-TPR–PhiV-1 terminase
terminase ATPase
Avs3 CTD
terminase ATPase
terminase nuclease
Avs3 TPR
C
D
180°
Avs3 TPR
Avs3 TPR
N
C
504
1841

FIG. 28A


ATPase
Nuclease
SeAvs3 +
T7 terminase AlphaFold
ZL19 terminase AlphaFold
ST32 terminase AlphaFold
T4 terminase PDB: 3CPE

FIG. 28B

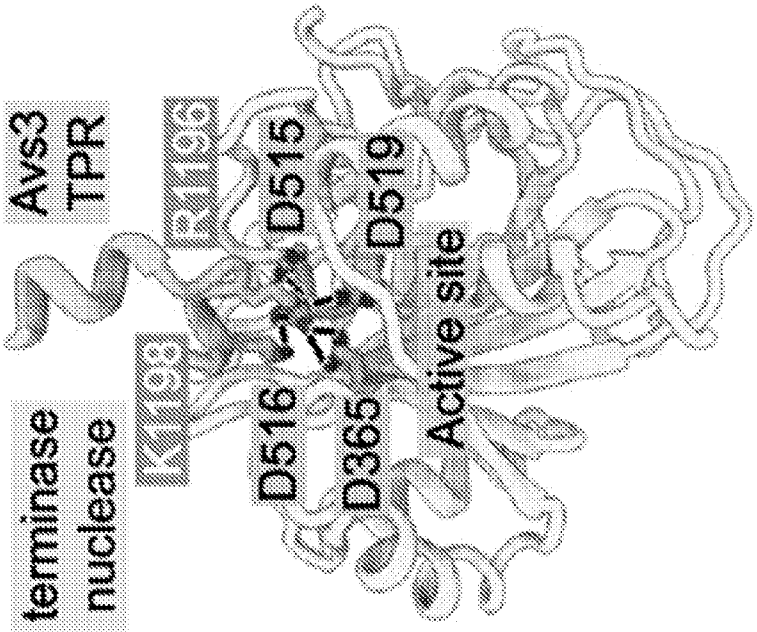


FIG. 28D

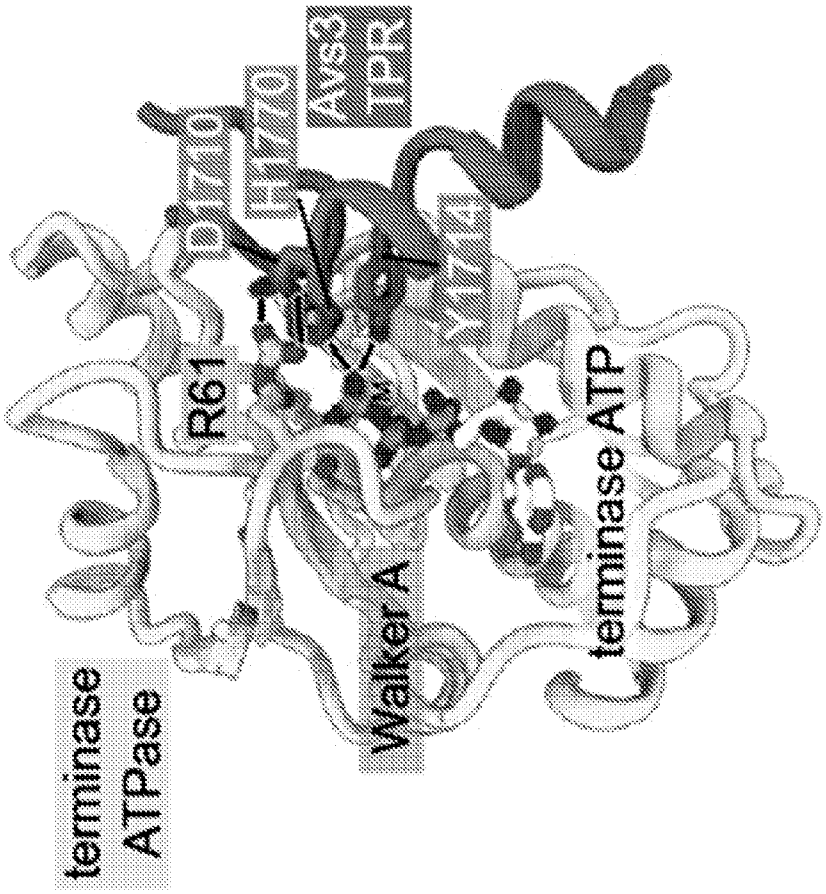


FIG. 28C

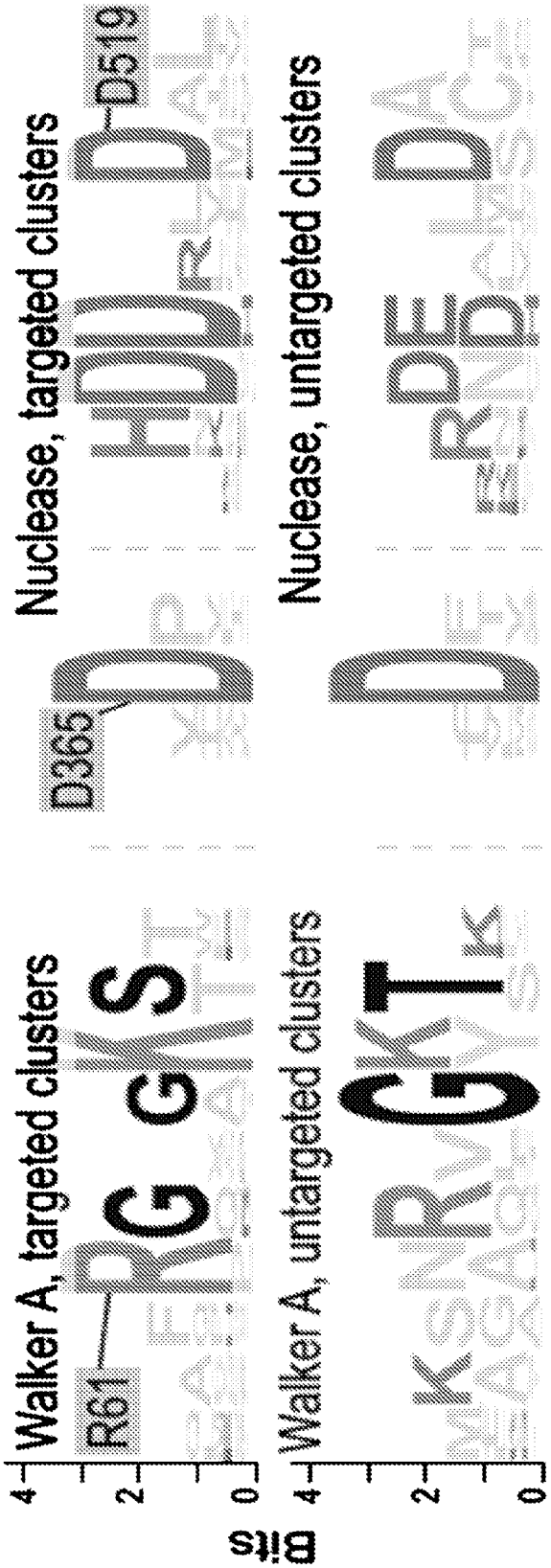

Walker A, targeted clusters
R61
Walker A, untargeted clusters
Nuclease, targeted clusters
D365
D519
Nuclease, untargeted clusters
Bits
4
2
0

FIG. 28E

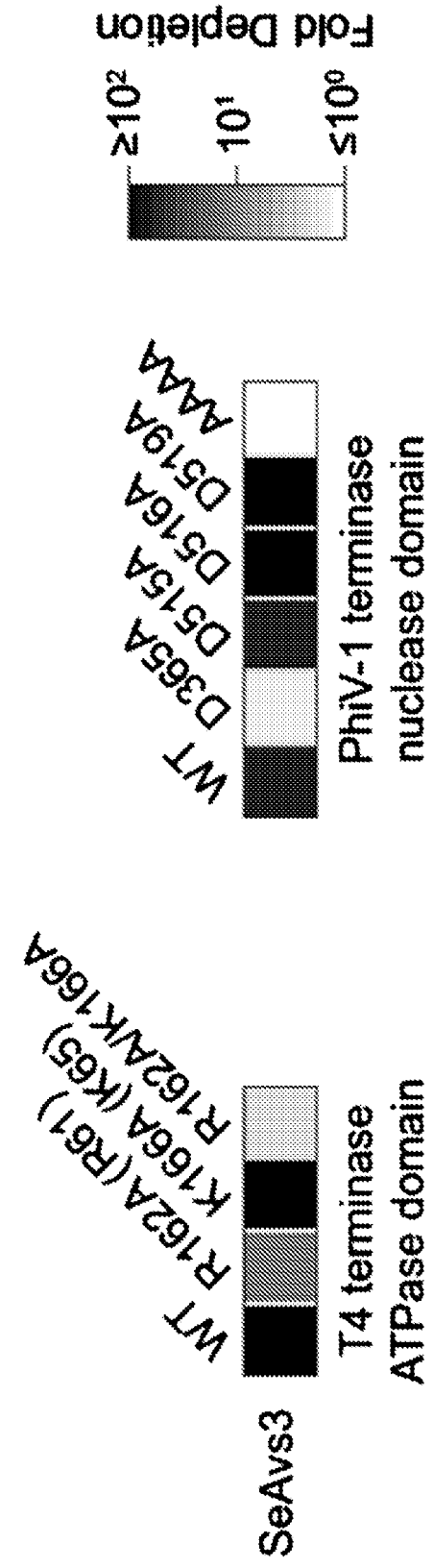

Fold Depletion
≥10²
10¹
≤10⁰
PhiV-1 terminase nuclease domain
WT
D365A
D515A
D516A
D519A
AAAA
T4 terminase ATPase domain
WT
R162A (R61)
K166A (K65)
R162A/K166A
SeAvs3

FIG. 28F

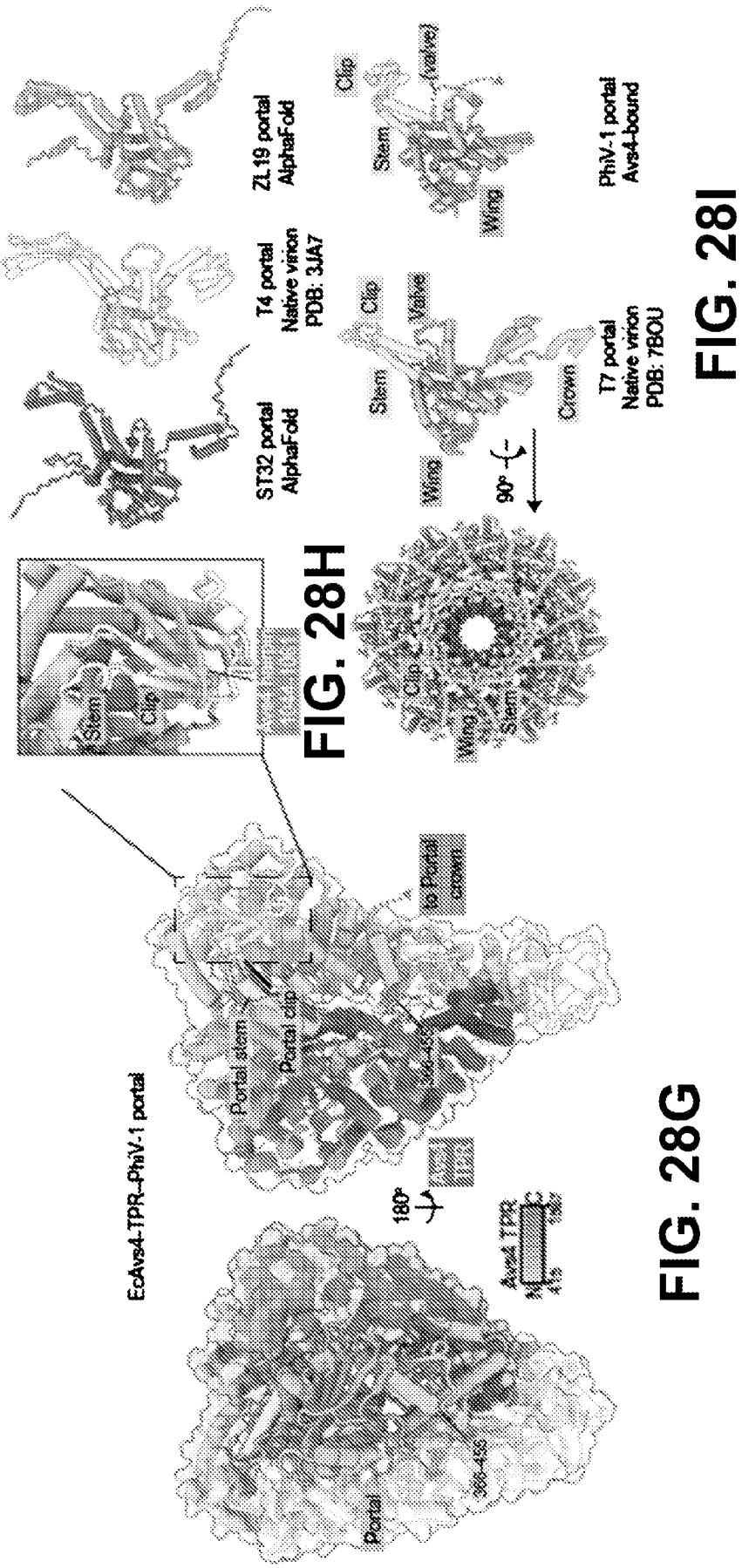
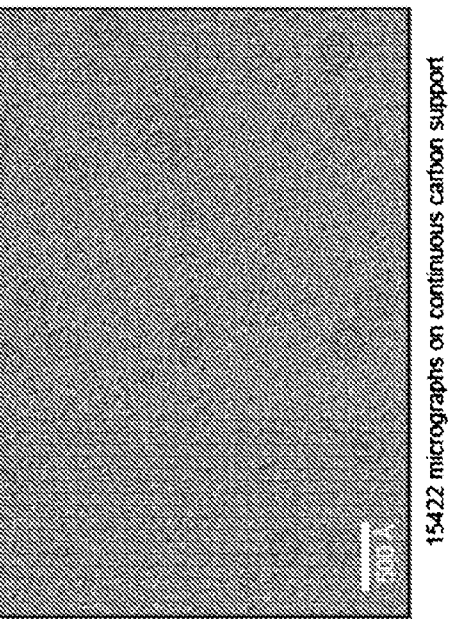


FIG. 28G

FIG. 28H

FIG. 28I



15422 micrographs on continuous carbon support Krios/K3

FIG. 29A

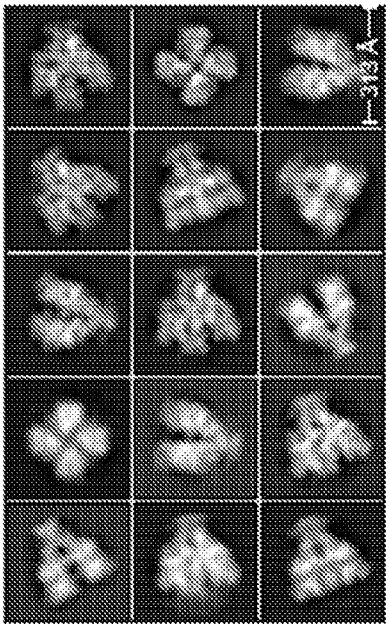
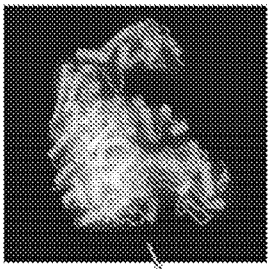
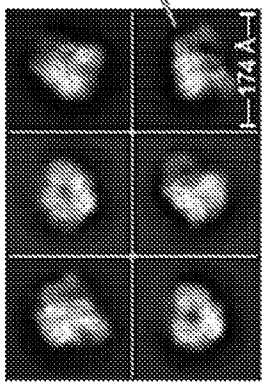


FIG. 29B





Apo SeAvs3, 5100 micrographs, Krios/K3

FIG. 29C

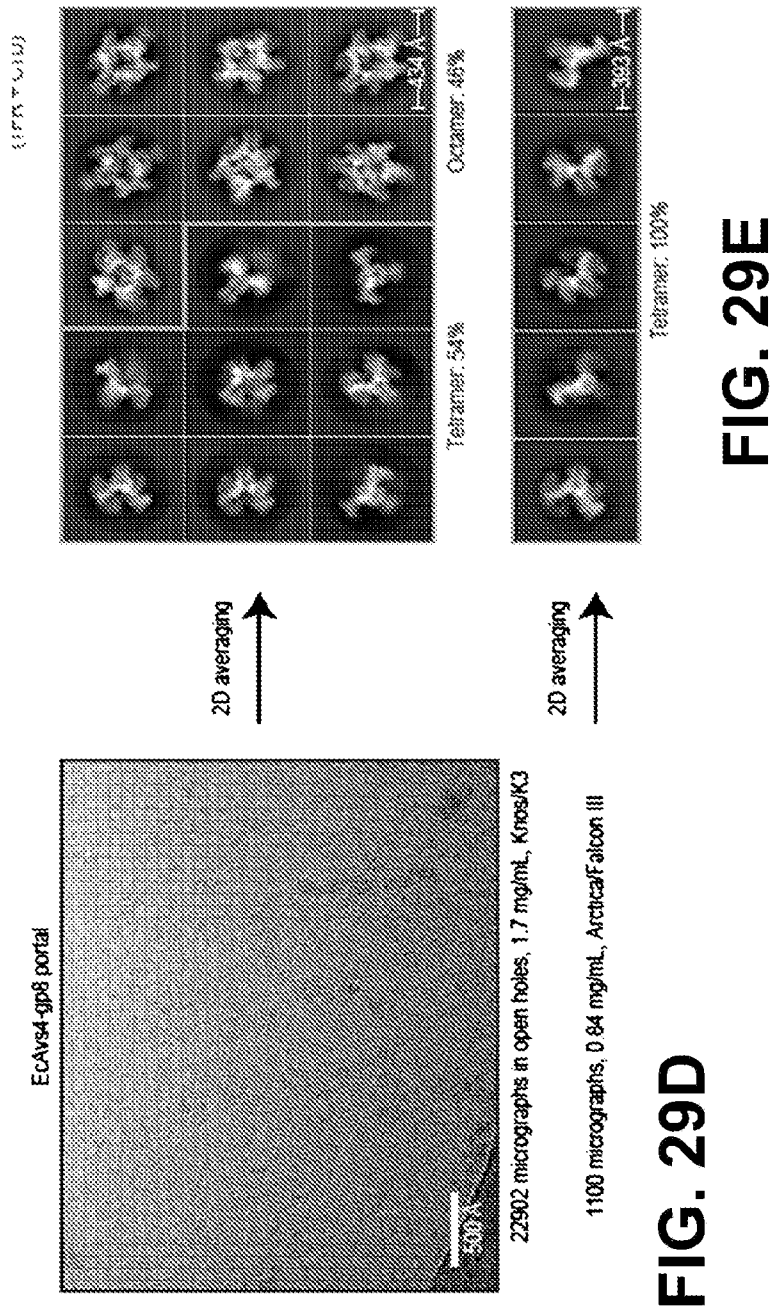


FIG. 29D

FIG. 29E

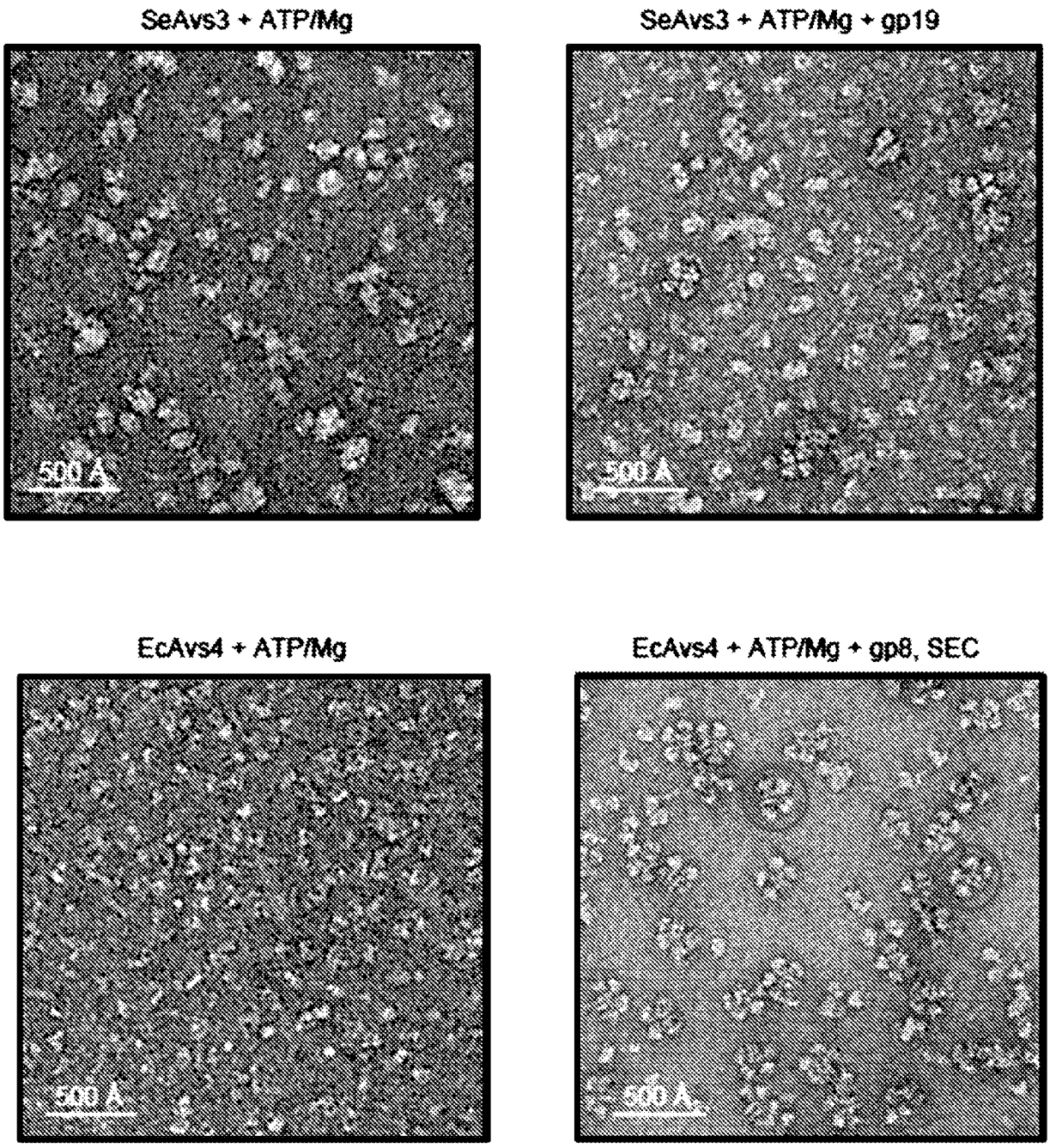

SeAvs3 + ATP/Mg
500 Å
SeAvs3 + ATP/Mg + gp19
500 Å
EcAvs4 + ATP/Mg
500 Å
EcAvs4 + ATP/Mg + gp8, SEC
500 Å

FIG. 29F

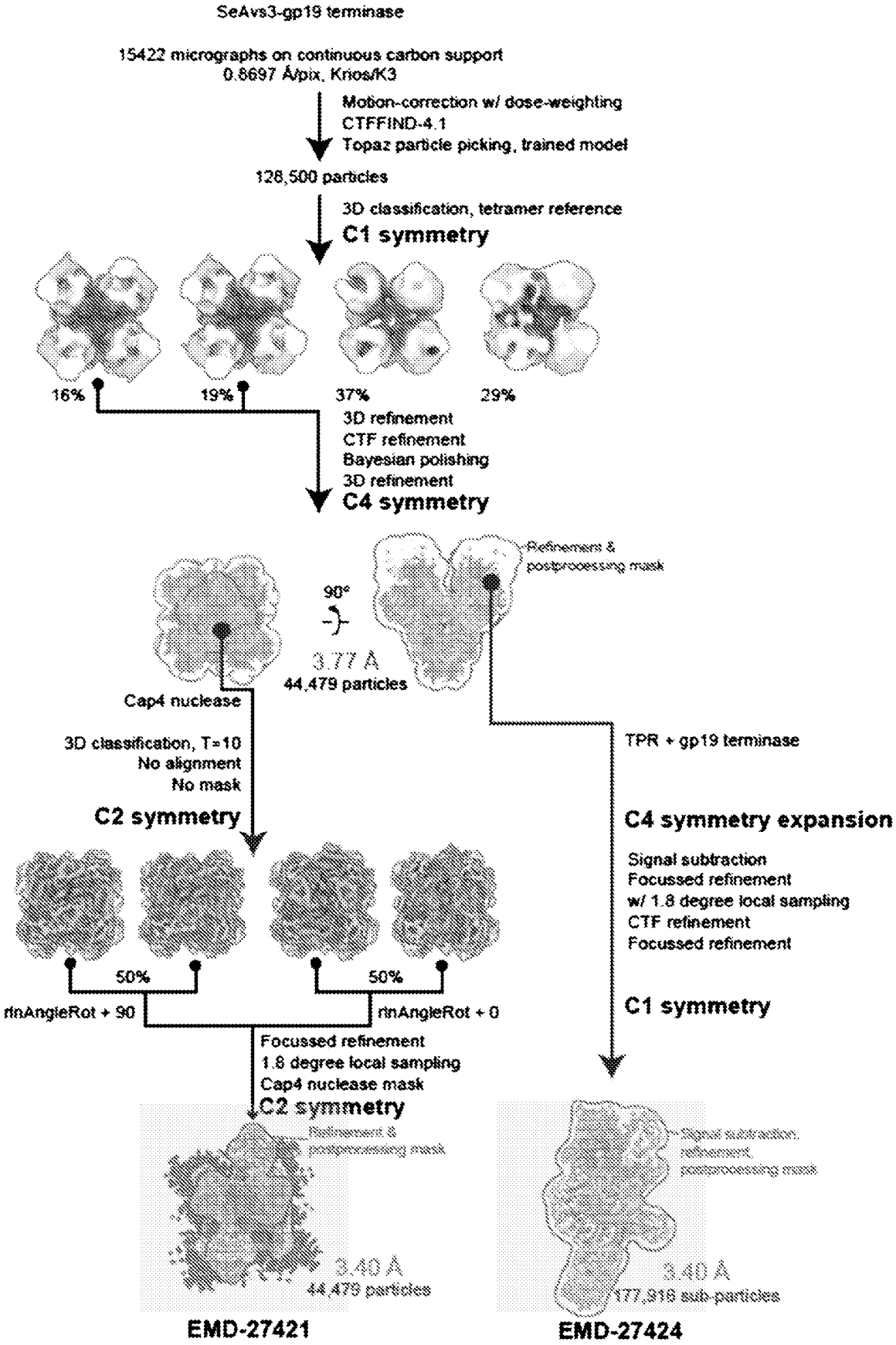

SeAvs3-gp19 terminase
15422 micrographs on continuous carbon support
0.8697 Å/pix, Krios/K3
Motion-correction w/ dose-weighting
CTFFIND-4.1
Topaz particle picking, trained model
128,500 particles
3D classification, tetramer reference
C1 symmetry
16%
19%
37%
29%
3D refinement
CTF refinement
Bayesian polishing
3D refinement
C4 symmetry
Refinement & postprocessing mask
90°
3.77 Å
44,479 particles
Cap4 nuclease
3D classification, T=10
No alignment
No mask
C2 symmetry
TPR + gp19 terminase
C4 symmetry expansion
Signal subtraction
Focussed refinement
w/ 1.8 degree local sampling
CTF refinement
Focussed refinement
C1 symmetry
50%
50%
rlnAngleRot + 90
rlnAngleRot + 0
Focussed refinement
1.8 degree local sampling
Cap4 nuclease mask
C2 symmetry
Refinement & postprocessing mask
3.40 Å
44,479 particles
Signal subtraction, refinement, postprocessing mask
3.40 Å
177,916 sub-particles
EMD-27421
EMD-27424

FIG. 30A

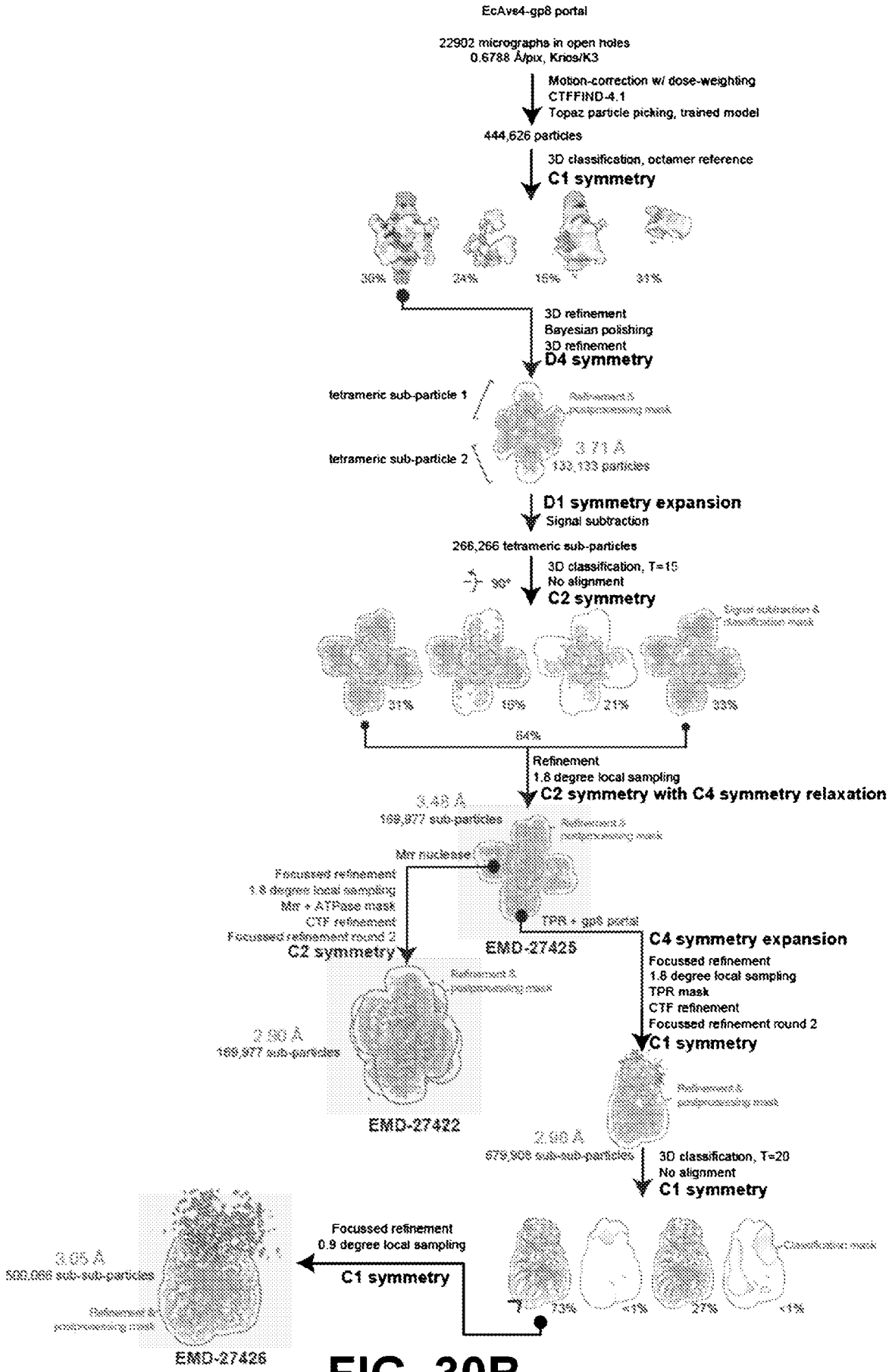

EcAvs4-gp8 portal
22902 micrographs in open holes
0.6788 Å/pix, Krios/K3
Motion-correction w/ dose-weighting
CTFFIND-4.1
Topaz particle picking, trained model
444,626 particles
3D classification, octamer reference
C1 symmetry
36%
24%
15%
31%
3D refinement
Bayesian polishing
3D refinement
D4 symmetry
tetrameric sub-particle 1
tetrameric sub-particle 2
3.71 Å
133,133 particles
D1 symmetry expansion
Signal subtraction
266,266 tetrameric sub-particles
3D classification, T=15
No alignment
C2 symmetry
31%
15%
21%
33%
64%
Refinement
1.8 degree local sampling
C2 symmetry with C4 symmetry relaxation
3.43 Å
168,977 sub-particles
Mtr nuclease
TPR + gp8 portal
EMD-27425
Focussed refinement
1.8 degree local sampling
Mtr + ATPase mask
CTF refinement
Focussed refinement round 2
C2 symmetry
2.80 Å
168,977 sub-particles
EMD-27422
C4 symmetry expansion
Focussed refinement
1.8 degree local sampling
TPR mask
CTF refinement
Focussed refinement round 2
C1 symmetry
2.66 Å
679,908 sub-sub-particles
3D classification, T=20
No alignment
C1 symmetry
73%
<1%
27%
<1%
Focussed refinement
0.9 degree local sampling
C1 symmetry
3.05 Å
500,066 sub-sub-particles
EMD-27426

FIG. 30B

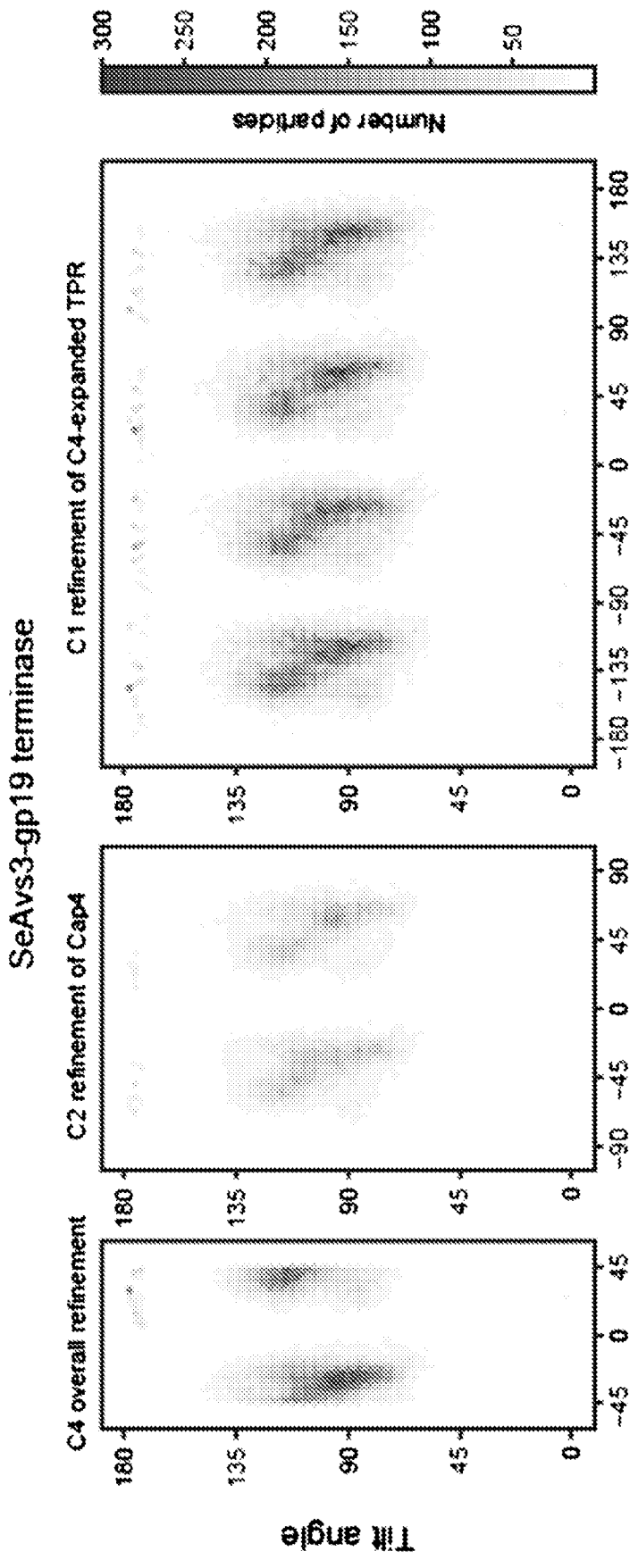

SeAvs3-gp19 terminase
C4 overall refinement
C2 refinement of Cap4
C1 refinement of C4-expanded TPR
Tilt angle
Rotation angle
Number of particles
300
250
200
150
100
50
180
135
90
45
0
-45
-90
-135
-180

FIG. 31A

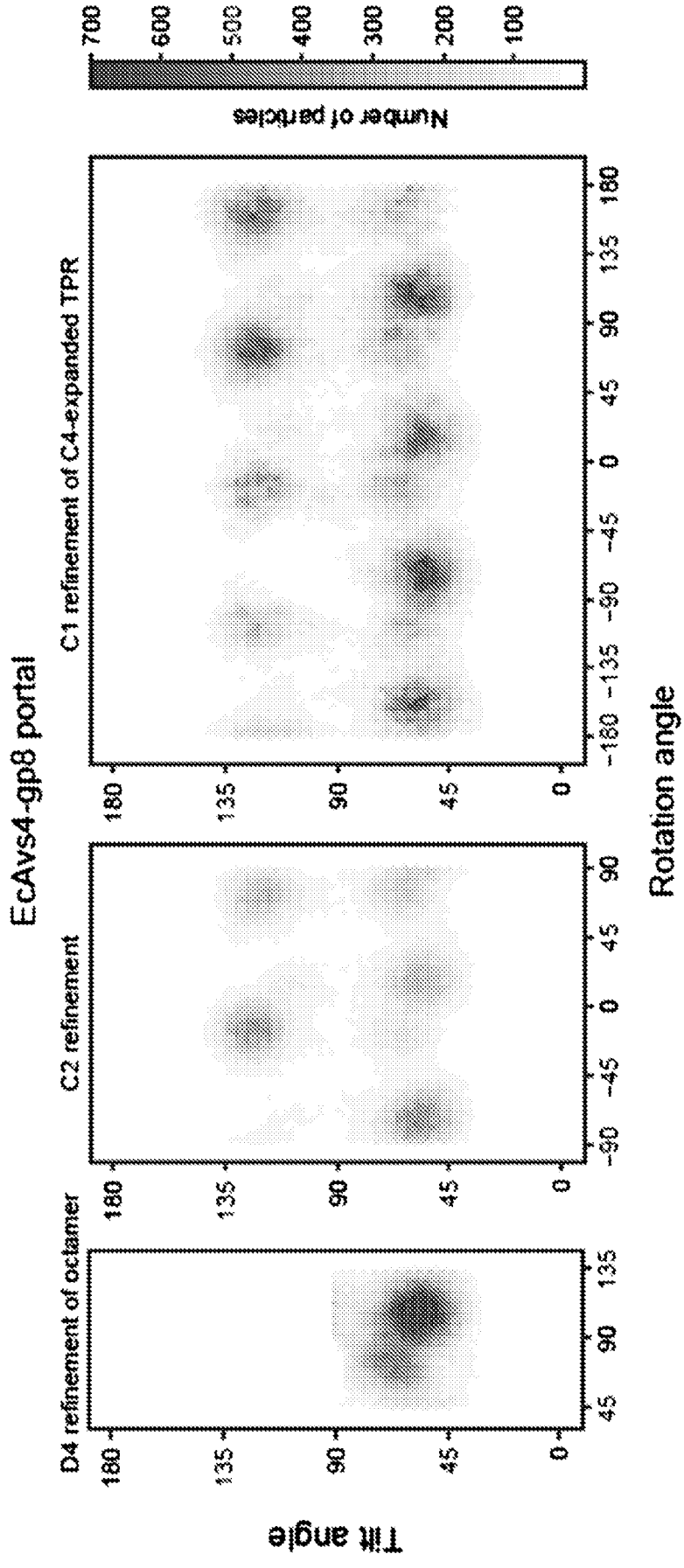

EcAvs4-gp8 portal
D4 refinement of octamer
C2 refinement
C1 refinement of C4-expanded TPR
Tilt angle
Rotation angle
Number of particles
700
600
500
400
300
200
100
180
135
90
45
0
-180
-135
-90
-45
0
45
90
135
180

FIG. 31B

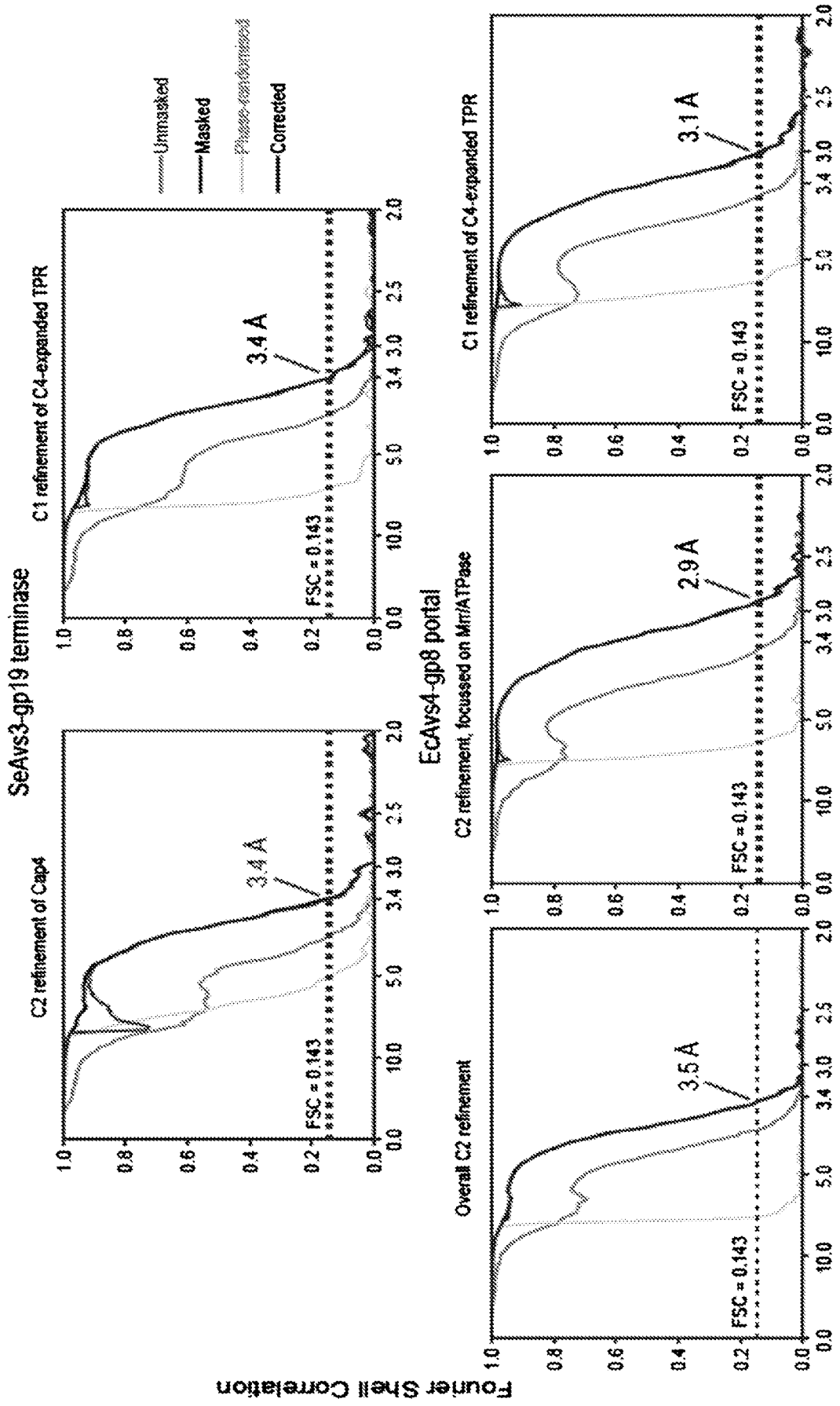

SeAvs3-gp19 terminase
EcAvs4-gp8 portal
C2 refinement of Cap4
C1 refinement of C4-expanded TPR
C2 refinement, focussed on MrrATPase
C1 refinement of C4-expanded TPR
Overall C2 refinement
3.4 Å
3.4 Å
2.9 Å
3.1 Å
3.5 Å
FSC = 0.143
Unmasked
Masked
Phase-randomised
Corrected
Fourier Shell Correlation
Resolution (Å)

FIG. 31C

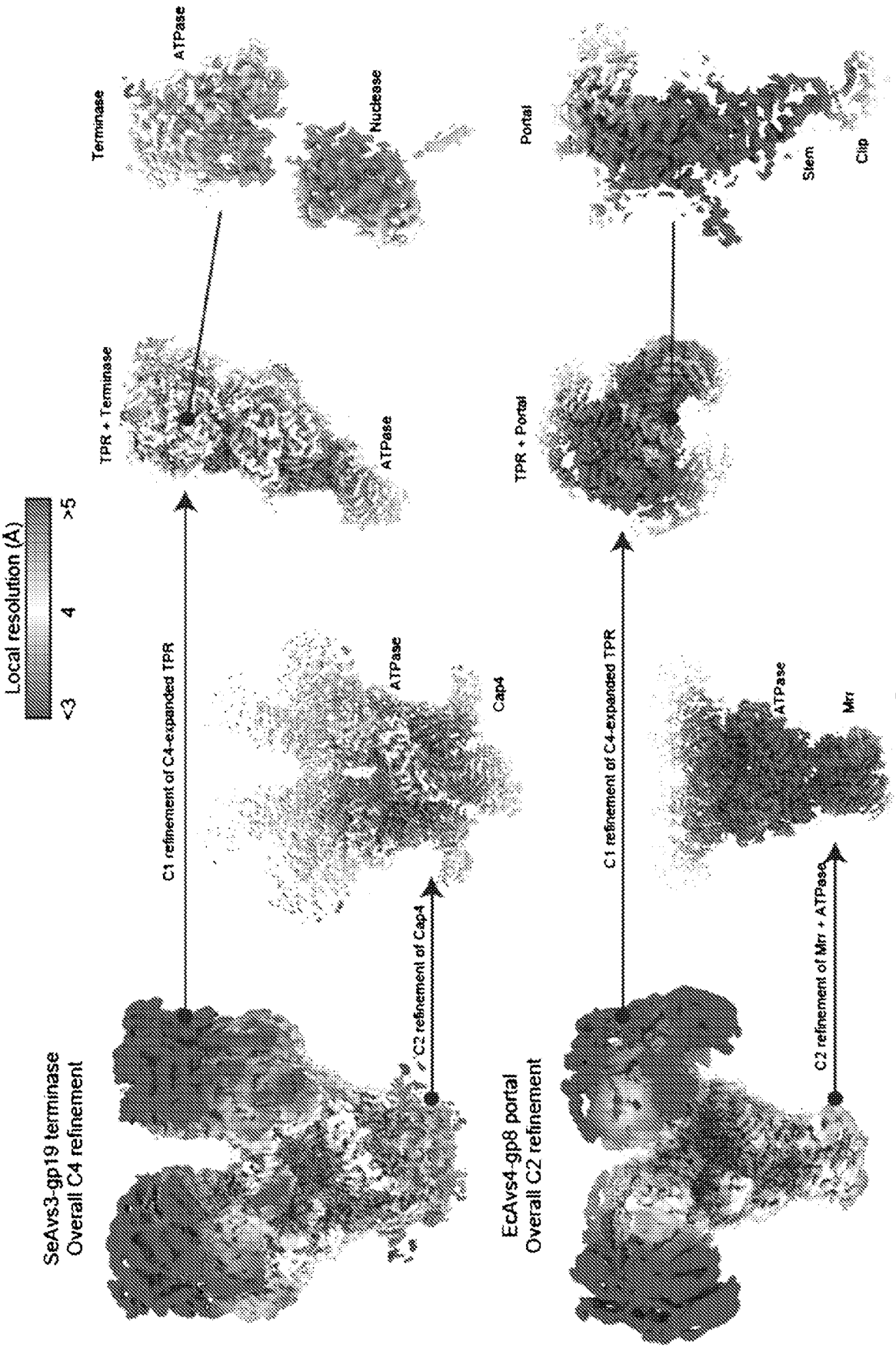

SeAvs3-gp19 terminase
Overall C4 refinement
C1 refinement of C4-expanded TPR
C2 refinement of Cap4
Local resolution (Å)
<3
4
>5
Terminase
ATPase
Nuclease
TPR + Terminase
ATPase
ATPase
Cap4
EcAvs4-gp8 portal
Overall C2 refinement
C1 refinement of C4-expanded TPR
C2 refinement of Mrr + ATPase
Portal
Stem
Clip
TPR + Portal
ATPase
Mrr

FIG. 32A

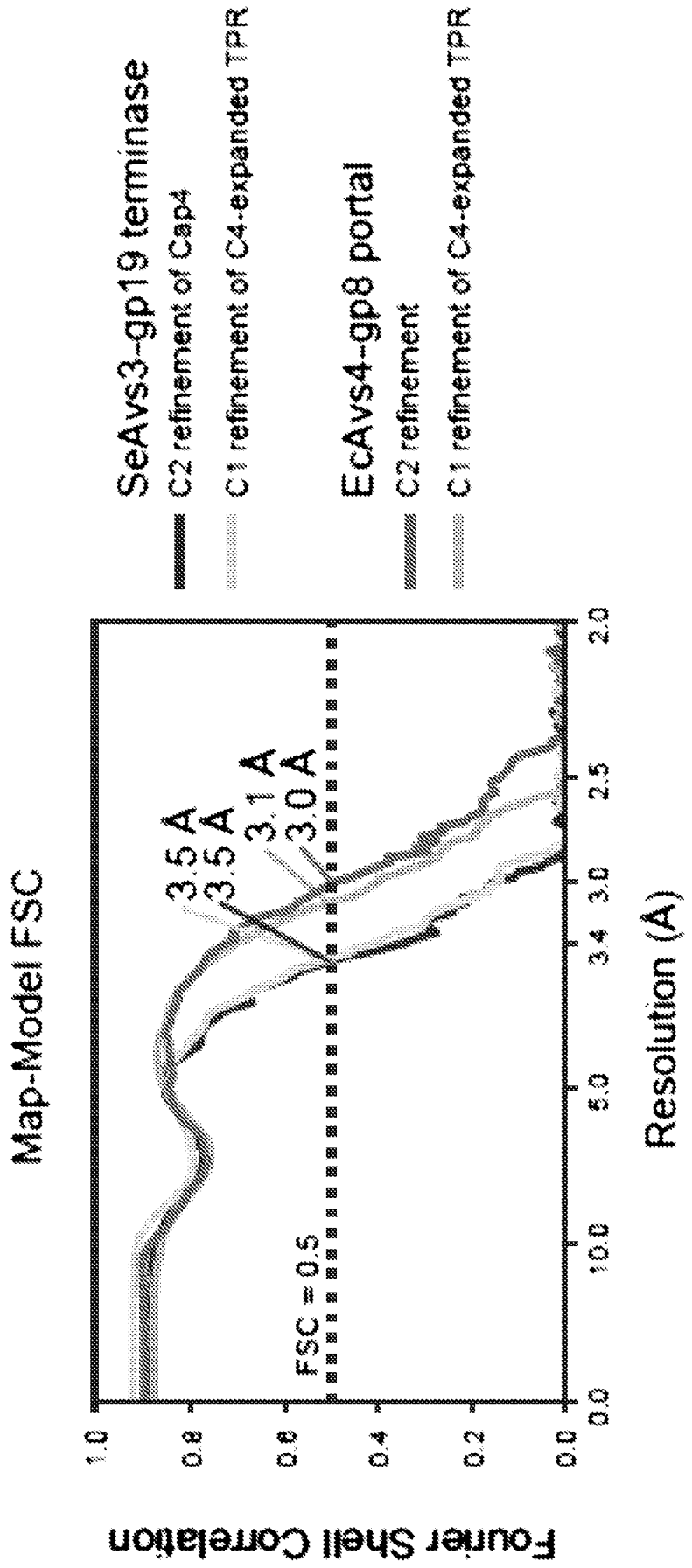

SeAvs3-gp19 terminase
C2 refinement of Cap4
C1 refinement of C4-expanded TPR
EcAvs4-gp8 portal
C2 refinement
C1 refinement of C4-expanded TPR
Map-Model FSC
Fourier Shell Correlation
Resolution (Å)
FSC = 0.5
3.5 Å
3.5 Å
3.1 Å
3.0 Å
1.0
0.8
0.6
0.4
0.2
0.0
0.0
10.0
6.0
3.4
3.0
2.5
2.0

FIG. 32B

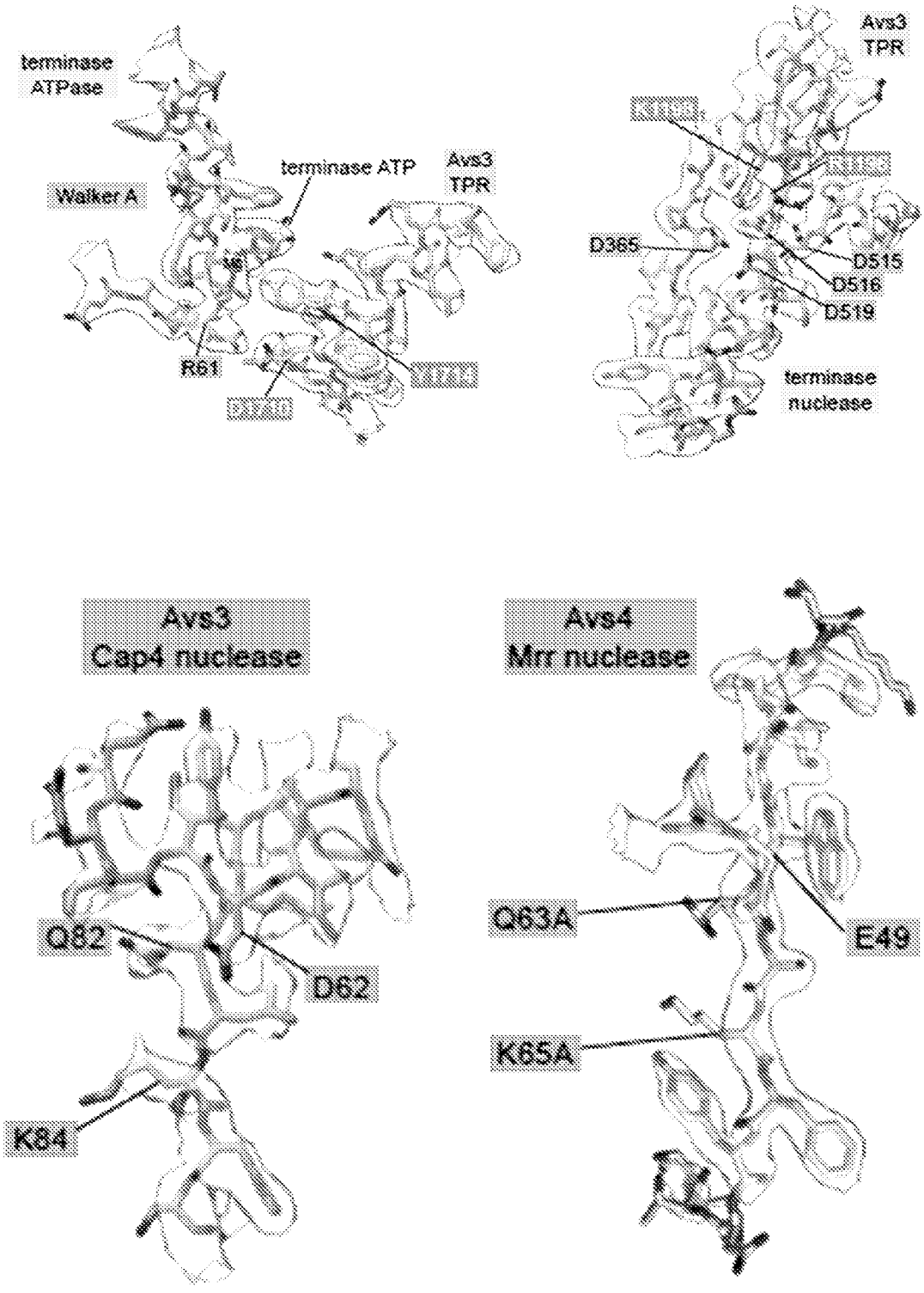

terminase ATPase
Walker A
terminase ATP
Avs3 TPR
R61
D365
Avs3 TPR
D365
D515
D516
D519
terminase nuclease
Avs3 Cap4 nuclease
Q82
D62
K84
Avs4 Mrr nuclease
Q63A
E49
K65A

FIG. 32C

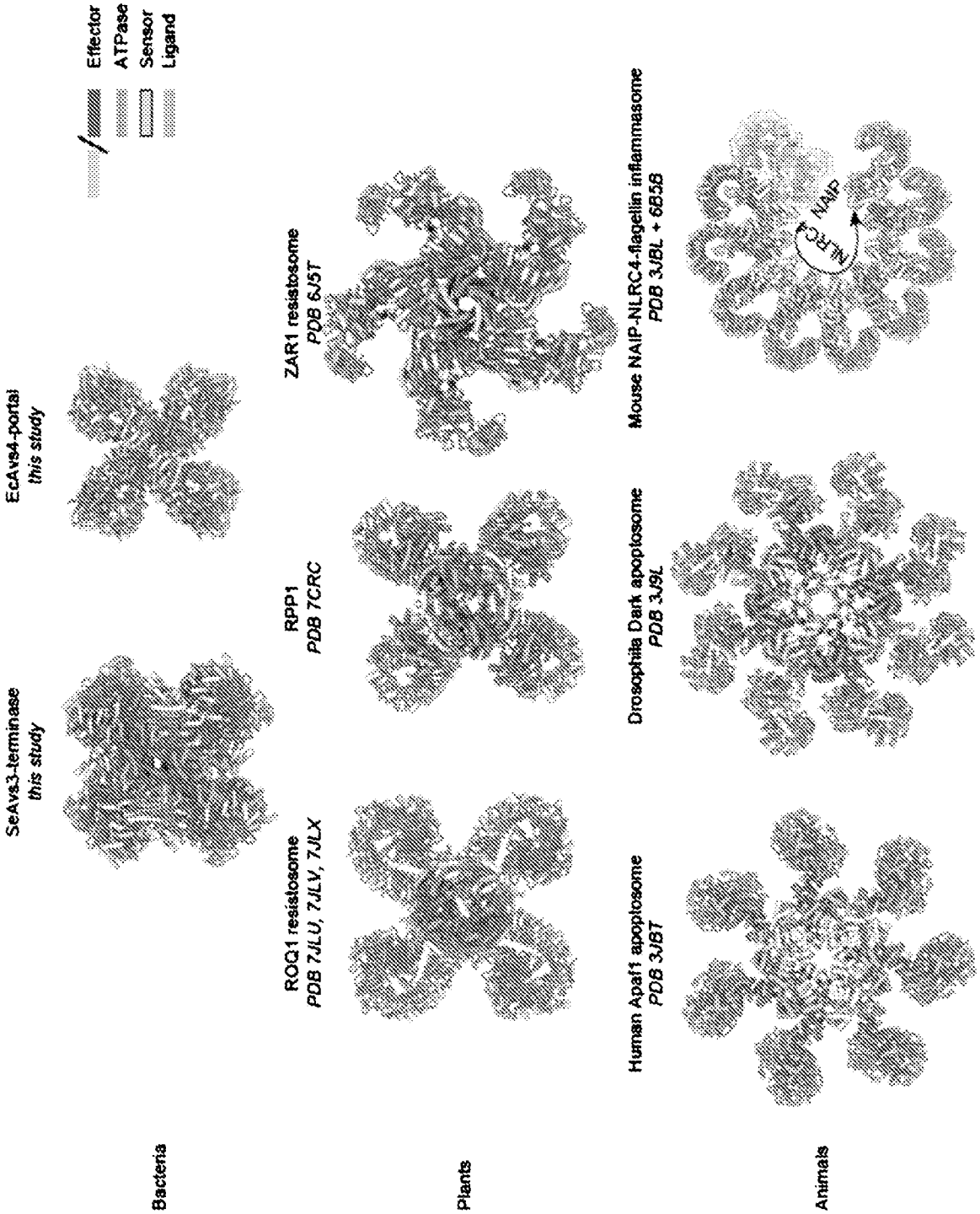

Effector
ATPase
Sensor
Ligand
Bacteria
SeAvs3-terminase
this study
EcAvs4-portal
this study
Plants
ROQ1 resistosome
PDB 7JLU, 7JLV, 7JLX
RPP1
PDB 7CRC
ZAR1 resistosome
PDB 6J5T
Animals
Human Apaf1 apoptosome
PDB 3J6T
Drosophila Dark apoptosome
PDB 3J9L
Mouse NAIP-NLRC4-flagellin inflammasome
PDB 3JBL + 6B5B
NAIP
NLRC4

FIG. 33A

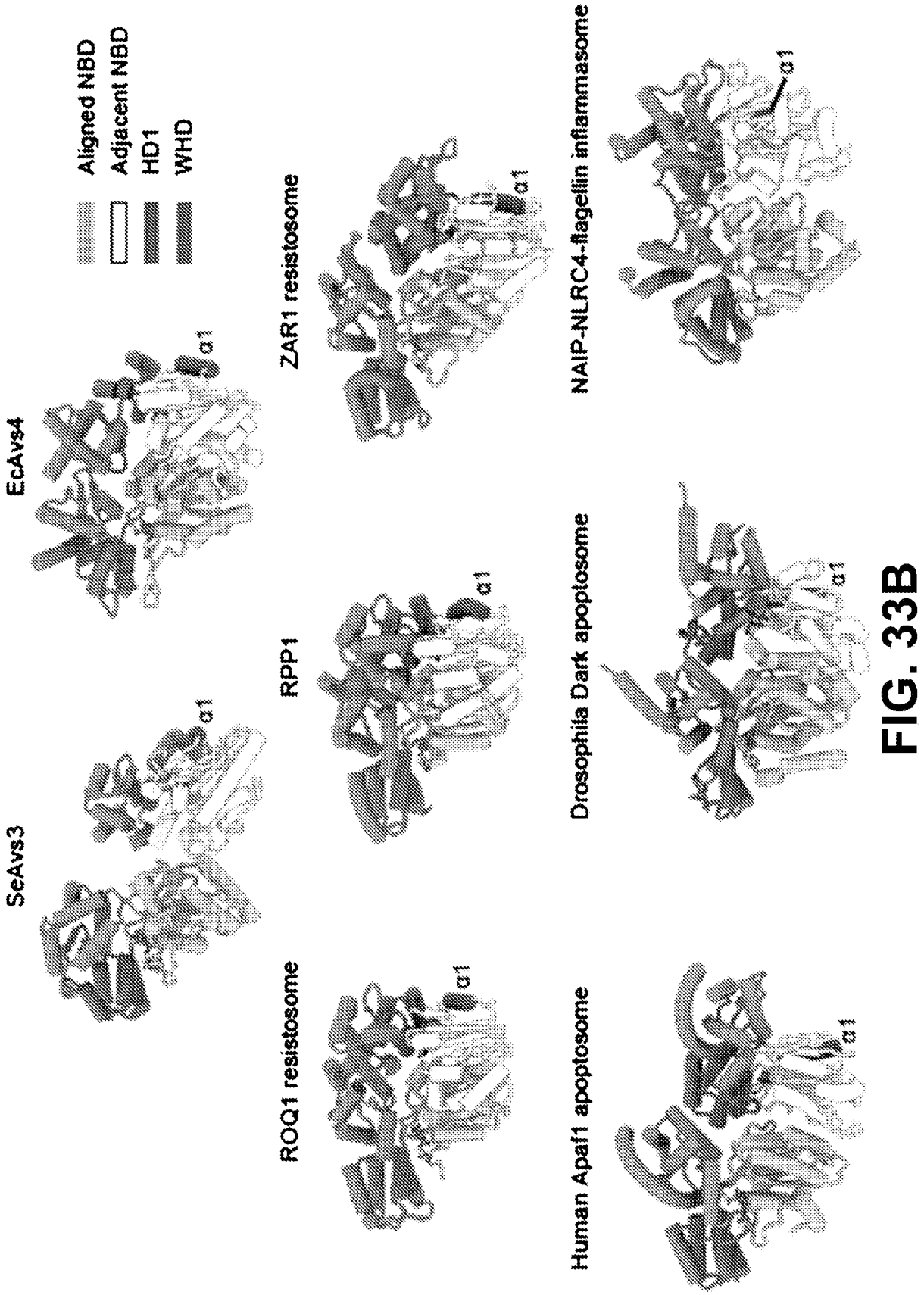
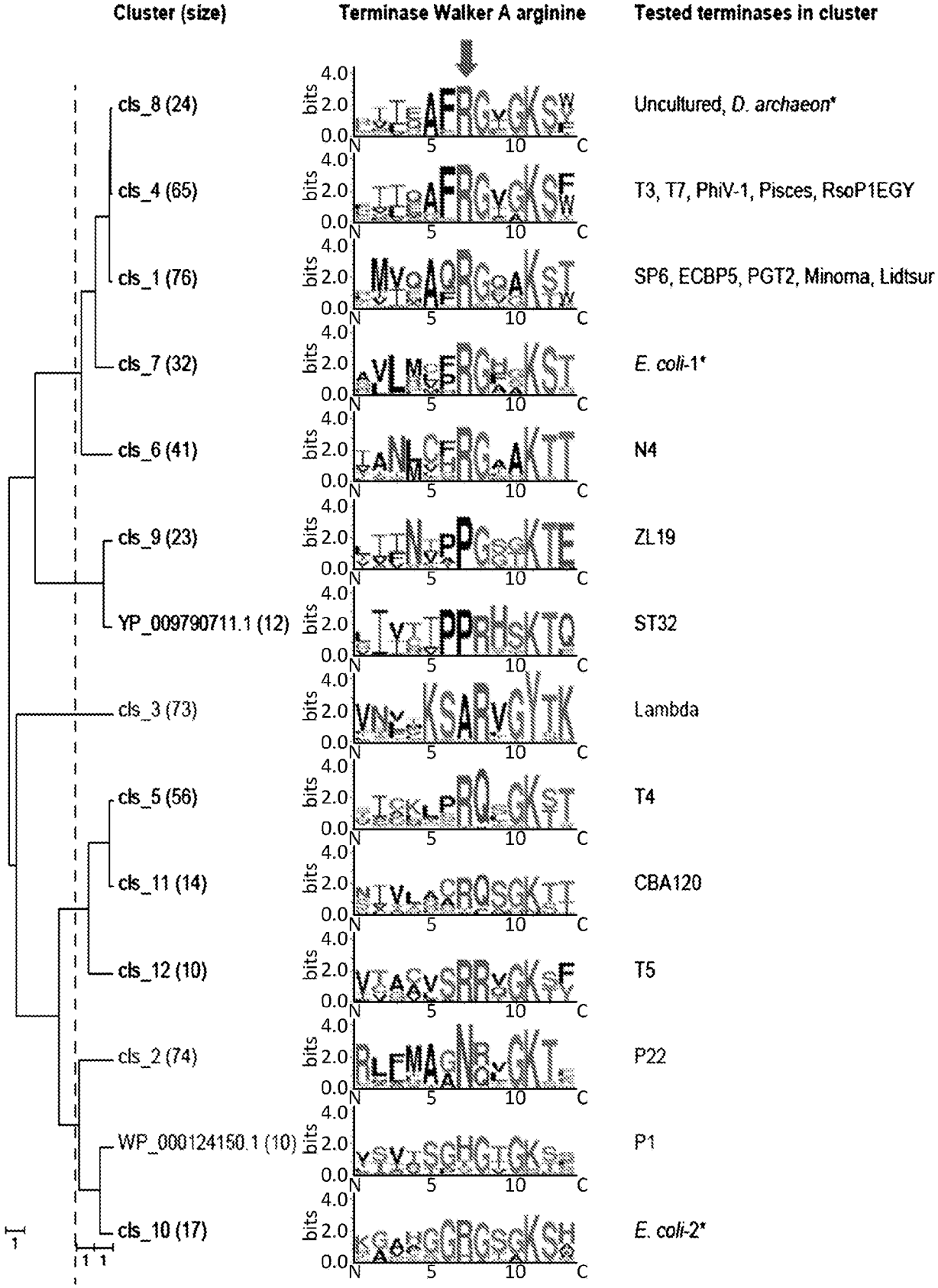

Aligned NBD
Adjacent NBD
HD1
WHD
EcAvs4
α1
SeAvs3
α1
ROQ1 resistosome
α1
RPP1
α1
ZAR1 resistosome
α1
Drosophila Dark apoptosome
α1
Human Apaf1 apoptosome
α1
NAIP-NLRC4-flagellin inflammasome
α1

FIG. 33B



Red – clusters with terminase proteins that do not activate SeAvs3

FIG. 34

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2023/071227

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - INV. - C07K 19/00; C12N 1/20, 15/62, 9/14, 9/22 (2023.01)
ADD. - C07K 14/195, 14/435; G01N 33/50, 33/52 (2023.01)

CPC - INV. - C07K 19/00; C12N 1/20, 15/62, 9/14, 9/22 (2023.08)

ADD. - C07K 14/195, 14/435, 2319/61; C12Y 306/01015; G01N 33/5005, 33/52 (2023.08)

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 database consulted during the international search (name of database 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 --- Y	US 2021/0130833 A1 (THE BROAD INSTITUTE INC. et al.) 06 May 2021 (06.05.2021) entire document	1-15, 17, 26, 27 --- 16, 18, 24, 25, 28, 29
Y	US 2022/0098250 A1 (UNIVERSITY OF WASHINGTON) 31 March 2022 (31.03.2022) entire document	16, 18, 24, 25, 28, 29
A	WO 2004/050870 A2 (LUDWIG-MAXIMILIANS-UNIVERSITÄT) 17 June 2004 (17.06.2004) entire document	1-18, 24-29
P, X	GAO et al., "Prokaryotic innate immunity through pattern recognition of conserved viral proteins," Science, 12 August 2022 (12.08.2022), Vol. 377, Iss. 6607: eabm4096, Pgs. 1-15. entire document	1-18, 24-29

☐ 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
19 October 2023	NOV 22 2023
Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Taina Matos Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2023/071227

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.: 19-23
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

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

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

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

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

Remark on Protest

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

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

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2023/071227

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

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

   a. ☒ forming part of the international application as filed.

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

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

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

3. Additional comments: