



(12) **DEMANDE DE BREVET CANADIEN**
CANADIAN PATENT APPLICATION

(13) **A1**

(86) Date de dépôt PCT/PCT Filing Date: 2019/08/09
(87) Date publication PCT/PCT Publication Date: 2020/02/13
(85) Entrée phase nationale/National Entry: 2021/02/08
(86) N° demande PCT/PCT Application No.: IB 2019/000946
(87) N° publication PCT/PCT Publication No.: 2020/030984
(30) Priorités/Priorities: 2018/08/09 (KR10-2018-0093336);
2018/10/30 (US62/752,950)

(51) Cl.Int./Int.Cl. *C12N 9/22* (2006.01),
A61K 38/00 (2006.01), *A61K 48/00* (2006.01),
A61P 35/00 (2006.01), *C12N 15/52* (2006.01)
(71) Demandeurs/Applicants:
G+FLAS LIFE SCIENCES, KR;
SEOUL NATIONAL UNIVERSITY R&DB FOUNDATION,
KR
(72) Inventeurs/Inventors:
CHOE, SUNGHWAN, KR;
PARK, JONGJIN, KP;
YOON, JIYOUNG, KR;
KIM, HAN SEONG, KR;
KIM, DONG WOOK, KR
(74) Agent: GOWLING WLG (CANADA) LLP

(54) Titre : COMPOSITIONS ET PROCEDES DE MODIFICATION DU GENOME AVEC DES PROTEINES CAS12A
(54) Title: COMPOSITIONS AND METHODS FOR GENOME ENGINEERING WITH CAS12A PROTEINS

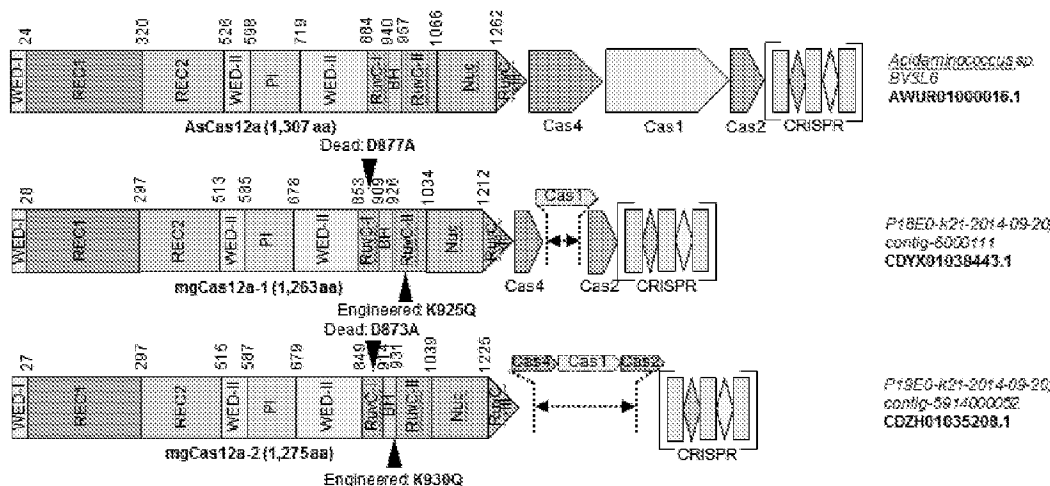


FIG. 20C

(57) **Abrégé/Abstract:**

The present disclosure provides novel Cas12a proteins, which cleave target nucleic acids and methods of use thereof.

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

(19) World Intellectual Property

Organization

International Bureau

(43) International Publication Date

13 February 2020 (13.02.2020)



(10) International Publication Number

WO 2020/030984 A3

(51) International Patent Classification:

C12N 9/22 (2006.01) A61P 35/00 (2006.01)

C12N 15/52 (2006.01) A61K 38/00 (2006.01)

A61K 48/00 (2006.01)

(21) International Application Number:

PCT/IB2019/000946

(22) International Filing Date:

09 August 2019 (09.08.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

10-2018-0093336 09 August 2018 (09.08.2018) KR

62/752,950 30 October 2018 (30.10.2018) US

(71) Applicants: **G+FLAS LIFE SCIENCES** [KR/KR]; 1F, 38 Nakseongdae-ro, Seoul, Gwanak-gu 08790 (KR). **SEOUL NATIONAL UNIVERSITY R&DB FOUNDATION** [KR/KR]; Seoul National University, Seoul, Gwanak-gu 08826 (KR).

(72) Inventors: **CHOE, Sunghwa**; 1F, 38 Nakseongdae-ro, Seoul, Gwanak-gu 08790 (KR). **PARK, Jongjin**; 1F, 38 Nakseongdae-ro, Seoul, Gwanak-gu 08790 (KR). **YOON, Jiyoung**; 1F, 38 Nakseongdae-ro, Seoul, Gwanak-gu 08790 (KR). **KIM, Han Seong**; 1F, 38 Nakseongdae-ro, Seoul, Gwanak-gu 08790 (KR). **KIM, Dong, Wook**; 1F, 38 Nakseongdae-ro, Seoul, Gwanak-gu 08790 (KR).

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

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

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))

(88) Date of publication of the international search report:

16 July 2020 (16.07.2020)

(54) Title: COMPOSITIONS AND METHODS FOR GENOME ENGINEERING WITH CAS12A PROTEINS

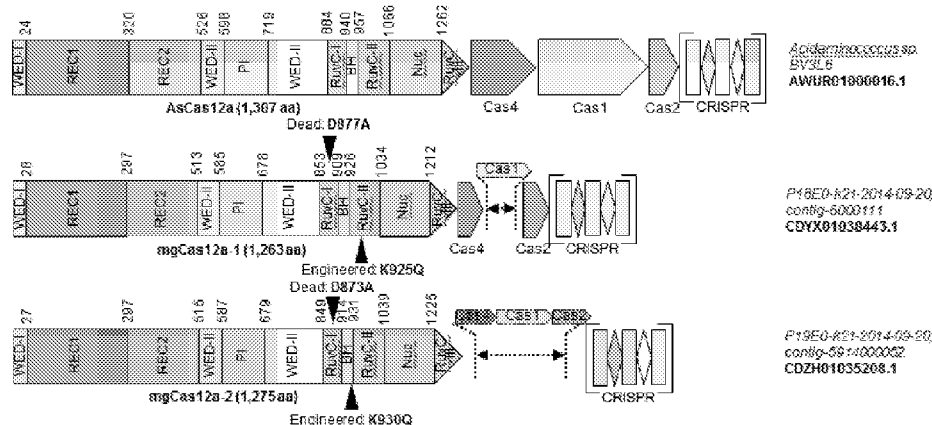


FIG. 20C

(57) Abstract: The present disclosure provides novel Cas12a proteins, which cleave target nucleic acids and methods of use thereof.



WO 2020/030984 A3

COMPOSITIONS AND METHODS FOR GENOME ENGINEERING WITH CAS12A PROTEINS

CROSS-REFERENCE

[0001] The present application claims priority to and benefit from Korean Patent Application 10-2018-0093336 filed August 9, 2018 and U.S. Provisional Application No. 62/752,950 filed October 30, 2018, the entire contents of each of which are herein incorporated by reference.

BACKGROUND

[0002] Genome editing is a genetic engineering technique that targets the genetic insert into a specific location of the genome, thereby modifying defective genes and/or introduction of functional genes. Recent advancement in Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-CRISPR-associated (Cas) technology using RNA guided endonuclease to cleave a nucleic acid sequence of interest further increases the efficiency and shortens the genome editing process. However, current CRISPR technology using the traditional endonuclease suffers from relatively low precision. Thus, there is still a significant need for novel, highly efficient endonucleases.

SUMMARY

[0003] The present disclosure provides a composition comprising at least one of i) - iv), and a guide RNA coupled the at least one of i) - iv): i) a polypeptide having at least 80% sequence identity with residue 829 through residue 991 of SEQ ID NO: 1, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 925 of SEQ ID NO: 1; ii) a polypeptide having at least 80% sequence identity with residue 825 through residue 996 of SEQ ID NO: 3, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 930 of SEQ ID NO: 3; iii) a polypeptide having at least 80% sequence identity with SEQ ID NO: 1, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 925 of SEQ ID NO: 1; or iv) a polypeptide having at least 80% sequence identity with SEQ ID NO: 3, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 930 of SEQ ID NO: 3. The guide RNA comprises a sequence that is reverse complementary to a eukaryotic nucleic acid sequence comprising from 6 to 60 bases.

[0004] In some embodiments, the a Lysine aligned to a Lysine at position 925 of SEQ ID NO: 1 or a Lysine aligned to a Lysine at position 930 of SEQ ID NO: 3 may not at positions 925 or 930, respectively, relative to SEQ ID NO: 1 or SEQ ID NO: 1 when optimally aligned.

[0005] In some embodiments, the eukaryotic nucleic acid sequence is a human nucleic acid sequence, which may comprise a region in KRAS, HER2/neu, PD-1, TCR, p53, CCR5, DNMT1, EMX1, and LKB1. Alternatively and/or additionally, the eukaryotic nucleic acid sequence is a plant nucleic acid sequence.

[0006] Preferably, the polypeptide is a type V CRISPR-associated protein. In such embodiment, it is preferred that the type V CRISPR-associated protein is a Cas12a protein.

[0007] In some embodiments, the polypeptide comprises a purification tag. The purification tag comprises at least one tag selected from the group consisting of a His tag, a FLAG tag, an AU1 epitope tag, an AU5 epitope tag, a bacteriophage T7 tag, a bacteriophage V5 epitope tag, a Bluetongue virus tag (B-tag), a Glu-Glu tag (EE-tag), an HSV epitope tag, a KT3 epitope tag, a Myc epitope tag, a PDZ ligand tag, a polyarginine tag, a polyaspartate tag, a polycysteine tag, a polyphenylalanine tag, a protein C tag, an S1-tag, an S-tag, a Step-tag, and a VSV-G tag.

[0008] In some embodiments, the guide RNA comprises an A-rich protospacer adjacent motif (PAM) sequence, a G-rich PAM sequence, a T-rich PAM sequence, or a C-rich PAM sequence. Optionally, the PAM sequence comprises 3 T nucleobases, 2 T nucleobases, 1 T nucleobase, or TTTN. Alternatively and/or additionally, guide RNA comprises a crRNA and a tracrRNA.

[0009] It is preferred that the composition exhibits at least 2-fold increased genome editing efficiency than AsCas12a, FnCas12a, or LbCas12a. In some embodiments, the composition further comprises an excipient. In some embodiments, the excipient has a pH of from 7 to 8. In some embodiments, the excipient is a buffer, which may comprise at least one of Bis-Tris Propane-HCl, MgCl₂, or bovine serum albumin.

[0010] The present disclosure also disclose a method of gene editing, comprising: providing a composition comprising at least one of i) - iv), and a guide RNA coupled the at least one of i) - iv): i) a polypeptide having at least 80% sequence identity with residue 829 through residue 991 of SEQ ID NO: 1, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 925 of SEQ ID NO: 1; ii) a polypeptide having at least 80% sequence identity with residue 825 through residue 996 of SEQ ID NO: 3, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 930 of SEQ ID NO: 3; iii) a polypeptide having at least 80% sequence identity with SEQ ID NO: 1, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 925 of SEQ ID NO: 1; or iv) a polypeptide having at least 80% sequence identity with SEQ ID NO: 3, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 930 of SEQ ID NO: 3. The guide RNA comprises a sequence that is reverse complementary to a eukaryotic nucleic acid sequence comprising from 6 to 60 bases.

[0011] In some embodiments, the eukaryotic nucleic acid sequence is a human nucleic acid sequence, which may comprise a region in KRAS, HER2/neu, PD-1, TCR, p53, CCR5, DNMT1, EMX1, and LKB1. Alternatively and/or additionally, the eukaryotic nucleic acid sequence is a plant nucleic acid sequence.

[0012] Preferably, the polypeptide is a type V CRISPR-associated protein. In such embodiment, it is preferred that the type V CRISPR-associated protein is a Cas12a protein.

[0013] In some embodiments, the polypeptide comprises a purification tag. The purification tag comprises at least one tag selected from the group consisting of a His tag, a FLAG tag, an AU1 epitope tag, an AU5 epitope tag, a bacteriophage T7 tag, a bacteriophage V5 epitope tag, a Bluetongue virus tag (B-tag), a Glu-Glu tag (EE-tag), an HSV epitope tag, a KT3 epitope tag, a Myc epitope tag, a PDZ ligand tag, a polyarginine tag, a polyaspartate tag, a polycysteine tag, a polyphenylalanine tag, a protein C tag, an S1-tag, an S-tag, a Step-tag, and a VSV-G tag.

[0014] In some embodiments, the guide RNA comprises an A-rich protospacer adjacent motif (PAM) sequence, a G-rich PAM sequence, a T-rich PAM sequence, or a C-rich PAM sequence. Optionally, the PAM sequence comprises 3 T nucleobases, 2 T nucleobases, 1 T nucleobase, or TTTN. Alternatively and/or additionally, guide RNA comprises a crRNA and a tracrRNA.

[0015] In some embodiments, the composition further comprises an excipient. In some embodiments, the excipient has a pH of from 7 to 8. In some embodiments, the excipient is a buffer, which may comprise at least one of Bis-Tris Propane-HCl, MgCl₂, or bovine serum albumin. It is preferred that an efficiency of the step of cleaving is at least 2-fold higher than genome editing efficiency of AsCas12a, FnCas12a, or LbCas12a.

[0016] The present disclosure also discloses a method of improving cleaving efficiency of a type V CRISPR-associated protein, comprising: providing the type V CRISPR-associated protein; identifying a residue of the type V CRISPR-associated protein that is aligned with at Lysine at position 925 of SEQ ID NO: 1 or at Lysine at position 930 of SEQ ID NO: 3; and mutating the residue to Lysine.

[0017] The present disclosure provides a composition comprising: a protein (or a polypeptide) having at least 80% sequence identity with residue 829 through residue 991 of SEQ ID NO: 1, wherein the protein comprises a Lysine at position 925; and a guide RNA comprising a sequence that is reverse complementary to a eukaryotic nucleic acid sequence comprising from 6 to 60 bases.

[0018] The present disclosure provides a composition comprising: a protein (or a polypeptide) having at least 80% sequence identity with residue 825 through residue 996 of SEQ ID NO: 3, wherein the protein comprises a Lysine at position 930; and a guide RNA comprising a sequence that is reverse complementary to a eukaryotic nucleic acid sequence comprising from 6 to 60 bases.

[0019] The present disclosure provides a composition comprising: a protein (or a polypeptide) having at least 80% sequence identity with SEQ ID NO: 1, wherein the protein comprises a Lysine at position 925; and a guide RNA comprising a sequence that is reverse complementary to a eukaryotic nucleic acid sequence comprising from 6 to 60 bases.

[0020] The present disclosure provides a composition comprising: a protein (or a polypeptide) having at least 80% sequence identity with SEQ ID NO: 3, wherein the protein comprises a Lysine at position 930; and a guide RNA comprising a sequence that is reverse complementary to a eukaryotic nucleic acid sequence comprising from 6 to 60 bases.

[0021] In these compositions, the eukaryotic nucleic acid sequence can be human nucleic acid sequence. Optionally, the human nucleic acid sequence is implicated in cancer. Alternatively, the eukaryotic nucleic acid sequence can be a plant nucleic acid sequence. In some embodiments, the nucleotide sequence encoding for SEQ ID NO: 1 comprises at least 80% sequence identity to SEQ ID NO: 2. Alternatively, and/or additionally, the nucleotide sequence encoding for SEQ ID NO: 3 comprises at least 80% sequence identity to SEQ ID NO: 4. It is contemplated that the protein is from, obtained from, or derived from the *Eubacteriaceae* family. In particular, disclosures provided herein, the protein comprises a nuclease.

[0022] Preferably, the protein comprises a nuclease. In such embodiments, the nuclease may comprise a type V CRISPR-associated protein. Further preferably, the type V CRISPR-associated protein may comprise a Cas12a protein. In some embodiments, such Cas12a protein is metagenomically mined.

[0023] In some embodiments, the composition comprises, or has pH ranged from 7 to 7.9, such as a pH of about or exactly 7. Alternatively, and/or additionally, the composition is formulated in a buffer, which may comprise Bis-Tris Propane-HCl, MgCl₂, and/or bovine serum albumin. For example, the buffer comprises from 0.1 to 50 mM Bis-Tris Propane-HCl. Additionally, the buffer comprises from 0.1 to 50 mM MgCl₂. The buffer may also comprise from 1 to 500 µg/ml bovine serum albumin. The buffer may, in particular, comprise about or exactly 10 mM Bis-Tris Propane-HCl. Further, the buffer may comprises about or exactly 10 mM MgCl₂. Additionally, the buffer comprises about or exactly 100 µg /ml of bovine serum albumin.

[0024] In some embodiments, the protein (or a polypeptide; protein and polypeptide can be used interchangeably herein) comprises a purification tag. For example, the purification tag comprises, or may include, at least one tag selected from the group consisting of a His tag, a FLAG tag, an AU1 epitope tag, an AU5 epitope tag, a bacteriophage T7 tag, a bacteriophage V5 epitope tag, a Bluetongue virus tag (B-tag), a Glu-Glu tag (EE-tag), an HSV epitope tag, a KT3 epitope tag, a Myc epitope tag, a PDZ ligand tag, a polyarginine tag, a polyaspartate tag, a polycysteine tag, a

polyphenylalanine tag, a protein C tag, an S1-tag, an S-tag, a Step-tag, and a VSV-G tag. It is consistent with the present disclosure for the guide RNA to comprise an A-rich protospacer adjacent motif (PAM) sequence, a G-rich PAM sequence, a T-rich PAM sequence, or a C-rich PAM sequence.

[0025] The guide RNA in the composition may comprise a T-rich PAM sequence. For example, the PAM sequence comprises 3 T nucleobases, 2 T nucleobases, or 1 T nucleobase. A PAM sequence compatible with the present disclosure comprises TTTN. In some embodiments, the guide RNA sequence comprises from 1 to 100 nucleotides.

[0026] Alternatively, and/or additionally, the target human nucleic acid sequence comprises a region in KRAS, HER2/neu, PD-1, TCR, p53, CCR5, DNMT1, EMX1, and LKB1.

[0027] Several cancers are contemplated and compatible with the present disclosure, for example, the cancer comprises a bladder cancer, a bone cancer, a blood cancer, a breast cancer, a black colored tumor, a thyroid cancer, a parathyroid cancer, a bone marrow cancer, a laryngopharyngeal cancer, a laryngeal cancer, a lung cancer, an esophagus cancer, a pancreatic cancer, a colorectal cancer, a gastric cancer, a tongue cancer, a skin cancer, a brain tumor, a uterine cancer, a head or neck cancer, a gallbladder cancer, an oral cancer, a central nervous system tumor, or a liver cancer.

[0028] The present compositions are capable of exhibiting improved genome editing efficiency to other Cas12a orthologs. For example, in some aspects, the composition exhibits at least 2-fold increased genome editing efficiency than AsCas12a, FnCas12a, or LbCas12a. The presently disclosed proteins are capable of being complexed with a crRNA having a 5' handle compatible with other Cas12a orthologs. For example, in any of the above compositions, the guide RNA comprises a crRNA and a tracrRNA. Further, the crRNA comprises a 5' repeat recognition sequence of AAUU. In some aspects, the protein exhibits cleavage activity in the presence of CaCl₂, CoCl₂, FeCl₂, MnSO₄, or any combination thereof.

[0029] The present disclosure also provides a method of gene editing, wherein the method comprises contacting a cell with any one of the above compositions; binding the guide RNA to the target human nucleic acid sequence; and cleaving the human nucleic acid sequence.

[0030] The present disclosure additionally provides a method of gene editing, wherein the method comprises providing a composition comprising a protein (or a polypeptide) having at least 80% sequence identity with residue 829 through residue 991 of SEQ ID NO: 1, wherein the protein comprises a Lysine at position 925; and a guide RNA comprising a sequence that is reverse complementary to a nucleic acid sequence comprising from 6 to 60 nucleotides; contacting a cell with the composition; binding the guide RNA to a eukaryotic nucleic acid sequence; and cleaving the nucleic acid sequence.

[0031] The present disclosure provides a method of gene editing, wherein the method comprises providing a composition comprising a protein (or a polypeptide) having at least 80% sequence identity with residue 825 through residue 996 of SEQ ID NO: 3, wherein the protein comprises a Lysine at position 930; and a guide RNA comprising a sequence that is reverse complementary to a nucleic acid sequence comprising from 6 to 60 nucleotides; contacting a cell with the composition; binding the guide RNA to a eukaryotic nucleic acid sequence; and cleaving the nucleic acid sequence.

[0032] The present disclosure additionally provides a method of gene editing, wherein the method comprises providing a composition comprising a protein (or a polypeptide) having at least 80% sequence identity with SEQ ID NO: 1, wherein the protein comprises a Lysine at position 925; and a guide RNA comprising a sequence that is reverse complementary to a nucleic acid sequence comprising from 6 to 60 nucleotides; contacting a cell with the composition; binding the guide RNA to a eukaryotic nucleic acid sequence; and cleaving the nucleic acid sequence.

[0033] The present disclosure also provides a method of gene editing, wherein the method comprises providing a composition comprising a protein (or a polypeptide) having at least 80% sequence identity with SEQ ID NO: 3, wherein the protein comprises a Lysine at position 930; and a guide RNA comprising a sequence that is reverse complementary to a nucleic acid sequence comprising from 6 to 60 nucleotides; contacting a cell with the composition; binding the guide RNA to a eukaryotic nucleic acid sequence; and cleaving the nucleic acid sequence.

[0034] In these compositions, the eukaryotic nucleic acid sequence can be human nucleic acid sequence. Optionally, the human nucleic acid sequence is implicated in cancer. Alternatively, the eukaryotic nucleic acid sequence can be a plant nucleic acid sequence. In some embodiments, the nucleotide sequence encoding for SEQ ID NO: 1 comprises at least 80% sequence identity to SEQ ID NO: 2. Alternatively, and/or additionally, the nucleotide sequence encoding for SEQ ID NO: 3 comprises at least 80% sequence identity to SEQ ID NO: 4. Oftentimes, it may be the case that the nucleic acid sequence comprises a human nucleic acid sequence or a plant nucleic acid sequence. The present disclosure provides that the contacting the cell with the composition comprises administering the composition to a subject in need thereof. For example, the administering comprises intravenous, subcutaneous, intramuscular, oral, or mucosal administration. In some aspects, the contacting the cell with the composition comprises administering the composition to the cell *ex vivo*.

[0035] It is contemplated that the protein (or a polypeptide) is from, obtained from, or derived from the *Eubacteriaceae* family. In particular disclosures provided herein, the protein comprises a nuclease. Preferably, the protein comprises a nuclease. In such embodiments, the nuclease may

comprise a type V CRISPR-associated protein. Further, preferably, the type V CRISPR-associated protein may comprise a Cas12a protein. In some embodiments, such Cas12a protein is metagenomically mined.

[0036] In some embodiments, the composition comprises, or has pH ranged from 7 to 7.9, such as a pH of about or exactly 7. Alternatively, and/or additionally, the composition is formulated in a buffer, which may comprise Bis-Tris Propane-HCl, MgCl₂, and/or bovine serum albumin. For example, the buffer comprises from 0.1 to 50 mM Bis-Tris Propane-HCl. The buffer also comprises from 0.1 to 50 mM MgCl₂. The buffer additionally comprises from 1 to 500 µg/ml bovine serum albumin. The buffer also comprises about or exactly 10 mM Bis-Tris Propane-HCl. Further, the buffer comprises about or exactly 10 mM MgCl₂. Additionally, the buffer comprises about or exactly 100 µg /ml of bovine serum albumin.

[0037] In some embodiments, the protein (or a polypeptide) comprises a purification tag. For example, the purification tag comprises, or may include, at least one tag selected from the group consisting of a His tag, a FLAG tag, an AU1 epitope tag, an AU5 epitope tag, a bacteriophage T7 tag, a bacteriophage V5 epitope tag, a Bluetongue virus tag (B-tag), a Glu-Glu tag (EE-tag), an HSV epitope tag, a KT3 epitope tag, a Myc epitope tag, a PDZ ligand tag, a polyarginine tag, a polyaspartate tag, a polycysteine tag, a polyphenylalanine tag, a protein C tag, an S1-tag, an S-tag, a Step-tag, and a VSV-G tag.

[0038] The guide RNA in the composition may comprise an A-rich protospacer adjacent motif (PAM) sequence, a G-rich PAM sequence, a T-rich PAM sequence, or a C-rich PAM sequence. The guide RNA in the composition may comprise a T-rich PAM sequence. For example, the PAM sequence comprises 3 T nucleobases, 2 T nucleobases, or 1 T nucleobase. A PAM sequence compatible with the present disclosure comprises TTTN. In some embodiments, the guide RNA sequence comprises from 1 to 100 bases. Alternatively, and/or additionally, the target human nucleic acid sequence comprises a region in KRAS, HER2/neu, PD-1, TCR, p53, CCR5, DNMT1, EMX1, and LKB1.

[0039] In some embodiments, the cell comprises cancer cells, preferably human cancer cells, including, but not limited to a bladder cancer cell, a bone cancer cell, a blood cancer cell, a breast cancer cell, a black colored tumor cell, a thyroid cancer cell, a parathyroid cancer cell, a bone marrow cancer cell, a laryngopharyngeal cancer cell, a laryngeal cancer cell, a lung cancer cell, an esophagus cancer cell, a pancreatic cancer cell, a colorectal cancer cell, a gastric cancer cell, a tongue cancer cell, a skin cancer cell, a brain tumor cell, a uterine cancer cell, a head or neck cancer cell, a gallbladder cancer cell, an oral cancer cell, a central nervous system tumor cell, or a liver cancer cell.

[0040] In some aspects, any of the above described methods results in an at least 2-fold increased genome editing efficiency than AsCas12a, FnCas12a, or LbCas12a. The presently disclosed methods further include a composition, wherein the guide RNA comprises a crRNA and a tracrRNA. Further, the crRNA comprises a 5' repeat recognition sequence of AAUU. Still further, in the presently disclosed methods the composition exhibit cleaving activity in the presence of CaCl₂, CoCl₂, FeCl₂, MnSO₄, or any combination thereof.

[0041] The present disclosure additionally provides a method of improving cleaving efficiency of a type V CRISPR-associated protein, the method comprising providing the type V CRISPR-associated protein; identifying a residue at position 925 or 930; and mutating the residue at position 925 or 930 to a Lysine, thereby improving cleaving efficiency of the type V CRISPR-associated protein.

INCORPORATION BY REFERENCE

[0042] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0043] The novel features of the invention are set forth with particularity in the appended claims. A better 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 are utilized, and the accompanying drawings of which:

[0044] FIG. 1 shows a process by which a CRISPR associated protein (CAS protein) of the present disclosure was identified by metagenomic mining.

[0045] FIG. 2 shows a dendrogram of Cas12a.

[0046] FIG. 3 shows an alignment of Cas12a proteins of the present disclosure, including SEQ ID NO: 1 (mgCas12a-1), SEQ ID NO: 3 (mgCas12a-2), SEQ ID NO: 9 (AsCas12a), SEQ ID NO: 10 (LbCas12a), and SEQ ID NO: 11 (FnCas12a).

[0047] FIG. 4 shows a continuation of the alignment of Cas12a proteins of the present disclosure, including SEQ ID NO: 1 (mgCas12a-1), SEQ ID NO: 3 (mgCas12a-2), SEQ ID NO: 9 (AsCas12a), SEQ ID NO: 10 (LbCas12a), and SEQ ID NO: 11 (FnCas12a) from **FIG. 3**.

[0048] FIG. 5 shows a continuation of the alignment of Cas12a proteins of the present disclosure, including SEQ ID NO: 1 (mgCas12a-1), SEQ ID NO: 3 (mgCas12a-2), SEQ ID NO: 9 (AsCas12a), SEQ ID NO: 10 (LbCas12a), and SEQ ID NO: 11 (FnCas12a) from **FIG. 4**.

[0049] **FIG. 6** shows a continuation of the alignment of Cas12a proteins of the present disclosure, including SEQ ID NO: 1 (mgCas12a-1), SEQ ID NO: 3 (mgCas12a-2), SEQ ID NO: 9 (AsCas12a), SEQ ID NO: 10 (LbCas12a), and SEQ ID NO: 11 (FnCas12a) from **FIG. 5**.

[0050] **FIG. 7** shows a continuation of the alignment of Cas12a proteins of the present disclosure, including SEQ ID NO: 1 (mgCas12a-1), SEQ ID NO: 3 (mgCas12a-2), SEQ ID NO: 9 (AsCas12a), SEQ ID NO: 10 (LbCas12a), and SEQ ID NO: 11 (FnCas12a) from **FIG. 6**.

[0051] **FIG. 8** shows a continuation of the alignment of Cas12a proteins of the present disclosure, including SEQ ID NO: 1 (mgCas12a-1), SEQ ID NO: 3 (mgCas12a-2), SEQ ID NO: 9 (AsCas12a), SEQ ID NO: 10 (LbCas12a), and SEQ ID NO: 11 (FnCas12a) from **FIG. 7**.

[0052] **FIG. 9A** shows a chart of characteristics of Cas12a proteins of the present disclosure, including Cas12a proteins discovered by metagenomics mining (*e.g.*, SEQ ID NO: 1 (mgCas12a-1) and SEQ ID NO: 3 (mgCas12a-2), AsCas12a, LbCas12a, and FnCas12a).

[0053] **FIG. 9B** shows a chart of amino acid sequence identities (%) between Cas12a orthologs. AsCas12 has less than 40% sequence identity to all other orthologs in the table. LbCas12a and FnCas12a have less than 40% sequence identity to mgCas12a-1 and mgCas12a-2. LbCas12a has between 40% and 50% sequence identity to FnCas12a. mgCas12a-1 has greater than 50% sequence identity to mgCas12a-2.

[0054] **FIG. 10** shows gel electrophoresis of target dsDNA #1 after exposure to mgCas12a-1 (SEQ ID NO: 1) or mgCas12a-2 (SEQ ID NO: 3) complexed with crRNA #1 in no buffer, NEBuffer 1.1, NEBuffer 2.1, or NEBuffer 3.1.

[0055] **FIG. 11** shows another gel electrophoresis of target dsDNA #2 after exposure to mgCas12a-1 (SEQ ID NO: 1) or mgCas12a-2 (SEQ ID NO: 3) complexed with crRNA #2 in no buffer, NEBuffer 1.1, NEBuffer 2.1, or NEBuffer 3.1.

[0056] **FIG. 12** shows gel electrophoresis of target dsDNA after exposure to mgCas12a-1 (SEQ ID NO: 1), mgCas12a-2 (SEQ ID NO: 3), AsCas12a, FnCas12a, and LbCas12a complexed with crRNA.

[0057] **FIG. 13** shows a diagram of the target double stranded DNA LsXTb12 tested in the gel electrophoresis experiments of **FIG. 10** to **FIG. 12** and the corresponding binding regions of crRNA.

[0058] **FIG. 14** shows the results of an *in vitro* cleavage assay using various nucleases including, FnCas12, AsCas12, LbCas12, He-MgCas12a-1 (humanized and engineered mgCas12a-1), and He-MgCas12a-2 (humanized and engineered mgCas12a-2), 1 hour after incubation of the target with the nucleases and 4 hours after incubation of the target with the nucleases.

[0059] FIGs. 15A-B illustrate genome editing efficiencies of FnCas12a and He-MgCas12a-1 in rice and *N.benthamiana*.

[0060] FIG. 15A illustrates that FnCas12a exhibited genome editing efficiencies in rice of 0.5%, 0.3%, and 0.9% in crRNA1-1, crRNA1-2, crRNA2, respectively and He-MgCas12a-1 exhibited genome editing efficiencies in rice of 1.9%, 0.7%, and 10.2% in crRNA1-1, crRNA1-2, crRNA2.

[0061] FIG. 15B illustrates that FnCas12a exhibited genome editing efficiencies in *N.benthamiana* of 0.8%, 1.4%, and 4.8% in crRNA1, crRNA2, and crRNA3, respectively and He-MgCas12a-1 exhibited genome editing efficiencies in *N.benthamiana* of 0.7%, 3.7%, and 3.4% in crRNA1, crRNA2, and crRNA3, respectively.

[0062] FIG. 16 shows an *in vitro* cleavage assay of Cas12a nucleases including mgCas12a-1 (wild-type), he_mgCas12a-1 (humanized and engineered mgCas12a-1), de_mgCas12a-1 (dead and engineered mgCas12a), mg Cas12a-2, he_mgCas12a-2, de_mgCas12a-2, AsCas12a, FnCas12a, and LbCas12a.

[0063] FIGs. 17A-B show an *in vitro* cleavage assay of Cas12a nucleases including mgCas12a-1 (wild-type), he_mgCas12a-1 (humanized and engineered mgCas12a-1), de_mgCas12a-1 (dead and engineered mgCas12a), mg Cas12a-2, he_mgCas12a-2, de_mgCas12a-2, AsCas12a, FnCas12a, and LbCas12a after reaction times of 12 h (**FIG. 17A**) and 24 h (**FIG. 17B**).

[0064] FIG. 18 shows an *in vitro* cleavage assay of target plasmid DNA with Cas12a nucleases including mgCas12a-1 (wild-type), he_mgCas12a-1 (humanized and engineered mgCas12a-1), de_mgCas12a-1 (dead and engineered mgCas12a), mg Cas12a-2, he_mgCas12a-2, de_mgCas12a-2, AsCas12a, FnCas12a, and LbCas12a.

[0065] FIGs. 19A-B shows an *in vitro* cleavage assay of target plasmid DNA with Cas12a nucleases including mgCas12a-1 (wild-type), he_mgCas12a-1 (humanized and engineered mgCas12a-1), de_mgCas12a-1 (dead and engineered mgCas12a), mg Cas12a-2, he_mgCas12a-2, de_mgCas12a-2, AsCas12a, FnCas12a, and LbCas12a after reaction times of 12 h (**FIG. 19A**) and 24 h (**FIG. 19B**).

[0066] FIGs. 20A-C show schematics of the mgCas12a proteins of the present disclosure.

[0067] FIG. 20A shows a condensed schematic relative to **FIG. 1**, showing the pipeline for mining Cas12a proteins from metagenome data.

[0068] FIG. 20B shows a phylogenetic tree of metagenome-derived Cas12a proteins of the disclosure and other Cas12a orthologs.

[0069] FIG. 20C shows schematics of functionally-characterized novel Cas12a's and AsCas12a (Yamano *et al.* 2016).

[0070] **FIG. 21** shows an unrooted and evolutionary distance-based phylogenetic tree of metagenome-derived Cas12a of the present disclosure and other orthologs.

[0071] **FIGs. 22A-B** show sequence-specific cleavage of dsDNA by crRNA guided-mgCas12a proteins of the present disclosure.

[0072] **FIG. 22A** shows sequence-specific cleavage of linear dsDNA by crRNA guided-Cas12a proteins including FnCas12a, WT mgCas12a-1 and WT mgCas12a-2.

[0073] **FIG. 22B** shows sequence-specific cleavage of circular dsDNA by crRNA guided-Cas12a proteins including FnCas12a, WT mgCas12a-1 and WT mgCas12a-2.

[0074] **FIGs. 23A-B** shows that the mgCas12a proteins of the present disclosure can utilize three different types of Cas12a handles.

[0075] **FIG. 23A** shows cleavage of target linear dsDNA by WT mgCas12a-1 complexed with a crRNA having a 5' handle from AsCas12a, FnCas12a, and LbCas12a.

[0076] **FIG. 23B** shows cleavage of target linear dsDNA by WT mgCas12a-2 complexed with a crRNA having a 5' handle from AsCas12a, FnCas12a, and LbCas12a.

[0077] **FIG. 24** shows that some Cas12a-RNPs exhibits uncontrolled dsDNase activity. Seven different Cas12a-RNPs were incubated with target dsDNA for 12 or 24 hours.

[0078] **FIGs. 25A-B** show that Cas12a exhibits random dsDNase activity.

[0079] **FIG. 25A** shows each Cas12a incubated with dsDNA for different time periods.

[0080] **FIG. 25B** shows a graph of time versus dsDNase activity of each Cas12a.

[0081] **FIGs. 26A-D** show the activity of each Cas12a-RNP in the presence of different divalent cation.

[0082] **FIG. 26A** shows the results from Cas12a-RNP cleavage of target, linear dsDNA in the presence of seven different divalent cations were given to each Cas12a-RNP.

[0083] **FIG. 26B** shows sequence-specific dsDNA cleavage of FnCas12a-RNP under presence of different divalent cations.

[0084] **FIG. 26C** shows sequence-specific dsDNA cleavage of each WT mgCas12a-1RNP under presence of different divalent cations.

[0085] **FIG. 26D** shows sequence-specific dsDNA cleavage of each WT mgCas12a-2-RNP under presence of different divalent cations.

[0086] **FIG. 27** shows a graph of editing efficiencies of mgCas12a-1, mgCas12a-2, and mock (negative control).

[0087] **FIG. 28** shows a graph of indel frequency targeting nuclease protein in *Nicotiana benthamiana*. The gene editing efficiency of mgCas12a-1 was twice that of FnCpf1.

DETAILED DESCRIPTION

Definitions

[0088] As used in the specification and appended claims, unless specified to the contrary, the following terms have the meaning indicated below.

[0089] As used herein, the term “comprise” or variations thereof such as “comprises” or “comprising” are to be read to indicate the inclusion of any recited feature but not the exclusion of any other features. Thus, as used herein, the term “comprising” is inclusive and does not exclude additional, unrecited features. In some embodiments of any of the compositions and methods provided herein, “comprising” may be replaced with “consisting essentially of” or “consisting of.” The phrase “consisting essentially of” is used herein to require the specified feature(s) as well as those which do not materially affect the character or function of the claimed disclosure. As used herein, the term “consisting” is used to indicate the presence of the recited feature alone.

[0090] Throughout this disclosure, various embodiments are presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of any embodiments. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well of any individual numerical values within that range to the tenth of the unit of the lower limit unless the context clearly dictates otherwise. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well of any individual values within that range, for example, 1.1, 2, 2.3, 5, and 5.9. This applies regardless of the breadth of the range. The upper and lower limits of these intervening ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention, unless the context clearly dictates otherwise.

[0091] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of any embodiment. As used herein, the singular forms “a,” “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms “comprises” and/or “comprising,” when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof. As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0092] Unless specifically stated or obvious from context, as used herein, the term “about” in reference to a number or range of numbers is understood to mean the stated number and numbers +/- 10% thereof, or 10% below the lower listed limit and 10% above the higher listed limit for the values listed for a range.

[0093] The terms “determining,” “measuring,” “evaluating,” “assessing,” “assaying,” and “analyzing” are often used interchangeably herein to refer to forms of measurement. The terms include determining if an element is present or not (for example, detection). These terms can include quantitative, qualitative or quantitative and qualitative determinations. Assessing can be relative or absolute. “Detecting the presence of” can include determining the amount of something present in addition to determining whether it is present or absent depending on the context.

[0094] As used herein, “treatment of” or “treating,” “applying”, or “palliating” or “ameliorating” are used interchangeably. These terms refer to an approach for obtaining beneficial or desired results including but not limited to therapeutic benefit and/or a prophylactic benefit. By “therapeutic benefit” is meant eradication or amelioration of the underlying disorder being treated. Also, a therapeutic benefit is achieved with the eradication or amelioration of one or more of the physiological symptoms associated with the underlying disorder such that an improvement is observed in the patient, notwithstanding that the patient is still afflicted with the underlying disorder. For prophylactic benefit, the compositions are, in some embodiments, administered to a patient at risk of developing a particular disease or condition, or to a patient reporting one or more of the physiological symptoms of a disease, even though a diagnosis of this disease has not been made.

[0095] The terms “subject,” “individual,” or “patient” are often used interchangeably herein. A “subject” can be a biological entity containing expressed genetic materials. The biological entity can be a plant, animal, or microorganism, including, for example, bacteria, viruses, fungi, and protozoa. The subject can be tissues, cells and their progeny of a biological entity obtained in vivo or cultured in vitro. The subject can be a mammal. The mammal can be a human. The subject may be diagnosed or suspected of being at high risk for a disease. In some cases, the subject is not necessarily diagnosed or suspected of being at high risk for the disease.

[0096] The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

[0097] The present disclosure provides novel endonuclease identified by metagenomics analysis. The identified novel endonuclease is preferentially DNA endonucleases, which may additionally cleave RNA. It is contemplated that the identified endonucleases can cleave double stranded DNA, single stranded DNA, or both double stranded DNA (dsDNA) and single stranded DNA (ssDNA).

Preferably, the identified endonuclease may include Class II, Type V CRISPR/Cas proteins, such as Cas12a endonucleases. Further, preferably, Cas12a proteins may comprise a RuvC-like domain and may lack the HNH domain found in Cas9.

[0098] The present disclosure also provides synthetic tools for genome editing comprising a clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system, which includes guide RNA (crRNA and/or tracrRNA) that directs a DNA endonuclease (Cas protein) to a region of double stranded DNA (dsDNA) to be cleaved. Guide RNAs (gRNA) are synthetically engineered and include the crRNA and tracrRNA, which are linked together with a linker to form gRNA. It is contemplated that the linkers can be of any length and with any combination of nucleobases, which may include G-rich linkers, A-rich linker, T-rich linkers, and C-rich linkers. Linkers of the present disclosure comprise 1 to 50 nucleobases, 5 to 50 nucleobases, 10 to 50 nucleobases, 15 to 50 nucleobases, 20 to 50 nucleobases, 25 to 50 nucleobases, 30 to 50 nucleobases, 35 to 50 nucleobases, 40 to 50 nucleobases, or 45 to 50 nucleobases. The present disclosure alternatively provides a linker having a sequence of GAAA to form gRNA. The present disclosure provides any guide RNA-directed endonuclease. Cas proteins of the present disclosure include any nuclease selected from Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10, Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas13a, Cas13b, Cas13c, Cas13d, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3 and Csf4. The present disclosure also provides one or more Cas proteins that are, in particular, Cas12a proteins, which is also referred to as Cpf1. Cas12a proteins do not require a tracrRNA and, thus, are complexed with a crRNA sequence to guide the endonuclease to a region of dsDNA of interest.

[0099] The present disclosure also provides Cas12a proteins that are complexed with a guide RNA that is smaller than would be used for Cas9 proteins, which comprises up to about 100 nucleotides. As such, CRISPR-Cas12a endonuclease of the present disclosure are much smaller than CRISPR/Cas9 systems and, thus, easier to package in viral vectors for delivery. Cas12a proteins of the present disclosure are able to recognize a protospacer adjacent motif (PAM) sequence, which is a recognition sequence adjacent to the target nucleic acid sequence to be bound. Some Cas12a proteins of the present disclosure recognize nucleic acid sequence stretches. Unlike Cas9 proteins that recognize G-rich PAM sequences, Cas12a proteins of the present disclosure are capable of recognizing T-rich PAM sequences increasing the number of regions of double stranded DNA that may be targeted and cleaved. Alternatively, Cas12a proteins of the present disclosure are also capable of recognizing A-rich PAM sequences, G-rich PAM sequences, and C-rich PAM

sequences. The PAM comprises a three nucleobase PAM such as a T-rich PAM. For example, some guide RNA-directed endonucleases, such as Cas12a proteins described herein, recognize a PAM sequence of 5'-TTTN-3'. Alternatively a PAM sequence having two T nucleobases or 1 T nucleobase is also consistent with the present disclosure. Also consistent with the present disclosure is a PAM sequence that is pyrimidine rich or purine rich.

[0100] Unlike Cas9 proteins that cleave close to the recognition site, Cas12a proteins of the present disclosure cleave farther away from the recognition site. This is particularly advantageous because resulting non-homologous end joining (NHEJ) results in preservation of the PAM sequence, allowing for further editing and improves homology directed repair (HDR). Finally, unlike Cas9 proteins which generate blunt ends in dsDNA upon cleavage, Cas12a endonucleases of the present disclosure generate staggered or sticky ends in dsDNA upon cleavage. Accordingly, some guide RNA-directed endonucleases, such as the Cas12a proteins of the present disclosure, facilitate genome editing by generating staggered or sticky ends that are more likely to be repaired through HDR rather than random NHEJ.

Cas12a Proteins

[0101] The present disclosure provides Cas12a proteins (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) identified by metagenomics mining. The present disclosure interchangeably refers to polypeptides and proteins. Thus, a polypeptide of the present disclosure is a Cas12a protein, which may also be referred to as a Cas12a polypeptide. For example, metagenome base sequences are downloaded from the NCBI Genbank BLAST database and saved as a local BLASTp database. A number of different Cas12a and Cas1 amino acid sequences are downloaded from the Uniprot database. MetaCRT is used to identify key CRISPR repeat and spacer sequences from the metagenome and all metagenome sequences having CRISPR sequences are extracted using the Prodigal program. Taxonomic hierarchies are built based on CRISPR sequences and novel Cas12a proteins are identified by search for homology to other Cas12a sequences. These Cas12a sequences are interchangeably referred to as Cas12a proteins or metagenome Cas12a endonucleases (mgCas12a).

[0102] In some embodiments, the Cas12a proteins are obtained or derived (or modified) from the *Eubacteriaceae* bacteria family. For example, the Cas12a proteins described herein are obtained or derived (or modified) from *Eubacterium rectale* or *Eubacterium eligens*.

[0103] In some embodiments, Cas12a proteins discovered via metagenomics mining can be wild-type proteins. However, the Cas12a proteins described herein can be humanized and/or engineered to be compatible for administration to a subject in need thereof. In some cases, a humanized Cas12a protein may include mutations in a wild-type Cas12a protein sequence, which may make the

humanized Cas12a protein less likely to elicit immunogenicity relative to a wild-type Cas12a protein without impacting protein function. If administered using a plasmid, a humanized Cas12a sequence are codon optimized to facilitate expression in a mammalian or human system. An engineered Cas12a protein includes mutations in the sequence, which improve nuclease activity and function. In some cases, humanization and engineering of a Cas12a disclosed herein increases cleavage efficiency. Alternatively, or in combination, humanization or engineering can increase cleavage specificity. In some cases, a dead and engineered (“de”) Cas12a protein is used as a control to compare relative activity to a wild-type or humanized and/or engineered Cas12a protein. A dead Cas12a protein includes mutations in a wild-type Cas12a protein sequence which make the protein non-functional and inactive.

[0104] The Cas12a proteins of the present disclosure (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) cleave a nucleic acid sequence at a particular pH. For example, a Cas12a endonuclease cleaves a nucleic acid sequence from about pH 7 to about pH 7.9, or from about pH 7 to about pH 8.0. In some cases, a Cas12a endonuclease cleaves at about pH 7, about pH 7.1, about pH 7.2, about pH 7.3, about pH 7.4, about pH 7.5, about pH 7.6, about pH 7.7, about pH 7.8, or about pH 7.9. A Cas12a endonuclease described herein is capable of cleaving a nucleic acid sequence from about pH 7 to about pH 7.3, about pH 7.3 to about pH 7.6, or about pH 7.6 to about pH 7.9. Cas12a endonucleases disclosed herein does not exhibit cleaving activity and is, thus, inactive at a pH greater than 7.9. For example, Cas12a endonuclease described herein (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) is activated and capable of cleaving cellular environments comprising a pH of from about 7.0 to about 7.9. Accordingly, Cas12a endonucleases described herein (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) is inactivated and becomes inert if the environmental pH is outside of from about 7.0 to about 7.9. Alternatively, Cas12a proteins of the present disclosure is capable of cleaving a nucleic acid sequence at any pH, for example, a Cas12a endonuclease described herein is capable of cleaving a nucleic acid sequence at about pH 2, about pH 2.5, about pH 3, about pH 3.5, about pH 4, about pH 4.5, about pH 5, about pH 5.5, about pH 6, about pH 6.5, about pH 7, about pH 7.5, about pH 8, about pH 8.5, about pH 9, about pH 9.5, about pH 10, about pH 10.5, about pH 11, about pH 11.5, about pH 12, about pH 2 to about pH 12, about pH 2 to about pH 2.5, about pH 3 to about pH 3.5, about pH 4 to about pH 4.5, about pH 5 to about pH 5.5, about pH 6 to about pH 6.5, about pH 7 to about pH 7.5, about pH 8 to about pH 8.5, about pH 9 to about pH 9.5, about pH 10 to about pH 10.5, about pH 11 to about pH 11.5, or about pH 12 to about pH 12.5. Cas12a proteins of the present disclosure may also be inactivated and inert at any environmental pH, for example, a Cas12a endonuclease described herein may be inert at about pH

2, about pH 2.5, about pH 3, about pH 3.5, about pH 4, about pH 4.5, about pH 5, about pH 5.5, about pH 6, about pH 6.5, about pH 7, about pH 7.5, about pH 8, about pH 8.5, about pH 9, about pH 9.5, about pH 10, about pH 10.5, about pH 11, about pH 11.5, about pH 12, about pH 2 to about pH 12, about pH 2 to about pH 2.5, about pH 3 to about pH 3.5, about pH 4 to about pH 4.5, about pH 5 to about pH 5.5, about pH 6 to about pH 6.5, about pH 7 to about pH 7.5, about pH 8 to about pH 8.5, about pH 9 to about pH 9.5, about pH 10 to about pH 10.5, about pH 11 to about pH 11.5, or about pH 12 to about pH 12.5.

[0105] Thus, it is preferred that the composition comprising the Cas12a proteins may have pH in a range of from about pH 7 to about pH 7.9, from about pH 7 to about pH 7.3, about pH 7.3 to about pH 7.6, or about pH 7.6 to about pH 7.9, or about pH 7, about pH 7.1, about pH 7.2, about pH 7.3, about pH 7.4, about pH 7.5, about pH 7.6, about pH 7.7, about pH 7.8, or about pH 7.9. Yet, where the Cas12a proteins can maintain its activity (*e.g.*, can cleave a nucleic acid sequence) at any other pH range, it is also contemplated that the composition comprising the Cas12a proteins may have other pH range (*e.g.*, at about pH 2, about pH 2.5, about pH 3, about pH 3.5, about pH 4, about pH 4.5, about pH 5, about pH 5.5, about pH 6, about pH 6.5, about pH 7, about pH 7.5, about pH 8, about pH 8.5, about pH 9, about pH 9.5, about pH 10, about pH 10.5, about pH 11, about pH 11.5, about pH 12, about pH 2 to about pH 12, about pH 2 to about pH 2.5, about pH 3 to about pH 3.5, about pH 4 to about pH 4.5, about pH 5 to about pH 5.5, about pH 6 to about pH 6.5, about pH 7 to about pH 7.5, about pH 8 to about pH 8.5, about pH 9 to about pH 9.5, about pH 10 to about pH 10.5, about pH 11 to about pH 11.5, or about pH 12 to about pH 12.5).

[0106] The Cas12a proteins of the present disclosure (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) cleave a nucleic acid sequence at a particular temperature. For example, a Cas12a endonuclease cleaves a nucleic acid sequence from about 0° to about 100°C, from about 10 °C to about 20 °C, from about 20 °C to about 30 °C, from about 30 °C to about 40 °C, from about 40 °C to about 50 °C, from about 50 °C to about 60 °C, from about 60 °C to about 70 °C, from about 70 °C to about 80 °C, from about 80 °C to about 90 °C, from about 90 °C to about 100 °C, from about 100 °C to about 110 °C, from about 110 °C to about 120 °C, or from about 120 °C to about 130 °C.

[0107] A Cas12a endonuclease described in the present disclosure (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) is formulated in an excipient, preferably pharmaceutically acceptable excipient (*e.g.*, diluent, carrier, buffer). In other words, the composition comprising the Cas12a endonuclease may further comprise an excipient, preferably pharmaceutically acceptable excipient (*e.g.*, diluent, carrier, buffer). In some embodiments, the buffer comprises Bis-Tris propane-HCl. The buffer alternatively, additionally, or optionally,

comprises MgCl_2 . The buffer, additionally or optionally, is supplemented with protein, for example, a buffer used with the Cas12 endonucleases described herein may further comprise bovine serum albumin (BSA). Cas12a endonucleases described herein (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) is formulated, for example, in a buffer comprising from 0.1 to 50 mM Bis-Tris Propane-HCl, from 0.1 to 5 mM Bis-Tris Propane-HCl, from 5 to 15 mM Bis-Tris Propane-HCl, from 15 to 25 mM Bis-Tris Propane-HCl, from 25 to 35 mM Bis-Tris Propane-HCl, from 35 to 45 mM Bis-Tris Propane-HCl, from 45 to 50 mM Bis-Tris Propane-HCl, less than 0.1 mM Bis-Tris Propane-HCl, less than 0.5 mM Bis-Tris Propane-HCl, less than 1 mM Bis-Tris Propane-HCl, less than 5 mM Bis-Tris Propane-HCl, less than 10 mM Bis-Tris Propane-HCl, less than 15 mM Bis-Tris Propane-HCl, less than 20 mM Bis-Tris Propane-HCl, less than 25 mM Bis-Tris Propane-HCl, less than 30 mM Bis-Tris Propane-HCl, less than 35 mM Bis-Tris Propane-HCl, less than 40 mM Bis-Tris Propane-HCl, less than 45 mM Bis-Tris Propane-HCl, or less than 50 mM Bis-Tris Propane-HCl. The buffer additionally comprises 0.1 to 50 mM MgCl_2 , from 0.1 to 5 mM MgCl_2 , from 5 to 15 mM MgCl_2 , from 15 to 25 mM MgCl_2 , from 25 to 35 mM MgCl_2 , from 35 to 45 mM MgCl_2 , from 45 to 50 mM MgCl_2 , less than 0.1 mM MgCl_2 , less than 0.5 mM MgCl_2 , less than 1 mM MgCl_2 , less than 5 mM MgCl_2 , less than 10 mM MgCl_2 , less than 15 mM MgCl_2 , less than 20 mM MgCl_2 , less than 25 mM MgCl_2 , less than 30 mM MgCl_2 , less than 35 mM MgCl_2 , less than 40 mM MgCl_2 , less than 45 mM MgCl_2 , or less than 50 mM MgCl_2 . Additionally, the buffer comprise 1 to 500 $\mu\text{g/ml}$ BSA, from 1 to 50 $\mu\text{g/ml}$ BSA, from 50 to 100 $\mu\text{g/ml}$ BSA, from 100 to 150 $\mu\text{g/ml}$ BSA, from 150 to 200 $\mu\text{g/ml}$ BSA, from 200 to 250 $\mu\text{g/ml}$ BSA, from 250 to 300 $\mu\text{g/ml}$ BSA, from 300 to 350 $\mu\text{g/ml}$ BSA, from 350 to 400 $\mu\text{g/ml}$ BSA, from 400 to 450 $\mu\text{g/ml}$ BSA, from 450 to 500 $\mu\text{g/ml}$ BSA, less than 1 $\mu\text{g/ml}$ BSA, less than 10 $\mu\text{g/ml}$ BSA, less than 20 $\mu\text{g/ml}$ BSA, less than 30 $\mu\text{g/ml}$ BSA, less than 40 $\mu\text{g/ml}$ BSA, less than 50 $\mu\text{g/ml}$ BSA, less than 100 $\mu\text{g/ml}$ BSA, less than 150 $\mu\text{g/ml}$ BSA, less than 200 $\mu\text{g/ml}$ BSA, less than 250 $\mu\text{g/ml}$ BSA, less than 300 $\mu\text{g/ml}$ BSA, less than 350 $\mu\text{g/ml}$ BSA, less than 400 $\mu\text{g/ml}$ BSA, less than 450 $\mu\text{g/ml}$ BSA, or less than 500 $\mu\text{g/ml}$ BSA,. The present disclosure additionally describes Cas12a endonucleases formulated in a buffer comprising 10 mM Bis-Tris Propane-HCl, 10 mM MgCl_2 , and 100 $\mu\text{g/ml}$ of BSA. Alternatives outside of the ranges provided above for Bis-Tris Propane, MgCl_2 , and BSA are also consistent with the invention.

[0108] A Cas12a protein of the present disclosure cleaves 1% to 100%, 10% to 90%, 50% to 100%, 1% to 10%, 10% to 20%, 20% to 30%, 30% to 40%, 40% to 50%, 50% to 60%, 60% to 70%, 70% to 80%, 80% to 90%, 90% to 100%, or 80% to 100% of the target DNA. Target DNA compatible with the present invention includes linear DNA or plasmid DNA. Cas12a proteins of the present disclosure cleave 1% to 100%, 10% to 90%, 50% to 100%, 1% to 10%, 10% to 20%, 20% to 30%,

30% to 40%, 40% to 50%, 50% to 60%, 60% to 70%, 70% to 80%, 80% to 90%, 90% to 100%, or 80% to 100% of the target DNA in 0 to 1 hour, 1 hour to 5 hours, 5 hours to 10 hours, 10 hours to 20 hours, 20 hours to 30 hours, 30 hours to 40 hours, 40 hours to 50 hours, 50 hours to 60 hours, 60 hours to 70 hours, 70 hours to 80 hours, 80 hours to 90 hours, or 90 hours to 100 hours. The Cas12a proteins of the present disclosure cleave 90 to 100% of the target DNA within 1 hour. In addition, Cas12a proteins of the present disclosure cleave 90 to 100% of the target DNA within 12 hours. Cas12a proteins of the present disclosure cleave 90 to 100% of the target DNA within 24 hours. The above described cleavage efficiencies are achieved by reacting 0.1 pmol to 100 pmol, 0.1 to 10 pmol, 10 pmol to 20 pmol, 20 pmol to 30 pmol of a Cas12a protein of the present disclosure, 30 pmol to 40 pmol, 40 pmol to 50 pmol, 50 pmol to 60 pmol, 60 pmol to 70 pmol, 70 pmol to 80 pmol, 80 pmol to 90 pmol, or 90 to 100 pmol of a Cas12a protein of the present disclosure with 1 ng to 1000 ng of target DNA, 1 ng to 100 ng of target DNA, 100 ng to 200 ng of target DNA, 200 ng to 300 ng of target DNA, 300 ng to 400 ng of target DNA, 400 ng to 500 ng target DNA, 500 ng to 600 ng of target DNA, 600 ng to 700 ng of target DNA, 700 ng to 800 ng of target DNA, 800 ng to 900 ng of target DNA, or 900 ng to 1000 ng of target DNA. The above described cleavage efficiencies are achieved by reacting 6 pmol of a Cas12a protein of the present disclosure with 300 ng of target DNA. The above described cleavage efficiencies are achieved at 37°C.

[0109] In some cases, a Cas12a protein of the present disclosure cleaves the target DNA into cleaved products without fully degrading the DNA, or in other words, without exhibiting interminable DNase activity. Target DNAs that are cleaved by a Cas12a protein of the present disclosure include any double stranded DNA (dsDNA), any linearized dsDNA, and any plasmid dsDNA.

[0110] Preferably, a functional domain of the CAS12A protein includes residues 829 through 991 of SEQ ID NO: 1 or residues 825 through 996 of SEQ ID NO: 3. Alternatively, a Cas12a protein has at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the residues 829 through 991 of SEQ ID NO: 1 or residues 825 through 996 of SEQ ID NO: 3.

[0111] Alternatively and/or additionally a Cas12a protein of the present disclosure has at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to a functional domain in SEQ ID NO: 1.

[0112] Alternatively, a Cas12a protein of the present disclosure has at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least

95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to a functional domain in SEQ ID NO: 3.

[0113] As another example, some proteins disclosure herein share at least 20 consecutive residues, at least 25 consecutive residues, at least 40 consecutive residues, at least 55 consecutive residues, at least 70 consecutive residues, at least 85 consecutive residues, at least 100 consecutive residues, at least 115 consecutive residues, at least 130 consecutive residues, at least 145 consecutive residues, at least 160 consecutive residues, at least 175 consecutive residues, at least 190 consecutive residues, at least 205 consecutive residues, at least 220 consecutive residues, at least 235 consecutive residues, at least 250 consecutive residues, at least 265 consecutive residues, at least 280 consecutive residues, at least 295 consecutive residues, at least 310 consecutive residues, at least 325 consecutive residues, at least 340 consecutive residues, at least 355 consecutive residues, at least 370 consecutive residues, at least 385 consecutive residues, at least 400 consecutive residues, at least 415 consecutive residues, at least 430 consecutive residues, at least 445 consecutive residues, at least 460 consecutive residues, at least 475 consecutive residues, at least 490 consecutive residues, at least 505 consecutive residues, at least 520 consecutive residues, at least 535 consecutive residues, at least 550 consecutive residues, at least 565 consecutive residues, at least 580 consecutive residues, at least 595 consecutive residues, at least 610 consecutive residues, at least 625 consecutive residues, at least 640 consecutive residues, at least 655 consecutive residues, at least 670 consecutive residues, at least 685 consecutive residues, at least 700 consecutive residues, at least 715 consecutive residues, at least 730 consecutive residues, at least 745 consecutive residues, at least 760 consecutive residues, at least 775 consecutive residues, at least 790 consecutive residues, at least 805 consecutive residues, at least 820 consecutive residues, at least 835 consecutive residues, at least 850 consecutive residues, at least 865 consecutive residues, at least 880 consecutive residues, at least 895 consecutive residues, at least 910 consecutive residues, at least 925 consecutive residues, at least 940 consecutive residues, at least 955 consecutive residues, at least 970 consecutive residues, at least 985 consecutive residues, at least 1000 consecutive residues, at least 1015 consecutive residues, at least 1030 consecutive residues, at least 1045 consecutive residues, at least 1060 consecutive residues, at least 1075 consecutive residues, at least 1090 consecutive residues, at least 1105 consecutive residues, at least 1120 consecutive residues, at least 1135 consecutive residues, at least 1150 consecutive residues, at least 1165 consecutive residues, at least 1180 consecutive residues, at least 1195 consecutive residues, at least 1210 consecutive residues, at least 1225 consecutive residues, at least 1240 consecutive residues, or at least 1250 consecutive residues, with a Cas12a protein of the present disclosure.

[0114] A Cas12a protein of the present disclosure has at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to at least 50 consecutive residues, at least 100 consecutive residues, at least 150 consecutive residues, at least 200 consecutive residues, at least 250 consecutive residues at least 300 consecutive residues, at least 350 consecutive residues, at least 400 consecutive residues, at least 450 consecutive residues, at least 500 consecutive residues, at least 550 consecutive residues, at least 600 consecutive residues, at least 650 consecutive residues, at least 700 consecutive residues, at least 750 consecutive residues, at least 800 consecutive residues, at least 850 consecutive residues, at least 900 consecutive residues, at least 950 consecutive residues, at least 1000 consecutive residues, at least 1050 consecutive residues, at least 1100 consecutive residues, at least 1150 consecutive residues, at least 1200 consecutive residues, at least 1250 consecutive residues, or at least 1263 consecutive residues of SEQ ID NO: 1.

[0115] A Cas12a protein of the present disclosure has at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to at least 50 consecutive residues, at least 100 consecutive residues, at least 150 consecutive residues, at least 200 consecutive residues, at least 250 consecutive residues at least 300 consecutive residues, at least 350 consecutive residues, at least 400 consecutive residues, at least 450 consecutive residues, at least 500 consecutive residues, at least 550 consecutive residues, at least 600 consecutive residues, at least 650 consecutive residues, at least 700 consecutive residues, at least 750 consecutive residues, at least 800 consecutive residues, at least 850 consecutive residues, at least 900 consecutive residues, at least 950 consecutive residues, at least 1000 consecutive residues, at least 1050 consecutive residues, at least 1100 consecutive residues, at least 1150 consecutive residues, at least 1200 consecutive residues, at least 1250 consecutive residues, or at least 1275 consecutive residues of SEQ ID NO: 3.

[0116] A Cas12a protein of the present disclosure has at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the full length sequence of SEQ ID NO: 1. Alternatively, a Cas12a protein of the present disclosure has at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the full length sequence of SEQ ID NO: 3.

[0117] A Cas12a protein of the present disclosure has at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to residues 1 through 100, residues 100 through 200, residues 200 through 300, residues 300 through 400, residues 400 through 500, residues 500 through 600, residues 600 through 700, residues 700 through 800, residues 800 through 900, residues 900 through 1000, residues 1000 through 1100, residues 1100 through 1200 or residues 1200 through 1263 of SEQ ID NO: 1.

[0118] A Cas12a protein of the present disclosure has at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to residues 1 through 100, residues 100 through 200, residues 200 through 300, residues 300 through 400, residues 400 through 500, residues 500 through 600, residues 600 through 700, residues 700 through 800, residues 800 through 900, residues 900 through 1000, residues 1000 through 1100, residues 1100 through 1200 or residues 1200 through 1275 of SEQ ID NO: 3.

[0119] The present disclosure provides a Cas12a protein having an amino acid sequence comprising SEQ ID NO: 1 and is also referred to herein as mgCas12a-1. Additionally, a Cas12a protein of the present disclosure is encoded for by a nucleotide sequence comprising SEQ ID NO: 2. A nucleotide sequence of SEQ ID NO: 2 encodes for an mgCas12a protein of SEQ ID NO: 1.

[0120] The present disclosure additionally provides a Cas12a protein having an amino acid sequence comprising SEQ ID NO: 3 and is also referred to herein as mgCas12a-2. Additionally a Cas12a protein of the present disclosure is encoded for by a nucleotide sequence comprising SEQ ID NO: 4. A nucleotide sequence of SEQ ID NO: 4 encodes for an mgCas12a protein of SEQ ID NO: 3.

[0121] The present disclosure provides a Cas12a protein of the present disclosure has a Lysine (K) at position (residue) 925, as in SEQ ID NO: 1. Additionally, a Cas12a protein of the present disclosure has a Lysine (K) at position (residue) 930, as in SEQ ID NO: 3. Thus, in some embodiments, the Cas12a protein or a variant thereof may have Lysine (K) at position (residue) 925 or 930, or in a position other than position (residue) 925 or 930, when it is optimally aligned with SEQ ID NO: 1 or with SEQ ID NO: 3, respectively. In other words, a Cas12a protein of the present disclosure having a Lysine (K) at position (residue) 925 includes a Cas12a protein or Cas12a protein variants (a portion of Cas12a protein, a polypeptide that includes some domains of Cas12a protein, etc.) that has a Lysine (K) in a position corresponding to residue 925 of SEQ ID NO: 1, a Lysine (K) in a position aligned to residue 925 of SEQ ID NO: 1, or residue 925 relative to SEQ ID NO: 1. For example, if amino acid residues 915-930 of the Cas12a protein or its variant are aligned

with amino acid residues 918-933 of the SEQ ID NO: 1, a Lysine (K) is preferably located in amino acid residue 922 that is relative to, equivalent to, or aligned to residue 925 of SEQ ID NO: 1.

[0122] A Cas12a polypeptide of the present disclosure can also have an amino acid sequence comprising a Lysine aligned to a Lysine at position 925 of SEQ ID NO: 1, a Lysine aligned to a Lysine at position 930 of SEQ ID NO: 3, a Lysine aligned to a Lysine at position 925 of SEQ ID NO: 1, a Lysine aligned to a Lysine at position 930 of SEQ ID NO: 3, or any combination thereof. For example, the present disclosure provides a composition comprising a polypeptide having at least 80% sequence identity with residue 829 through residue 991 of SEQ ID NO: 1, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 925 of SEQ ID NO: 1; a polypeptide having at least 80% sequence identity with residue 825 through residue 996 of SEQ ID NO: 3, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 930 of SEQ ID NO: 3; a polypeptide having at least 80% sequence identity with SEQ ID NO: 1, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 925 of SEQ ID NO: 1; or a polypeptide having at least 80% sequence identity with SEQ ID NO: 3, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 930 of SEQ ID NO: 3. In the Cas12a polypeptides disclosed herein, a first residue aligned to a second residue at a position can indicate that the first residue is in a candidate polypeptide sequence and the alignment is in reference to a second residue and a position in a reference polypeptide sequence. For example, a polypeptide having at least 80% sequence identity with residue 829 through residue 991 of SEQ ID NO: 1, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 925 of SEQ ID NO: 1 can refer to the candidate polypeptide having an Lysine that can be aligned to Lys925 of SEQ ID NO: 1. Alignment can be done using any of the methods disclosed herein and aligned residues will be readily apparent to one of skill.

[0123] Additionally, a Cas12a protein of the present disclosure has an amino acid sequence comprising SEQ ID NO: 5, which comprises the SEQ ID NO: 1 having a Lysine to glutamine mutation at position 925 (K925Q). The present disclosure additionally provides a Cas12a protein having an amino acid sequence comprising SEQ ID NO: 6, which comprises the SEQ ID NO: 3 having a Lysine to glutamine mutation at position 930 (K930Q).

[0124] It is consistent with the present disclosure for the Lysine at position 925 of SEQ ID NO: 1 to be substituted with any other amino acid residue. The Lysine at position 925 of SEQ ID NO: 1 may also be substituted with arginine (Arg), histidine (His), aspartic acid (Asp), glutamic acid (Glu), serine (Ser), threonine (Thr), asparagine (Asn), glutamine (Gln), tyrosine (Tyr), alanine (Ala), isoleucine (Ile), leucine (Leu), valine (Val), phenylalanine (Phe), methionine (Met),

tryptophan(e) (Trp), glycine (Gly), proline (Pro), or cysteine (Cys). Additionally, the present disclosure also describes that the Lysine at position 930 of SEQ ID NO: 3 is substituted with any other amino acid residue. The Lysine at position 925 of SEQ ID NO: 1 may be substituted with arginine (Arg), histidine (His), aspartic acid (Asp), glutamic acid (Glu), serine (Ser), threonine (Thr), asparagine (Asn), glutamine (Gln), tyrosine (Tyr), alanine (Ala), isoleucine (Ile), leucine (Leu), valine (Val), phenylalanine (Phe), methionine (Met), tryptophan(e) (Trp), glycine (Gly), proline (Pro), or cysteine (Cys).

[0125] A base sequence of a Cas12a of the present disclosure (*e.g.*, mgCas12a-1) is human codon optimized, which in some cases can improve protein expression, and has a sequence comprising SEQ ID NO: 7. A base sequence of a Cas12a of the present disclosure (*e.g.*, mgCas12a-2) is human codon optimized, which in some cases can improve protein expression, and has a sequence comprising SEQ ID NO: 8.

[0126] Cas12a proteins disclosed herein (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) cleave DNA or RNA.

[0127] **TABLE 1** shows exemplary Cas12a sequences of the present disclosure, including SEQ ID NO: 1 – SEQ ID NO: 8.

TABLE 1 – Exemplary Cas12a Sequences

SEQ ID NO	Sequence
SEQ ID NO: 1	MNGTNNFQNFIFIGISSLQKTLRNALIPTETTQQFIVKNGIIKEDELRGENRQIL KDIMDDYYRGFISSETLSSIDDIDWTSLEFEMEIQKNGDNKDTLIKEQAEKRKA IYKKFADDDRFKNMFSAKLISDILPEFVIHNNNYSASEKEEKTQVIKLFSEFAT SFKDYFKNRANCFSSADDISSSSCHRVNDNAEIFFSNALVYRRIVKNLSNDDIN KISGDIKDSLKEMSLEEIYSYKEYGEFITQEGISFYNDICGKVNFSMNLVCQKN KENKNLYKLRKLHKQILCIADTSYEVYPYKFESDEEVYQSVNGFLDNISSKHIVE RLRKIGDNYNGYNLDKIYIVSKFYESVSQKTYRDWETINTALEIHNNILPGNG KSKADKVKKAVKNDLQKSITEINELVSNYKLCPPDNKAETIHEISHILNFE AQELKYNPEIHLVESELKASELKNVLDVIMNAFWCSVFMTEELVDKDNFYAE LEEIYDEIYTVISLYNLVRNYVTQKPYSTKKIKLNFGIPTLADGWSKSKEYSNN AIIIMRDNLYLGI FNAKNKPKKII EGNTSENKGDYKMIYNLLPGPNKMIPK VFLSSKTGVETYPKPSAYILEGYKQNKHLKSSKDFDITFCHDLIDYFKNCIAIHP EWKNFGFDFSDTSTYEDISGFYREVELQGYKIDWTYISEKDIDLLQEKGLYLF QIYNKDFSKKSTGNDNLHTMYLKNLFSEENLKDIVLKLNGEAEIFFRKSSIKNP IIHKKGSILVNRITYEAEKQDFGNIQIVRKTIPENIYQELYKYFNDKSDKELSD EAAKLKNVVGHEAATNIVKDYRYTYDKYFLHMPITINFKANKTSFINDRILQY IAKEKNLHVIGIDRGERNLIYVSVIDTCGNIVEQKSFNIVNGYDYQIKLQKQEG ARQIARKEWKEIGIKIKEIGYLSLVIHEISKMVIKYNAI IAMEDLSYGFKKGR FKVERQVYQKFETMLINKLNLVFKDISITENGGLLKGYQLTYIPDKLKNVGHQ CGCIFVPAAYTSKIDPTTGfVNI FKFKDLTVDAKREFIKKFDSIRYDSEKKLF CFTFDYNNFITQNTVMSKSSWSVYTYGVRIKRRFVNGRFSNESDTIDITKMEK TLEMTDINWRDGHDLRQDIIDYEIVQHI FEIFRLTVQMRNSLSELED RDYDRLI SPVLNENNI FYDSAKAGDALPKDADANGAYCIALKGLYEIKQITENWKEDGKFS RDKLKISNKDWFDFIQNKRYL

SEQ ID NO	Sequence
SEQ ID NO: 2	ATGAATAACGGAACAAATAACTTTTCAGAACTTTATCGGAATTTCTTCTTTGCAG AAGACTCTTAGGAATGCTCTCATTCCAACAGAAACAACACAGCAATTTATTGTT AAAAATGGAATAATTAAAGAAGATGAACTCAGAGGAGAAAATCGTCAGATACTT AAAGATATCATGGATGATTATTACAGAGGTTTCATTTAGAAAATTTATCGTCA ATTGATGATATTGACTGGACCTCTTTATTTGAGAAAATGGAAATTCAGTTAAAA AATGGAGATAATAAAGACACTCTTATAAAAAGAACAGGCTGAAAAACGTAAGGCA ATCTATAAAAAATTTGCAGATGATGATAGATTTAAAAATATGTTTCAGTGCAAAA TTAATCTCAGATATTCTTCCTGAATTTGTCATTTCATAACAATAATTATTCTGCA TCAGAAAAGGAAGAAAAAACACAGGTAATTAATTATTTTCCAGATTTGCAACA TCATTCAAGGACTATTTTAAAAACAGGGCTAATTGTTTTTCTGCTGATGATATA TCTTCTTCTTCTTGTGATAGAAATAGTTAATGATAATGCAGAAATATTTTTTAGT AATGCATTGGTGTATAGGAGAATTGTAAAAAATCTTCAAATGATGATATAAAT AAAATATCCGGAGATATTAAGGATTCATTAAAGGAAATGTCTCTGGAGGAAATT TATTCTTATGAAAAATATGGGGAATTTATTACACAGGAAGGTATATCTTTTTAT AATGATATATGCGGTAAAGTAAATTCATTTATGAATTTATATTGCCAGAAAAAT AAAGAAAACAAAAATCTCTATAAGCTGCGAAAGCTTCATAAACAGATACTGTGC ATAGCAGATACTTCTTATGAGGTGCCGTATAAATTTGAATCAGATGAAGAGGTT TATCAATCAGTGAATGGATTTTTGGACAATATTAGTTCAAAACATATCGTTGAA AGATTGCGTAAGATTGGAGACAATAACGGCTACAATCTTGATAAGATTTAT ATTGTTAGTAAATTCTATGAATCAGTTTCCAAAAGACATATAGAGATTGGGAA ACAATAAAATACTGCATTAGAAATTCATTACAACAATATATTACCCGGAAATGGT AAATCTAAAGCTGACAAGGTAAAAAAGCGGTAAAGAATGATCTGAAAAAAGC ATTACTGAAATCAATGAGCTTGTTAGCAATTATAAATTATGTCCGGATGATAAT ATTAAAGCAGAGACATATATACATGAAATATCACATATTTTGAATAATTTTGAA GCACAGGAGCTTAAGTATAATCCTGAAATTCATCTGGTGGAAAGTGAATTGAAA GCATCTGAATTAAAAAATGTTCTCGATGTAATAATGAATGCTTTTCATTGGTGT TCGGTTTTTCATGACAGAGGAGCTGGTAGATAAAGATAATAATTTTTATGCGGAG TTAGAAGAGATATATGACGAAATATATACGGTAATTTTCATTGTATAATCTTGTG CGTAATTATGTAACGCAGAAAGCCATATAGTACAAAAAAAATTAATTTGAATTTT GGTATTCCTACACTAGCGGATGGATGGAGTAAAAGTAAAGAATATAGTAATAAT GCAATTATTCTCATGCGTGATAATTTGTACTATTTAGGAATATTTAATGCAAAA AATAAGCCTGACAAAAAGATAATTGAAGGTAATACATCAGAAAATAAAGGGGAT TATAAGAAGATGATTTATAATCTTCTGCCAGGACCAAATAAAATGATCCCCAAG GTATTCCTCTCTTCAAAAACCGGAGTGGAAACATATAAGCCGTCTGCCTATATA TTGGAGGGCTATAAACAAAACAAGCATCTTAAATCCTCTAAGGATTTTGATATA ACGTTTTGTGACGATTTGATTGATTATTTTAAAGAACTGTATAGCAATACATCCT GAATGGAAGAATTTTGGCTTTGATTTTTCTGACACCTCCACATATGAAGATATC AGCGGATTTTACAGAGAAGTCGAATTGCAAGGTTATAAAATTGACTGGACATAT ATCAGCGAAAAGGATATTGATTTGTTGCAGGAAAAGGACAGTTATATTTATTT CAAATATATAACAAAGATTTTCCAAAGAAAAGTACCGGAAATGATAATCTTCAT ACTATGTATTTGAAGAATTTGTTTAGCGAAGAGAATTTAAAGGATATTGTACTG AAATTAACCGGTGAGGCGGAAATCTTCTTTAGAAAATCAAGCATAAAGAATCCA ATAATTCATAAAAAAGGCTCTATTCTTGTTAATAGAACATATGAAGCAGAGGAA AAAGATCAATTTGGAAATATCCAGATAGTCAGAAAAACCATAACCGGAAAATATA TATCAGGAGCTTTATAAATATTTCAATGATAAAAGTGATAAAGAATTTTCGGAT GAAGCAGCTAAGCTTAAGAATGTAGTAGGTCATCATGAGGCTGCTACAAACATA GTAAAAGATTATAGATATACATATGATAAATATTTTCTTCATATGCCTATTACA ATCAATTTTAAAGCCAATAAGACAAGCTTTATTAATGACAGAATATTACAATAT ATTGCTAAAGAAAAGAATTTGCATGTAATAGGCATTGATCGTGGTGAAAAGAAC CTGATATATGTTTCAGTAATTGATACTTGTGGAAATATTGTTGAACAAAAATCG TTTAACATTGTTAATGGATATGATTATCAGATTAAGCTCAAGCAGCAGGAGGGG

SEQ ID NO	Sequence
	GCGCGACAAATCGCACGAAAAGAATGGAAAAGAAATCGGCAAAATAAAAAGAAATT AAAGAAGGCTATTTATCTCTTGTAAATTCATGAAATTTCAAAGATGGTTATTAAA TATAATGCCATAATTGCAATGGAGGATTTAAGCTACGGATTTAAAAAAGGTCGT TTCAAGGTTGAGCGACAGGTTTACCAGAAGTTTGAGACAATGCTTATCAACAAA CTCAACTATCTGGTATTTAAAGATATATCCATAACTGAAAACGGTGGTCTTCTA AAGGGATATCAGCTTACATATATTCAGATAAACTGAAAAATGTGGGTGATCAA TGTGGTTGTATATTTTACGTACCTGCTGCCTATACATCAAAAAATAGATCCTACA ACCGGATTTGTAAATATATTCAAATTTAAAGATTTAACAGTTGATGCAAAGAGA GAATTTATAAAAAAATTTGACAGTATCAGATATGATTCAGAAAAAAAACGTGTTT TGTTTTACATTTGATTATAATAACTTTATTACGCAAAATACTGTTATGTCAAAG TCAAGCTGGAGTGTATATACGTACGGAGTTAGGATAAAAAGAAGATTTGTCAAT GGCAGGTTCTCAAATGAATCGGATACAATTGATATAACAAAAGATATGGAAAA ACCCTCGAAATGACAGATATAAATTGGAGAGATGGTCATGATCTGAGGCAGGAT ATTATTGATTATGAAATCGTACAACACATATTTGAGATTTTATGATTGACTGTA CAAATGAGAAACAGTTTAAGTGAATTAGAAGACAGGGATTATGACCGTTTGATT TCTCCGGTGCTCAATGAAAATAATATATTTTATGATTGAGCTAAAGCAGGAGAT GCGTTACCTAAAGACGCAGATGCTAATGGTGCATATTGTATAGCTCTAAAAGGC TTGTATGAAATCAAACAAATTACAGAGAATTGGAAAGAAGACGGTAAGTTTTCA AGAGATAAACTTAAATTTCCAATAAGGACTGGTTTGACTTTATTCAAATAAAA AGGTATTTATAA
SEQ ID NO: 3	MGKNQNFQEFIGVSPLOKTLRNELIPTETTKKNI TQLDLLTEDEIRAQNREKLK EMDDYYRNVIDSTLHVGIADVWSYLFSCMRNHLRENSKESKRELERQTQDSIRS QIHNKFAERADFKDMFGASIIITKLLPTYIKQNSEYSEYDESMEILKLYGKFTT SLTDYFETRKNIFSKEKISSAVGYRIVEENAEI FLQNQNAYDRICKIAGLDLHG LDNEITAYVDGKTLKEVCSDEGFAKAITQEGIDRYNEAIGAVNQYMNLLCQKNK ALKPGQFKMKRLHKQILCKGTTSFDPKPFENDKQVYDAVNSFTEIVTKNNDLK RLLNITQNANDYDMNKIYVVADAYSMISQFISKKNWNLIEECLLDYSDNLPKGK NAKENKVKKAVKEETYRSVSQ LNEVIEKYVVEKTGQSVWKVESYISSLAEMI KL ELCHEIDNDEKHNLIEDDEKISEIKELLDMYMDVFHIIKVFRVNEVLNFDETFY SEMDEIYQDMQEI VPLYNHVRNYVTQKPYKQEKYRLYFHTPTLANGWSKSKEYD NNAIILVREDKYLLGILNAKKKPSKEIMAGKEDCSEHAYAKMNYLLPGANKML PKVFLSKKGIQDYHPSSYIVEGYNEKKHIKGSKNFDIRFCRDLIDYFKECIKKH PDWNKFNFEFSATETYEDISVFYREVEKQGYRVEWTYINSEDIQKLEEDGQLFL FQIYNKDFAVGSTGKPNLHTLYLKNLFSEENLRDIVLKLNGEAEIFFRKSSVQK PVIHKCGSILVNRTYEITESGTTTRVQSIPESEYMELYRYFNSEKQIELSDEAKK YLDKVQCCKAKTDIVKDYRYTMDKFFIHLPIITINFKVDKGNNVNAIAQQYIAGR KDLHVIGIDRGERNLIYVSVIDMYGRILEQKS FNLVEQVSSQGT KRYDYKEKL QNREERDKARKSWKTIGKIKELKEGYLSSVIHEIAQMVKYNAI IAMEDLNYG FKRGRFKVERQVYQKFETMLISKLNLYLADKSQAVDEPGGILRGYQMTYVPDNIK NVGRQCGIIFVYPAAYTSKIDPTTGFINAFKRDVVSNDAKENFLMKFDSIQYD IEKGLFKFSFDYKNFATHKLTAKTKWDVYTNGTRI QNMKVEGHWLSMEVELTT KMKELDDSHIPYEEGQNILDDLREMKDITTI VNGILEIFWLT VQLRNSRIDNP DYDRIISPVLNKNGEFFDSDEYNSYIDAQKAPLPIDADANGAFICIALKGMYTAN QIKENWVEGEKLPADCLKIEHASWLA FMQGERG
SEQ ID NO: 4	ATGGGTAAAAATCAAATTTTCAGGAATTTATTGGGGTATCACCACCTCAAAG ACTTTAAGAAACGAATTAATCCCAACAGAAACAACAAAAAAGAATATTACTCAG CTTGATCTTTTGACTGAGGATGAAATCCGCGCGCAAAATCGAGAGAAGCTGAAA GAGATGATGGATGACTACTACCGGAATGTGATTGATAGCACTTTGCATGTGGGT ATAGCTGTTGATTGGAGCTATTTATTTTCGTGTATGCGAAATCATCTAAGGGAG AATTCCAAAGAGTCAAAGCGGGAATTGGAACGAACACAGGATTCTATTCTGTTCA CAAATCCATAATAAGTTTGCTGAACGAGCGGATTTAAGGATATGTTTGAGCA

SEQ ID NO	Sequence
	TCGATAATAACAAAATTACTTCCGACATATATAAAACAGAATTCAGAATATTCC GAGCGGTATGACGAGAGCATGGAAATTTTGAAACTGTATGGAAAATTCACAACA TCGTTGACCGATTACTTTGAGACAAGAAAGAATATCTTTTCTAAAGAGAAAATA TCTTCTGCCGTTGGATATCGAATCGTAGAGGAAAATGCTGAGATCTTCTTGACAG AATCAGAATGCTTACGACAGAATCTGTAAGATAGCGGGACTGGATTTACATGGA TTGGATAATGAAATAACAGCATATGTTGATGGAAAAACATTAAAAGAAGTATGT TCGGATGAAGGATTTGCAAAGGCTATTACACAAGAAGGGATTGATCGCTACAAC GAGGCAATCGGTGCAGTAAATCAATATATGAATCTGTTATGCCAGAAGAATAAG GCATTAAAACCGGGACAATTTAAGATGAAGCGGCTACATAAACAGATTCTTTGC AAAGGAACAACCTCTTTTCGATATTTCCAAAGAAGTTTGAAAATGATAAACAGGTG TATGACGCAGTTAATTCTTTTACAGAGATAGTAACGAAGAATAATGATTTGAAG CGACTGTTAAATATTACACAGAATGCAAATGATTATGACATGAATAAAATCTAT GTAGTAGCCGATGCATATAGTATGATTTTACAGTTTATCAGTAAAAAATGGAAT CTGATTGAAGAATGCTTGCTGGATTATTATAGCGATAATTTGCCGGGAAAAGGA AATGCGAAAAGAAAACAAAGTTAAAAAGGCGGTAAAGGAAGAAACGTATCGCAGT GTTTCACAGTTGAATGAAGTTATTGAGAAATATTATGTGGAAAAGACCGGACAG TCAGTATGGAAAGTGGAAAGTTATATTTCTAGTCTGGCAGAAATGATTAAGCTG GAATTGTGCCACGAGATAGATAACGATGAGAAGCATAATCTGATTGAAGATGAT GAGAAGATATCCGAGATTAAGGAAGTGTGGATATGTACATGGATGTATTTTCAT ATTATAAAAGTGTTCGGGTGAATGAAGTATTGAATTTTCGATGAAACCTTTTAT TCGGAGATGGATGAGATCTATCAGGATATGCAGGAAATCGTTCCATTATACAAT CATGTTTCGAACTATGTTACACAGAAACCATATAAGCAGGAGAAATATCGTTTA TATTTCCACACTCCAACATTGGCAAATGGCTGGTCCAAGAGTAAGGAATATGAC AACAACGCAATTATATTGGTGCGAGAAGATAAATATTATTTAGGTATTCTGAAT GCGAAAAGAAACCATCGAAAGAAATTATGGCGGGCAAAGAGGATTGTTTCAGAA CATGCATATGCAAAGATGAATTATTATTTGTTGCCGGGCGCGAACAAGATGCTT CCAAAAGTATTTTTATCTAAGAAAGGAATACAGGACTATCACCCATCATCATAT ATTGTTGAAGGATATAATGAAAAGAAACATATTAAAGGTTCCAAGAATTTTGAT ATCCGGTTTTGTAGGGATTTGATTGACTACTTCAAGGAATGCATTAAAAAACAT CCGGATTGGAATAAGTTTTAACTTTGAATTTTCTGCGACAGAAACATATGAGGAT ATCAGTGTCTTTTATCGCGAAGTTGAAAAGCAAGGATATCGCGTAGAGTGGACG TATATCAATAGTGAAGATATTCAGAACTGGAAGAAGATGGACAGTTGTTTTTA TTTCAGATATATAACAAAGATTTTGCTGTGGGAAGTACAGGTAAACCAAATCTT CATACATTGTATCTGAAAAATCTGTTTCAGCGAAGAAAAATTTGCGGGACATTGTA TTAAAACATAATGGGGAAGCAGAAATATTCTTCCGTAAATCAAGTGTTCAAAAA CCGGTGATTTCATAAGTGCGGCAGTATTTTAGTCAATCGTACCTATGAGATTACC GAGAGTGGAACAACACGGGTACAATCAATTCCGGAAAGTGAATACATGGAATTA TATCGCTACTTTAATAGTGAAGAGCAGATAGAATTATCAGATGAGGCAAAAAAA TATTTGGACAAGGTGCAATGTAATAAGGCAAAGACAGATATTGTGAAAGACTAC CGATACACCATGGACAAGTTTTTTATTTCATCTTCCGATTACGATTAATTTTAAG GTTGATAAGGGTAACAATGTTAATGCCATTGCACAGCAATATATTGCAGGGCGG AAAGATTTACATGTGATAGGAATTGATCGAGGAGAACGGAATCTGATTTACGTT TCTGTAATTGACATGTATGGTAGAATTTTAGAGCAGAAATCCTTTAACCTTGTTG GAACAGGTATCGTCGAGGGAACGAAGCGATATTACGATTACAAAGAAAAATTA CAGAACC GGGAAGAGGAACGGGATAAAGCAAGAAAGAGTTGGAAGACAATCGGC AAGATTAAGGAATTAAGAGAGGGGTATCTGTCGTCAGTAATTCATGAGATTGCA CAGATGGTTCGTAAAGTATAACGCAATCATTTGCAATGGAAGATTTGAATTATGGA TTTAAGCGGGGAAGATTCAAAGTAGAGCGCCAGGTATATCAGAAATTTGAAACG ATGTTGATCAGTAAGTTGAATTATCTGGCAGATAAATCTCAGGCTGTGGATGAA CCGGGAGGTATATTACGGGGATATCAGATGACTTATGTGCCGGATAATATTAAG AATGTTGGAAGACAATGTGGAATAATCTTTTATGTGCCGGCAGCATATACCTCC

SEQ ID NO	Sequence
	AAGATTGATCCGACAACCGGATTTATCAATGCATTTAAGCGGGATGTGGTATCA ACAAATGATGCAAAAGAGAATTTCTGATGAAGTTTGATTCTATTTCAGTACGAT ATAGAAAAAGGCTTATTTAAGTTTTTCATTTGATTACAAAAATTTTGCCACACAT AACTTACACTTGCGAAGACAAAATGGGACGTATATACAAATGGAACCTCGAATA CAAAACATGAAAGTTGAAGGACATTGGCTTTCAATGGAAGTTGAACCTACAACG AAAATGAAAGAGTTGCTGGATGACTCGCATATTCCATATGAAGAAGGACAGAAT ATATTGGATGATTTGCGGGAGATGAAAGATATAACAACCATTTGTGAATGGTATA TTGGAAATCTTCTGGTTGACAGTCCAGCTTCGGAATAGCAGGATAGATAATCCG GATTACGATAGAATTATCTCACCGGTATTGAATAAAAAATGGAGAATTTTTTTGAT TCTGATGAATATAATTTCATATATTGATGCGCAAAAGGCACCGTTACCGATAGAT GCCGATGCAAATGGCGCATTTTGCATTGCATTAAAAGGAATGTATACTGCCAAT CAGATCAAAGAAAACCTGGGTTGAAGGGGAGAACTTCGGCGGGATTGCTTGAAG ATCGAACATGCGAGTTGGTTAGCATTTATGCAAGGAGAAAGGGGATAG
SEQ ID NO: 5	MNNGTNNFQNFQFIGISSLQKTLRNALIPTETTQQFIVKNGIIKEDELNGENRQIL KDIMDDYYRGFISETLSSIDDIDWTSLFKMEIQLKNGDNKDTLIKEQAEKRKA IYKKFADDDRFKNMFSAKLISDILPEFVIHNNNYSASEKEEKTQVIKLFSEFAT SFKDYFKNRANCFSADDISSSSCHRIVNDNAEIFFSNALVYRRIVKNLSNDDIN KISGDIKDSLKEMSLEEIYSYEKGEFITQEGISFYNDICGKVNSFMNLYCQKN KENKNLYKLRLHKQILCIADTSYEVYPKFESDEEVYQSVNGFLDNISSKHIVE RLRKIGDNYNGYNLDKIYIVSKFYESVSQKTYRDWETINTALEIHYNILPGNG KSKADKVKKAVKNDLQKSITEINELVSNYKLCPPDNKAETYIHEISHILNNFE AQELKYNPEIHLVESELKASELKNVLDVIMNAFWCSVFMTEELVDKDNFYAE LEEIYDEIYTVISLYNLVRNYVTQKPYSTKKIKLNFGIPTLADGWSKSKEYSNN AIIIMRDNLYYLGI FNAKNKPDKKIIEGNTSENKGDYKMKIYNLLPGPNKMI PK VFLSSKTGVETYPKSAIILEGYKQNKHLKSSKDFDITFCHDLIDYFKNCIAIHP EWKNFGFDFSDTSTYEDISGFYREVELQGYKIDWTYISEKDIDLLQEKGLYLF QIYNKDFSKSTGNDNLHTMYLKNLFSEENLKDIVLKLNGEAEIFFRKSSIKNP IIHKKGSILVNRTYEAEEKDQFGNIQIVRKTIPENIYQELYKYFNDKSDKELSD EAKLKNVVGHEAATNIVKDYRYTYDKYFLHMPITINFKANKTSFINDRILQY IAKEKNLHVIGIDRGERNLIYVSVIDTCGNIVEQKSFNIVNGYDYQIKLKQOEG ARQIARQEWKEIGKIKEIKEGYLSLVIHEISKMVIKYNAI IAMEDLSYGFKKGR FKVERQVYQKFETMLINKLNYLVFKDISITENGGLLKGYQLTYIPDKLKNVGHQ CGCIFYVPAAYTSKIDPTTGfVNI FKFKDLTVDAKREFIKKFDSIRYDSEKKLF CFTFDYNNFITQNTVMSKSSWSVYTYGVRIKRRFVNGRFSNESDTIDITKMEK TLEMTDINWRDGHDLRQDIIDYEIVQHI FEIFRLTVQMRNSLSELED RDYDRLI SPVLNENNI FYDSAKAGDALPKDADANGAYCIAKGLYEIKQITENWKEDGKFS RDKLKISNKDWFDFIQNKRYL
SEQ ID NO: 6	MGKNQNFQEFIGVSPLOKTLRNELIPTETTKKNITQLDLLTEDEIRAQNREKLK EMDDYYRNVIDSTLHVGIADVWSYLFSCMRNHLRENSKESKRELERTQDSIRS QIHNFKAERADFKDMFGASIIITKLLPTYIKQNSEYSEYDESMEILKLYGKFTT SLTDYFETRKNIFSKEKISSAVGYRIVEENAEIFLQONQAYDRICKIAGLDLHG LDNEITAYVDGKTLKEVCSDEGFAKAITQEGIDRYNEAIGAVNQYMNLLCQKNK ALKPGQFKMKRLHKQILCKGTTSFDPKPFENDKQVYDAVNSFTEIVTKNNDLK RLNITQONANDYDMNKIYVVADAYSMSISQFISKKWNLIEECLLDYSDNLPKGK NAKENKVKKAVKEETYRSVSQLENEVIEKYYVEKTGQSVWKVESYISSLAEMI KL ELCHEIDNDEKHNLIEDDEKISEIKELLDMYMDVFHIIKVFRVNEVLNDFETFY SEMDEIYQDMQEI VPLYNHVRNYVTQKPYKQEKYRLYFHTPTLANGWSKSKEYD NNAIILVREDKYLLGILNAKKPSKEIMAGKEDCSEHAYAKMNYLLPGANKML PKVFLSKKGIQDYHPSSYIVEGYNEKKHIKSGKNFDIRFCRDLIDYFKECIKKH PDWNKFNFEFSATETYEDISVFYREVEKQGYRVEWYINSEDIQKLEEDGQLFL FQIYNKDFAVGSTGKPNLHTLYLKNLFSEENLRDIVLKLNGEAEIFFRKSSVQK

SEQ ID NO	Sequence
	PVIHKCGSILVNRTYEITESGTRVQSIPESEYMELYRYFNSEKQIELSDEAKK YLDKVQC�KAKTDIVKDYRTMDKFFIHLPIITINFKVDKGNVNVAIAQQYIAGR KDLHVIGIDRGERNLIYVSVIDMYGRILEQKSFNLVEQVSSQGTKRYDYKEKL QNREEERDKARQSWKTIGKIKELKEGYLSSVIHEIAQMVKYNAILAMEDLNYG FKRGRFKVERQVYQKFETMLISKLNLYLADKSQAVDEPGGILRGYQMTYVPDNIK NVGRQCGIIFYVPAAYTSKIDPTTGFINAFKRDVVSNDAKENFLMKFDSIQYD IEKGLFKFSFDYKNFATHKLTAKTKWDVYTNGTRIQNMKVEGHWLSMEVELTT KMKELLDDSHIPYEEGQNILDDLREMKDITTTIVNGILEIFWLTQLRNSRIDNP DYDRIISPVLNKNGEFFDSDEYNSYIDAQKAPLPIDADANGAFICIALKGMYTAN QIKENWVEGEKLPADCLKIEHASWLAFMQGERG
SEQ ID NO: 7	ATGAACAATGGCACCAACAATTTCCAGAACTTTATCGGAATTAGCAGTCTGCAA AAGACTCTCCGGAATGCCCTTATACCCACCGAGACAACCCAGCAGTTTCATCGTG AAAAACGGGATTATCAAGGAAGACGAGCTGCGCGGCGAAAATCGGCAAAATTTG AAAGATATAATGGACGATTATTACCGCGGTTTTATCTCTGAGACTCTGAGCTCC ATTGACGATATCGACTGGACCTCACTCTTCGAAAAGATGGAGATTTCAGCTTAAA AACGGCGATAATAAGGACACACTGATAAAAGAACAGGCTGAGAAGCGGAAAGCC ATCTATAAGAAATTTGCAGATGACGATCGCTTCAAGAACATGTTTAGCGCCAAA TTGATTAGTGACATCCTGCCGGAATTCGTTATTACACAATAACAATTACTCTGCT AGCGAGAAGGAAGAGAAAACCCAAGTCATAAAGCTCTTTTCCCGGTTCCGCACT TCATTTAAAGATTATTTCAAGAACCGCGCAAATTGCTTTAGCGCCGACGATATC AGTTCTAGCTCCTGTCATCGGATTGTGAACGACAATGCTGAAATCTTCTTTTCA AACGCCCTTGTATACCGCCGGATTGTGAAAAATCTGAGCAACGATGACATAAAT AAGATCAGTGGAGATATTAAAGACTCTTTGAAGGAGATGAGCCTGGAAGAGATC TATTCCTACGAAAAATATGGGGAGTTCAATTACCCAGGAAGGCATATCATTTTAC AACGATATCTGCGGTAAGGTTAATAGCTTCATGAACCTCTATTGTGAGAAAAAT AAGGAGAACAAAAATCTTTACAAGCTGCGCAAATTGCACAAGCAAATTCTGTGC ATCGCAGACACAAGTTATGAAGTCCCTTACAAATTTGAGTCTGATGAAGAGGTG TATCAGAGCGTAAACGGCTTCCTCGACAATATTTCTCAAAGCATATAGTGGAA CGGCTTCGCAAAATCGGAGATAACTACAATGGGTATAACCTGGACAAGATTTAC ATCGTTAGCAAATTTTATGAGAGTGTCTCTCAGAAGACCTACCGGGATTGGGAA ACTATTAATACCGCCTTGGAGATACACTATAACAATATCCTGCCCGGCAACGGT AAAAGCAAGGCTGACAAAGTGAAGAAAGCCGTAAAGAATGATCTCCAAAAATCC ATTACAGAAATCAACGAGCTTGTGTCAAATTACAAGCTGTGTCCGGACGATAAC ATTAAGCAGAAACCTATATACATGAGATCAGCCACATTTTGAATAACTTCGAA GCCCAGGAGCTGAAGTACAATCCAGAAATCCATCTCGTTGAGAGTGAACCTAAA GCTTCTGAGCTGAAGAACGTCTTGGACGTGATTATGAATGCCTTTCACTGGTGC AGCGTATTCATGACTGAAGAGCTGGTGGATAAAGACAACAATTTTATGCAGAA CTCGAGGAAATATACGATGAGATCTATACCGTTATTTCCCTTTACAACCTGGTC CGCAATTATGTGACACAGAAGCCCTACTCAACCAAAAAGATCAAATTGAACTTC GGCATTCCGACTCTGGCCGACGGATGGAGCAAGAGTAAAGAATATTCTAATAAC GCTATAATCCTCATGCGGGATAATCTTTACTATCTGGGGATTTTAAACGCCAAG AATAAACCTGACAAGAAAATCATTGAGGGCAACACCAGCGAAAATAAGGGTGAT TACAAAAAGATGATATATAACTTGCTGCCCGGCCGAATAAAATGATCCCAAAG GTATTCCTCTCCTCAAAAACAGGAGTGGAGACCTACAAGCCCAGCGCATATATT CTTGAAGGGTACAAACAAAACAAGCATCTGAAAAGTTCTAAGGACTTTGATATC ACTTTCTGTACGACTTGATTGATTATTTTAAAAATTGCATAGCCATCCATCCG GAGTGGAAAGAACTTCGGCTTTGACTTCAGCGATACCTCCACATACGAAGACATT TCAGGTTTTTATCGCGAGGTTGAACTGCAGGGCTACAAAATCGATTGGACCTAT ATTAGCGAGAAGGACATAGATCTCCTTCAGGAAAAAGGACAACCTGTACTTGTTT CAGATCTATAATAAGGACTTTAGTAAAAAGTCTACTGGGAACGATAATCTGCAC ACCATGTACCTCAAAAACCTTTTCAGCGAGGAAAATCTGAAGGACATTGTCTTG

SEQ ID NO	Sequence
	AAACTGAACGGCGAGGCTGAAATCTTTTCCGGAAGTCCTCAATTAAAAATCCT ATAATCCATAAGAAAGGTAGCATTCTCGTGAACCGCACATATGAGGCCGAAGAG AAGGATCAGTTTGGCAATATCCAAATTGTACGGAAAACCATACCCGAAAACATC TACCAGGAGCTTTATAAGTACTTCAATGACAAAAGTGATAAGGAACTGTCTGAC GAGGCAGCCAAATTGAAGAACGTGGTTGGACACCATGAAGCTGCCACTAATATT GTCAAAGATTATCGCTACACCTATGACAAGTACTTTCTGCACATGCCGATCACA ATTAACTTCAAAGCAAATAAGACCAGCTTTATAAACGATCGGATTCTCCAGTAT ATTGCCAAAGAGAAGAATCTTCATGTGATCGGGATTGACCGCGGCGAACGGAAC CTGATATACGTATCCGTGATCGATACTTGTGGTAATATTGTTGAGCAAAAATCA TTCAACATCGTCAATGGCTATGACTACCAGATTAAGTTGAAACAGCAAGAAGGA GCTCGCCAGATAGCCCGGCAGGAGTGGAAGGAAATCGGGAAAATTAAGGAGATC AAAGAAGGCTATCTGAGCCTCGTGATTACGAGATAAGTAAGATGGTAATCAAA TACAACGCAATTATCGCCATGGAAGATCTTCTTATGGTTTTAAGAAAGGCCGC TTCAAGGTGGAGCGGCAAGTTTACCAGAAATTTGAAACCATGCTGATTAATAAG TTGAACTATCTGGTCTTCAAAGACATAAGCATCACAGAGAATGGAGGGCTCCTT AAGGGCTACCAGCTGACCTATATTCAGATAAATTGAAGAACGTGGGTCATCAA TGCGGCTGTATCTTTTACGTACCGCTGCCTATACTTCCAAAATTGACCCGACC ACAGGATTCGTGAATATATTTAAGTTCAAAGATCTGACCGTTGACGCAAAGCGC GAATTTATCAAAAAGTTCGATTCAATTCGGTACGACAGCGAGAAAAAGCTCTTT TGCTTCACTTTTGATTATAACAATTTTCATCACCAGAACACAGTCATGAGTAAA TCTAGCTGGTCCGTGTACACCTATGGGGTACGCATTAAGCGGCGCTTTGTGAAT GGCCGGTTCTCAAACGAAAGCGACACTATAGATATCACCAGAACATGGAGAAG ACACTTGAAATGACCGATATTAATTGGCGCGACGGTCACGATCTGCGGCAGGAC ATCATTGATTACGAGATAGTTCAACATATCTTTGAAATTTTCCGCTTGACTGTC CAGATGCGGAACAGTCTGTCTGAGCTCGAAGACCGCGATTATGACCGGCTTATC AGCCCTGTGCTGAATGAGAACAATATTTTTTACGATTCGCCCAAAGCTGGCGAC GCCTTGCCCAAGGATGCAGACGCCAACGGAGCTTATTGTATAGCCCTGAAAGGG CTCTACGAAATCAAGCAGATTACCGAGAATTGGAAAGAAGATGGCAAGTTCTCA CGCGACAAACTTAAGATCAGCAACAAAGATTGGTTTGACTTCATTCAAAAATAAG CGGTATCTG
SEQ ID NO: 8	ATGGGCAAAAACCAAATTTCCAAGAATTTATCGGAGTGAGCCCCCTGCAGAAG ACCCTCCGGAACGAGCTTATTCCGACTGAGACCACAAAGAAAAATATAACCCAG CTGGACTTGCTGACTGAAGATGAGATCCGCGCCAGAACCGGGAAAAGCTCAAA GAGATGATGGACGATTATTACCGCAATGTTATTGACAGTACCCTTCACGTCGGG ATCGCTGTGGATTGGTCTTATCTGTTACGCTGCATGCGGAACCATTGTGCGGAA AATTCCAAGGAGTCAAACGCGGAAGTGGAGCGCACACAGGACAGCATTTCGGAGT CAGATACACAACAAGTTTGCCGAACGCGCAGATTTCAAAGACATGTTTGGCGCC TCTATCATTACCAAGCTCCTTCTACTTACATCAAACAAAATAGCGAGTATTCC GAACGGTACGATGAGTCAATGGAAATTCTGAAGTTGTATGGTAAATTCACCACA AGCCTGACCGACTACTTTGAGACTCGCAAGAACATATTCAGTAAAGAAAAGATC TCTAGCGCTGTAGGCTATCGGATTGTGGAGGAAAATGCCGAGATCTTTCTCCAG AACCAGAATGCATACGATCGCATTTGTAAAATAGCCGGACTTGACCTGCATGGG TTGGATAACGAAATCACCGCTTATGTTGACGGCAAGACACTGAAAGAGGTCTGC TCCGATGAAGGTTTCGCCAAGGCAATTACCCAAGAGGGCATCGACCGGTACAAT GAAGCCATTGGAGCTGTGAACCAGTATATGAATCTCCTTTGTGAGAAAACAAG GCCCTGAAACCCGGGCAATTTAAGATGAAACGCTTGACACAAGCAGATACTGTGC AAAGGCACCTACCTCATTCGATATCCCGAAGAAATTTGAGAATGACAAGCAGGTA TACGATGCAGTGAACAGCTTCACAGAAATTGTTACCAAAAATAACGACCTCAAG CGGCTTCTGAATATCACTCAAACGCCAATGATTATGACATGAACAAAATTTAC GTCGTGGCTGATGCCTATAGTATGATATCTCAGTTTATCAGCAAGAAATGGAAT TTGATTGAGGAATGTCTGCTCGACTACTATTCGATAACCTTCAGGTAAGGGC

SEQ ID NO	Sequence
	AATGCAAAAAGAGAACAAGGTAAAAAAGGCCGTGAAAGAAGAGACCTACCGCTCA GTTAGCCAGCTGAATGAAGTCATCGAGAAGTATTACGTGGAAAAAACAGGACAA AGTGTATGGAAGGTGGAGTCTTATATTAGCTCCTTGGCTGAAATGATAAACTG GAGCTCTGCCATGAAATCGACAACGATGAGAAGCACAATCTTATTGAAGACGAT GAGAAAATCTCAGAAATTAAGGAGCTGTTGGACATGTACATGGATGTTTTCCAT ATAATCAAAGTCTTTCGGGTGAACGAAGTACTGAATTTTCGACGAGACCTTTTAT AGCGAAATGGATGAGATTTACCAGGACATGCAGGAAATCGTGCCCCTCTATAAC CACGTTTCGCAATTACGTCACTCAAAAGCCGTATAAACAGGAGAAGTACCGGCTT TATTTCCATACCCCTACACTGGCCAACGGGTGGAGTAAATCTAAGGAATACGAT AATAACGCAATTATATTGGTGCGCGAGGACAAATATTACCTGGGCATCCTCAAT GCCAAGAAAAAGCCAGCAAAGAAATTATGGCTGGTAAGGAGGATTGTTCCGAA CACGCCTATGCAAAAATGAAGTACTATCTTCTGCCGGGCGCCAATAAGATGTTG CCAAAAGTATTTCTGTCAAAGAAAGGAATCCAGGACTACCATCCCAGCAGTTAT ATTGTGGAGGGGTACAACGAAAAGAAACACATAAAGGGCTCTAAAAATTTTCGAT ATCCGGTTTTTGCCGCGACCTCATTGATTATTTCAAGGAGTGTATCAAAAAGCAT CCGGAAGTGAACAAATTTAATTTTCAATTTAGCGCTACCGAGACTTACGAAGAT ATTTCCGTTTTCTATCGGGAGGTGCAAAAGCAAGGTTACCGCGTGGAGTGGACC TATATAAACTCAGAAGACATCCAGAACTTGAGGAAGATGGCCAGCTGTTTTTG TTCCAAATTTACAATAAGGACTTTGCCGTAGGAAGCACAGGGAAACCTAACCTG CACACCCTCTATCTTAAGAATCTGTTCAAGTGAAGGAAACTTGCGGGATATCGTG CTGAAACTCAATGGCGAGGCAGAAATTTTTTTTCCGCAAGTCTAGCGTTTCAGAAA CCCGTCATACATAAGTGCGGTTCCATCCTTGTGAACCGGACTTACGAGATTACC GAATCAGGCACAACCCGCGTACAGAGCATCCCGGAGAGTGAATATATGGAGCTG TACCGGTATTTTAATTCTGAAAAACAAATTGAGTTGAGCGACGAAGCCAAGAAA TACCTGGATAAGGTGCAGTGTAACAAAGCTAAGACTGACATAGTTAAAGATTAT CGCTACACCATGGACAAGTTCTTTATCCACCTCCCAATTACAATCAATTTCAA GTCGATAAGGGAAACAATGTGAACGCCATTGCACAGCAATATATAGCCGGGCGG AAAGACCTTCATGTAATCGGCATTGATCGCGGTGAGCGGAATCTGATCTACGTG TCCGTTATTGACATGTATGGCCGCATATTGGAACAGAAGTCATTTAACCTGGTC GAGCAGGTGAGCAGTCAAGGAACCAACGGTACTATGATTACAAGGAAAACTC CAGAATCGCGAGGAAGAGCGGGACAAGGCTCGCCAGTCTTGGAAGAACTATCGGG AAGATTAAAGAACTTAAGGAGGGCTATCTGAGCTCCGTAATCCACGAAATTGCC CAAATGGTGGTTAAATACAACGCAATAATCGCCATGGAGGATTTGAATTATGGT TTCAAGCGGGGCGCTTTAAAGTCGAACGGCAGGTGTACCAGAAGTTTCGAGACC ATGCTGATTTCAAACCTCAACTATCTTGCTGACAAGAGCCAAGCCGTAGATGAA CCCGGAGGGATTCTGCGCGGCTACCAGATGACATATGTGCCGGACAATATTA AACGTTGGTCGGCAGTGCGGCATAATCTTTTACGTCCCTGCAGCCTATACCAGT AAGATTGATCCCACTACCGGATTCATCAATGCTTTTAAACGCGACGTGGTATCT ACAAACGATGCCAAGGAGAATTTCTTGATGAAATTTGACAGCATTCAATACGAT ATAGAAAAGGGGCTGTTCAAATTTTCCTTCGACTATAAGAACTTTGCAACCCAT AAACTCACTCTTGCCAAGACCAAATGGGATGTGTACACAAATGGCACCCGGATT CAGAACATGAAGGTTGAGGGTCACTGGCTGTCAATGGAAGTCGAGTTGACTACC AAAATGAAGGAAGTCTCGACGATAGCCATATTCCGTATGAGGAAGGCCAGAAT ATCCTTGACGATCTGCGCGAGATGAAAGACATTACAACCATAGTGAACGGAATC TTGGAATTTTCTGGCTGACTGTACAACCTCGGAATAGTCGCATCGATAACCCA GACTACGATCGGATTATATCTCCCGTGCTTAATAAGAACGGGGAGTTTTTCGAC AGCGATGAATATAATTCTTACATCGACGCTCAGAAAGCCCCGCTGCCTATTGAT GCAGACGCCAACGGCGCTTTTTGTATCGCCTTGAAGGGTATGTATACCGCAAAT CAGATTAAAGAGAACTGGGTGTAAGGCGAGAAGCTGCCCCGCGATTGCCTCAA ATAGAACACGCTTCATGGCTTGCTTCATGCAAGGAGAGCGCGG

[0128] A Cas12a (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) protein of the present disclosure comprises one or more of the following stretches of amino acid residues: LIPTETT (SEQ ID NO: 12), MDDYYR (SEQ ID NO: 13), NAEIF (SEQ ID NO: 14), CQKNK (SEQ ID NO: 15), LHKQILC (SEQ ID NO: 16), KVKKAVK (SEQ ID NO: 17), VRNYVTQKPYPY (SEQ ID NO: 18), GWSKSKEY (SEQ ID NO: 19), DLIDYFK (SEQ ID NO: 20), DIVLKLNGEAEIFFRKSS (SEQ ID NO: 21), GSILVNRTYE (SEQ ID NO: 22), ELSDEA (SEQ ID NO: 23), IVKDYRYT (SEQ ID NO: 24), LHVIGIDRGERNLIYVSVID (SEQ ID NO: 25), KYNAIIAMEDL (SEQ ID NO: 26), GRFKVERQVYQKFETMLI (SEQ ID NO: 27), IFYVPAAYTSKIDPTTGF (SEQ ID NO: 28), or ISPVLN (SEQ ID NO: 29).

[0129] The present disclosure also provides a method of improving the cleaving efficiency of a type V CRISPR-associated protein, wherein the method comprises providing a type V CRISPR-associated protein, identifying a residue at a critical position, such as position 925 of SEQ ID NO: 1 or position 930 of SEQ ID NO: 3, and mutating the residue at the critical position to a Lysine, thereby improving cleaving efficiency of the endonuclease.

[0130] The metagenomically mined Cas12a proteins of the present disclosure (*e.g.*, mgCas12a-1 or mgCas12a-2) exhibit superior editing efficiency as compared to other Cas12a orthologs (*e.g.*, AsCas12a, FnCas12a, or LbCas12a). For example, in some cases, the metagenomically mined Cas12a proteins of the present disclosure exhibit at least 1.5 fold, at least 2 fold, at least 5 fold, at least 10 fold, at least 15 fold, at least 20 fold, at least 25 fold, at least 30 fold, at least 35 fold, at least 40 fold, at least 45 fold, at least 50 fold, at least 60 fold, at least 70 fold, at least 80 fold, at least 90 fold, at least 100 fold, at least 150 fold, at least 200 fold, at least 250 fold, at least 300 fold, at least 350 fold, at least 400 fold, at least 450 fold, at least 500 fold, at least 600 fold, at least 700 fold, at least 800 fold, at least 900 fold, at least 1000 fold, at least 2000 fold, from 1.5 to 5 fold, from 5 to 10 fold, from 10 to 15 fold, from 15 to 20 fold, from 20 to 25 fold, from 25 to 30 fold, from 30 to 35 fold, from 35 to 40 fold, from 40 to 45 fold, from 45 to 50 fold, from 50 to 60 fold, from 60 to 70 fold, from 70 to 80 fold, from 80 to 90 fold, from 90 to 100 fold, from 100 to 150 fold, from 150 to 200 fold, from 200 to 250 fold, from 250 to 300 fold, from 300 to 350 fold, from 350 to 400 fold, from 400 to 450 fold, from 450 to 500 fold, from 500 to 600 fold, from 600 to 700 fold, from 700 to 800 fold, from 800 to 900 fold, from 900 to 1000 fold, from 1000 to 2000 fold, from 1.5 to 2000 fold, from 1.5 to 20 fold, from 5 to 10 fold, from 10 to 40 fold, from 20 to 40 fold, from 5 to 15 fold, from 5 to 100 fold, from 50 to 80 fold, from 2 to 4 fold, from 4 to 8 fold, from 8 to 10 fold, or from 10 to 14 fold more efficient genome editing than other Cas12a orthologs, such as FnCas12a, AsCas12, and/or LbCas12a. Editing efficiency can be measured by any known methods of quantification, such as measuring percent indels or amplicon targeted deep sequencing.

[0131] It is contemplated that metagenomically mined Cas12a proteins of the present disclosure can exhibit low random DNase activity, at least similar random DNase activity to other Cas12a orthologs, or even lower random DNase activity than other Cas12a orthologs, depending on the reaction conditions, while also achieving higher editing efficiency than other Cas12a orthologs or while maintaining comparable editing efficiency to other Cas12a orthologs. For example, as shown in FIG. 24, and FIG. 25A, dsDNA substrate and cleaved product were almost entirely degraded, potentially due to random DNase activity of certain Cas12a orthologs (e.g., FnCas12a, LbCas12a) after incubation of the reaction for 12 hours and 24 hours, while both substrate and cleaved products were detected in the 12 hour reaction with metagenomically mined Cas12a proteins of the present disclosure (WT mg-1 and WT mg-2). FIG. 25B shows a graph of time versus dsDNase activity of each Cas12a, demonstrating that the target substrate dsDNA remains at later time points for mgCas12a-1, indicating lower random DNase activity of the mgCas12a-1.

[0132] In some embodiments, metagenomically mined Cas12a proteins (e.g., mgCas12a-1 or mgCas12a-2) of the present disclosure display cleavage activity in the presence of a wide range of divalent cations. For example, metagenomically mined Cas12a proteins (e.g., mgCas12a-1 or mgCas12a-2) of the present disclosure display can display cleavage activity in the presence of CaCl₂, CoCl₂, FeCl₂, MnSO₄, or any combination thereof.

[0133] TABLE 2 below shows sequences of other Cas12a proteins including *Acidaminococcus* sp. Cas12a (AsCas12a) as set forth in SEQ ID NO: 9, *Lachnospiraceae* sp. (LbCas12a) as set forth in SEQ ID NO: 10, and *Francisella tularensis* subsp. *novicidia* (FnCas12a) as set forth in SEQ ID NO: 11.

TABLE 2 – Additional Cas12a Proteins

SEQ ID NO	Sequence
SEQ ID NO: 9	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQQFIEEDKARNDHYKELKPI IDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAI HDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALL RSFDKFTTYFSGFYENRKNVFS AEDISTAI PHRIVQDNFPKF KENCHI FTRLI TAVPSLREHFENVKKAIGIFVSTSI EEVFSFPFYNQLLTQTQIDLYNQLLGGI SREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSF ILEEFKSDEEVIQS FCKYKTL LRNENVLETAEALF NELNSIDLTHIFISHKKL ETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEII SAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQDLSLGL YHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKPKYSVEK FKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPT EKTSEGFDMYYDYFPDAAKMIPK CSTQLKAVTAHFQTH TPIILLSNNFIEPL EITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTK TTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYL FQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAE LFYRPKSRM KRMAHRLGEKMLNKKLKDQKTPIDTLYQELYDVVNHRLSHDLSDEARALLPN

SEQ ID NO	Sequence
	VITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPE TPIIGIDRGERNLIYITVIDSTGKILEQSRSLNTIQQFDYQKKLDNREKERVAA RQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVLENLNFQFKSKRTGIA EKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPNYQLTDQFTSFAKMGTSQS FLFYVPAPYTSKIDPLTGFDVDFVWKTIKNHESRKHFLLEGFDLHYDVKTGDF ILHFKMNRNLSFQRLPGFMPAWDIVFEKNETQFQDAKGTPFIAGKRIVPVIEN HRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDSDSHAITMVALI RSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPWPMDADANGAYHIAL KGQLLNLHLKESKDLKLQNGISNQDWLAYIQELRN
SEQ ID NO: 10	AASKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLEVEDEKRAEDYKGVKK LLDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEI AKAFKGAAGYKSLFKKDIETILPEAADDKDEIALVNSFNGFTTAFSTGFFDNR ENXFSEEAKSTSIAFRFCINENLTRYISNXDIFEKVDAIFDKHEVQEIKEKILN SDYDVEDFFEGEFFNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNA KTKQALPKFKPLYKQVLSDRSLSFYGEGYTSDEEVLEVFRNTLNKNSEIFSS IKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNLIRDKWNAEYDDIH LKKKAVVTEKYEDDRRSFKKIGSFSLEQLQEYADADLSVVEKLKEIIIQKVD EIIYKVYGSSEKLFDAADFVLEKSLKKNDVVAIXKDLLDSVKSFENYIKAFFGE GKETNRDESFYGDFVLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLYFQNPQ FXGGWDKDKETDYRATILRYGSKYYLAIXDKKYAKCLQKIDKDDVNGNYEIKIN YKLLPGPNKXLPKVFFSKWXAYNPSEDIQKIYKNGTFFKGDVFNLDCHKL IDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQGYKVSFESASKK EVDKLVEEGKLYXFQIYNKDFSDKSHGTPNLHTXYFKLLFDENNHGQIRLSGG AELFXRRASLKKEELVVHPANSPIANKNPDNPKTTTSLSYDVYKDKRFSEDQY ELHIPIAINKCPKNI FKINTEVRVLLKHDDNPYVIGIDRGERNLLYIVVVDGK GNIVEQYSLNEIINNENFNGIRIKTDYHSLLDKKEKERFEARQNWTSIENIKELK AGYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKXLIDK LNYXVDKKSNPCATGGALKGYQITNKFESFKSXSTQNGFIIFYIPAWLTSKIDP STGTFVNLLKTKYTSIADSKKFISSFDRIXVPEEDLFEFALDYKNFSRTDADY IKKWKLYSYGNRIRIFAANKNNVFAWEEVCLTSAYKELFNKYGINYQQGDIR ALLCEQSDKAFYSSFALXSLXLQXRNSITGRTDVDFLISPVKNSDGIIFYDSR NYEAQENAILPKNADANGAYNIARKVLWAIQGFKADEKLDKVKIAISNKEW LEYAQTSVK
SEQ ID NO: 11	MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQI IDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIK KQISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDI TDIDEALEI IKSFKGWTTYFKGFHENRKNVYSSNDIPTSIYRIVDDNLPKFL ENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVFE IANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKK YKMSVLFKQILSDTESKSFVIDKLEDDSDVVTMQSFYEQIAAFKTVEEKS IKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQQ IAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALFENKHRDIDKQCRFEE ILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKDLLQASAEDDVKAIKDLL DQTNLLHLKLIKFIHSQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIR NYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYLLGVMNKK NNKIFDDKAIKENKGEYKIVYKLLPGANKMLPKVFFSAKSIKFYNPSEDI L RIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSIKHPPEWKDFGFRFS DTQRYNSIDEFYREVENQGYKLTFFENISESYIDSVVNQGKLYLFIYKNKDFSA YSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSI PKKITHPAKEA IANKNKDNPKKESVFEYDLIKDKRFTEDKFFHCPITINFKSSGANKFNDEIN LLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYH

SEQ ID NO	Sequence
	DKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVEYNAIVVFED LNFGFKRGRFKVEKQVYQKLEKMLIEKLNLYLVFKDNEFDKTGGVLRAYQLTAP FETFKKMGKQTGIIYYVPAGFTSKICPVTGTFVNQLYPKYESVSKSQEFFSKFD KICYNLDKGYFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTR EVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTILQMRN SKTGTELDYLI SPVADVNGNFFDSRQAPKNMPQDADANGAYHIGLKGLMLLGR IKNNQEGKKLNLVIKNEEYFEFVQNRNN

Purification Tags

[0134] In some embodiments, it is contemplated that the Cas12a proteins of the present disclosure form heterologous proteins, fused N-terminally or C-terminally to a purification tag. A Cas12a protein of the present disclosure (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) includes a purification tag incorporated at its N-terminus or C-terminus. A purification tag is incorporated at the N-terminus or C-terminus of a Cas12a protein of the present disclosure to provide a recombinant polypeptide that are easily purified. Said purification tags, thus, facilitate purification of the Cas12a proteins.

[0135] For example, a Cas12 protein of the present disclosure (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) has a His-tag, such as a 6xHis or a poly-His-tag. Said His-tags are capable of binding to several metal ions such as nickel, cobalt, and copper. Thus, a His-tagged Cas12a protein of the present disclosure (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) is rapidly purified using metal affinity chromatography. Alternatively, the present disclosure provides a Cas12a protein (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) having a FLAG®-tag, which comprises the FLAG® peptide having a sequence of DYKDDDDK (SEQ ID NO: 30). Said FLAG®-tag is recognized by an anti-FLAG® antibody. Thus, a FLAG®-tagged Cas12a protein of the present disclosure (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) is rapidly purified using affinity chromatography.

[0136] A Cas12a protein of the present disclosure (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) has a purification tag selected from the group consisting of a His tag, a FLAG tag, an AU1 epitope tag, an AU5 epitope tag, a bacteriophage T7 tag, a bacteriophage V5 epitope tag, a Bluetongue virus tag (B-tag), a Glu-Glu tag (EE-tag), an HSV epitope tag, a KT3 epitope tag, a Myc epitope tag, a PDZ ligand tag, a polyarginine tag, a polyaspartate tag, a polycysteine tag, a polyphenylalanine tag, a protein C tag, an S1-tag, an S-tag, a Step-tag, and a VSV-G tag.

Guide RNA for Cas12a Endonucleases

[0137] Cas12a endonucleases of the present disclosure (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) is coupled to a crRNA sequence, which guides the Cas12a endonuclease to a nucleic acid sequence of interest. Thus, the crRNA is also referred to as guide RNA, or gRNA for short. The crRNA sequence is reverse complementary to a nucleic acid sequence of interest, for example, a eukaryotic nucleic acid sequence. Eukaryotic nucleic acid sequence can be a human nucleic acid sequence or a plant nucleic acid sequence. In other words, the crRNA targets Cas12a endonucleases described herein to a region of interest in the genome. Specifically, crRNAs are designed to be reverse complementary to a nucleic acid sequence that exists solely in cancer cells, and not in normal cells. In some cases, the nucleic acid sequence existing solely in cancer cells, may only have a one base difference from a nucleic acid sequence existing in normal cells. In some cases, the nucleic acid sequence existing solely in cancer cells may comprise a portion of a gene or may comprise a gene, which has been partially substituted or loss. For example, a nucleic acid sequence that exists solely in cancer cells may be a nucleic acid sequence in cancer cells comprising a single nucleotide polymorphism (SNP). A guide RNA having a reverse complementary sequence to a nucleic acid sequence existing solely in cancer cells bind said nucleic acid sequence and will not bind to off-target nucleic acid sequences or nucleic acid sequences not in cancer cells.

[0138] The guide RNA, also referred to as the crRNA sequence, used to guide a Cas12a endonuclease of the present disclosure is from 10 to 60 nucleotides long, from 10 to 15 nucleotides long, from 15 to 20 nucleotides long, from 20 to 25 nucleotides long, from 25 to 30 nucleotides long, from 30 to 35 nucleotides long, from 35 to 40 nucleotides long, from 40 to 45 nucleotides long, from 45 to 50 nucleotides long, from 50 to 55 nucleotides long, from 55 to 60 nucleotides long, from 1 to 100 nucleotide long, from 1 to 200 nucleotides long, from 1 to 200 nucleotides long, from 1 to 300 nucleotides long, from 1 to 50 nucleotides long, from 50 to 100 nucleotides long, from 100 to 150 nucleotides long, from 150 to 200 nucleotides long, from 200 to 250 nucleotides long, from 250 to 300 nucleotides long, from 300 to 350 nucleotides long, from 350 to 400 nucleotides long, from 400 to 450 nucleotides long, or from 450 to 500 nucleotides long. The crRNA sequence used to guide a Cas12a endonuclease of the present disclosure is about 10 nucleotides in length, about 15 nucleotides in length, about 20 nucleotides in length, about 25 nucleotides in length, about 30 nucleotides in length, about 35 nucleotides in length, about 40 nucleotides in length, about 45 nucleotides in length, about 50 nucleotides in length, about 55 nucleotides in length, about 60 nucleotides in length, about 65 nucleotides long, about 70 nucleotides long, about 75 nucleotides long, about 80 nucleotides long, about 85 nucleotides long,

about 90 nucleotides long, about 95 nucleotides long, about 100 nucleotides long, about 150 nucleotides long, about 200 nucleotides long, about 250 nucleotides long, about 300 nucleotides long, about 350 nucleotides long, about 400 nucleotides long, about 450 nucleotides long, or about 500 nucleotides long. The crRNA sequence used to guide a Cas12a endonuclease of the present disclosure is about 42 nucleotides in length.

[0139] The metagenomically mined Cas12a proteins of the present disclosure possess flexible and superior properties due to their ability to maintain cleavage activity when complexed with crRNAs having 5' repeat recognition sequences (the nucleotides upstream of the step-loop region of the crRNA) of AsCas12a, FnCas12a, and/or LbCas12a. For example, a Cas12a protein of the present disclosure (*e.g.*, mgCas12a-1 or mgCas12a-2) can be complexed with a crRNA having a 5' handle of crRNA complexed to AsCas12a, FnCas12a, and/or LbCas12a, and still retain cleavage activity of target nucleic acids.

Delivery of CRISPR/Cas Systems

[0140] It is further contemplated that Cas12a and the guide RNA can form a CRISPR/Cas complex, and such formed complex can be delivered to the target cell via various methods. One exemplary method includes plasmids or viral vectors comprising a nucleic acid encoding the endonuclease and guide RNA. Exemplary viral vector-based delivery methods include lentiviral-based delivery or adeno-associated virus (AAV)-based delivery. Alternatively, CRISPR/Cas complex disclosed herein can be delivered using ribonucleoproteins (RNPs) for delivery of the intact protein assembled to the gRNA. Intact proteins assembled to the gRNA are then delivered directly to cells.

Applications

[0141] A CRISPR/Cas complex, or CRISPR/Cas12a endonuclease complex (or CRISPR/Cas12a complex), described herein is used to edit the genome of a nucleic acid sequence. The nucleic acid sequence is a mammalian nucleic acid sequence. For example, the nucleic acid sequence is a human nucleic acid sequence. The nucleic acid sequence is also a non-human nucleic acid sequence. The nucleic acid sequence is from any animal. For genome editing, the crRNA, also referred to as the guide RNA (gRNA), is reverse complementary to a nucleic acid sequence of interest. For example, the crRNA is reverse complementary to a human nucleic acid sequence.

[0142] For example, a CRISPR/Cas12a complex can be used for genome editing of a nucleic acid sequence in a cancer cell, which comprises cancer-specific sequences (*e.g.*, single nucleotide polymorphisms (SNPs) or cancer specific mutations) identified by sequence analysis of genomes of various cell tissues. crRNA sequences are synthesized that are reverse complementary to these nucleic acid sequences, which exist uniquely in cancer cells. CRISPR/Cas12a complexes can be

generated using the crRNA reverse-complementary to the cancer-specific sequence, and administered to a subject having a cancer in a dose and schedule effective to treat the tumor cells as a customized therapy.

[0143] The cancer that can be treated with the CRISPR/Cas complexes of the present disclosure can include bladder cancer, bone cancer, blood cancer, breast cancer, black colored tumor, thyroid cancer, parathyroid cancer, bone marrow cancer, rectal cancer, laryngopharyngeal cancer, laryngeal cancer, lung cancer, esophagus cancer, pancreatic cancer, colorectal cancer, gastric cancer, tongue cancer, skin cancer, brain tumor, uterine cancer, head or neck cancer, gallbladder cancer, oral cancer, colon cancer, cancer around the anus area, central nervous system tumor, liver cancer, or colorectal cancer. In particular instances, the cancer comprises gastric cancer, colorectal cancer, liver cancer, lung cancer, or breast cancer. As such, the cancer cell includes a bladder cancer cell, a bone cancer cell, a blood cancer cell, a breast cancer cell, a black colored tumor cell, a thyroid cancer cell, a parathyroid cancer cell, a bone marrow cancer cell, a laryngopharyngeal cancer cell, a laryngeal cancer cell, a lung cancer cell, an esophagus cancer cell, a pancreatic cancer cell, a colorectal cancer cell, a gastric cancer cell, a tongue cancer cell, a skin cancer cell, a brain tumor cell, a uterine cancer cell, a head or neck cancer cell, a gallbladder cancer cell, an oral cancer cell, a central nervous system tumor cell, or a liver cancer cell.

[0144] Alternatively, and/or additionally, a nucleic acid sequence targeted by the RNA-guided nucleases of the present disclosure include regions in genes of KRAS, BRCA1, HER2/neu, PD-1, TCR, p53, CCR5, DNMT1, EMX1, and LKB1.

[0145] For example, CRISPR/Cas12a complexes of the present disclosure include crRNAs designed to be reverse complementary to exon 10 or exon 11 of BRCA1, which is implicated in ovarian cancer and breast cancer. Two or more gRNAs that target exon 11 of BRCA1 are used in a CRISPR/Cas12a complex of the present disclosure. Thus, combinations of gRNA are selected based on the purpose of the cancer treatment or the type of cancer.

[0146] Disclosed herein are also methods of using the RNA-guided Cas12a endonuclease of the present disclosure for treating, curing, or ameliorating a symptom of cancer or the cancer itself. For example, said method includes identifying a subject in need thereof having a symptom of cancer and administering a CRISPR/Cas12a complex disclosed herein. Several cancer indications are consistent with the compositions disclosed herein. The present disclosure provides a Cas12a protein that is used in a CRISPR/Cas12a complex for genome editing of a nucleic acid sequence in a cancer cell from any cancer set forth in <https://www.cancer.gov/types>. For example, a Cas12a protein of the present disclosure (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3 or any sequence related thereto as disclosed herein) is used in a CRISPR/Cas12a complex for genome editing of a target nucleic acid

sequence in cancer cell from a cancer such as acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), adolescents, cancer in, adrenocortical carcinoma, childhood adrenocortical carcinoma, aids-related cancers, kaposi sarcoma (soft tissue sarcoma), aids-related lymphoma (lymphoma), primary cns lymphoma (lymphoma), anal cancer, appendix cancer, astrocytomas, childhood (brain cancer), atypical teratoid/rhabdoid tumor, childhood, central nervous system (brain cancer), basal cell carcinoma of the skin, bile duct cancer, bladder cancer, childhood bladder cancer, bone cancer (includes ewing sarcoma and osteosarcoma and malignant fibrous histiocytoma), brain tumors, breast cancer, childhood breast cancer, bronchial tumors, childhood, burkitt lymphoma, carcinoid tumor (gastrointestinal), childhood carcinoid tumors, carcinoma of unknown primary, childhood carcinoma of unknown primary, cardiac (heart) tumors, childhood, central nervous system, atypical teratoid/rhabdoid tumor, childhood (brain cancer), embryonal tumors, childhood (brain cancer), germ cell tumor, childhood (brain cancer), primary cns lymphoma, cervical cancer, childhood cervical cancer, childhood cancers, cancers of childhood, unusual, cholangiocarcinoma, chordoma, childhood, chronic lymphocytic leukemia (CLL), chronic myelogenous leukemia (CML), chronic myeloproliferative neoplasms, colorectal cancer, childhood colorectal cancer, craniopharyngioma, childhood (brain cancer), cutaneous t-cell lymphoma, ductal carcinoma in situ (DCIS), embryonal tumors, central nervous system, childhood (brain cancer), endometrial cancer (uterine cancer), ependymoma, childhood (brain cancer), esophageal cancer, childhood esophageal cancer, esthesioneuroblastoma (head and neck cancer), ewing sarcoma (bone cancer), extracranial germ cell tumor, childhood, extragonadal germ cell tumor, eye cancer, childhood intraocular melanoma, intraocular melanoma, retinoblastoma, fallopian tube cancer, fibrous histiocytoma of bone, malignant, and osteosarcoma, gallbladder cancer, gastric (stomach) cancer, childhood gastric (stomach) cancer, gastrointestinal carcinoid tumor, gastrointestinal stromal tumors (GIST) (soft tissue sarcoma), childhood gastrointestinal stromal tumors, germ cell tumors, childhood central nervous system germ cell tumors (brain cancer), childhood extracranial germ cell tumors, extragonadal germ cell tumors, ovarian germ cell tumors, testicular cancer, gestational trophoblastic disease, hairy cell leukemia, head and neck cancer, heart tumors, childhood, hepatocellular (liver) cancer, histiocytosis, langerhans cell, hodgkin lymphoma, hypopharyngeal cancer (head and neck cancer), intraocular melanoma, childhood intraocular melanoma, islet cell tumors, pancreatic neuroendocrine tumors, kaposi sarcoma (soft tissue sarcoma), kidney (renal cell) cancer, langerhans cell histiocytosis, laryngeal cancer (head and neck cancer), leukemia, lip and oral cavity cancer (head and neck cancer), liver cancer, lung cancer (non-small cell and small cell), childhood lung cancer, lymphoma, male breast cancer, malignant fibrous histiocytoma of bone and osteosarcoma, melanoma, childhood melanoma, melanoma, intraocular

(eye), childhood intraocular melanoma, merkel cell carcinoma (skin cancer), mesothelioma, malignant, childhood mesothelioma, metastatic cancer, metastatic squamous neck cancer with occult primary (head and neck cancer), midline tract carcinoma with nut gene changes, mouth cancer (head and neck cancer), multiple endocrine neoplasia syndromes, multiple myeloma/plasma cell neoplasms, mycosis fungoides (lymphoma), myelodysplastic syndromes, myelodysplastic/myeloproliferative neoplasms, myelogenous leukemia, chronic (CML), myeloid leukemia, acute (AML), myeloproliferative neoplasms, chronic, nasal cavity and paranasal sinus cancer (head and neck cancer), nasopharyngeal cancer (head and neck cancer), neuroblastoma, non-hodgkin lymphoma, non-small cell lung cancer, oral cancer, lip and oral cavity cancer and oropharyngeal cancer (head and neck cancer), osteosarcoma and malignant fibrous histiocytoma of bone, ovarian cancer, childhood ovarian cancer, pancreatic cancer, childhood pancreatic cancer, pancreatic neuroendocrine tumors (islet cell tumors), papillomatosis (childhood laryngeal), paraganglioma, childhood paraganglioma, paranasal sinus and nasal cavity cancer (head and neck cancer), parathyroid cancer, penile cancer, pharyngeal cancer (head and neck cancer), pheochromocytoma, childhood pheochromocytoma, pituitary tumor, plasma cell neoplasm/multiple myeloma, pleuropulmonary blastoma, pregnancy and breast cancer, primary central nervous system (CNS) lymphoma, primary peritoneal cancer, prostate cancer, rectal cancer, recurrent cancer, renal cell (kidney) cancer, retinoblastoma, rhabdomyosarcoma, childhood (soft tissue sarcoma), salivary gland cancer (head and neck cancer), sarcoma, childhood rhabdomyosarcoma (soft tissue sarcoma), childhood vascular tumors (soft tissue sarcoma), ewing sarcoma (bone cancer), kaposi sarcoma (soft tissue sarcoma), osteosarcoma (bone cancer), soft tissue sarcoma, uterine sarcoma, sézary syndrome (lymphoma), skin cancer, childhood skin cancer, small cell lung cancer, small intestine cancer, soft tissue sarcoma, squamous cell carcinoma of the skin -, squamous neck cancer with occult primary, metastatic (head and neck cancer), stomach (gastric) cancer, childhood stomach (gastric) cancer, t-cell lymphoma, cutaneous, testicular cancer, childhood testicular cancer, throat cancer (head and neck cancer), nasopharyngeal cancer, oropharyngeal cancer, hypopharyngeal cancer, thymoma and thymic carcinoma, thyroid cancer, transitional cell cancer of the renal pelvis and ureter (kidney (renal cell) cancer), childhood cancer of unknown primary, unusual cancers of childhood, ureter and renal pelvis, transitional cell cancer (kidney (renal cell) cancer, urethral cancer, uterine cancer, endometrial, uterine sarcoma, vaginal cancer, childhood vaginal cancer, vascular tumors (soft tissue sarcoma), vulvar cancer, wilms tumor, other childhood kidney tumors, or any combination thereof.

[0147] The present disclosure additionally provides a method of genome editing comprising the steps of contacting a cell with a CRISPR/Cas12a described herein, allowing the guide RNA to bind

to the human nucleic acid sequence of interest, and cleaving the human nucleic acid sequence. CRISPR/Cas12a endonucleases described herein are administered to a plurality of cells *ex vivo*, thereby allows for the generation of engineered cells. Said engineered cells are administered as a cell therapy to a subject in need thereof, wherein the subject is a human and has any of the above described cancers. CRISPR/Cas12a endonucleases described herein are directly administered to a subject in need thereof. Direct administration comprises intravenous administration, subcutaneous administration, intramuscular administration, oral administration, or mucosal administration.

[0148] In other cases, the crRNA is reverse complementary to a plant nucleic acid sequence. For example, a Cas12a protein of the present disclosure is used in a CRISPR/Cas12a complex for genome editing of a nucleic acid sequence in a plant cell. Non-limiting examples of plant cells of interest include rice and *N. benthamiana*.

[0149] The present disclosure additionally provides a method of improving the cleaving efficiency of a type V CRISPR-associated protein. For example, the methods provided herein include identifying a residue at a position corresponding to residue 925 or residue 930 of a type V CRISPR-associated protein and mutating these residues to Lysine. Either the mutation to Lysine at residue 925, the mutation to Lysine at residue 930, or both, can improve the cleaving efficiency of the type V CRISPR-associated protein. Lysines at residues corresponding to 925 or 930 of a type V CRISPR-associated protein (*e.g.*, SEQ ID NO: 1 or SEQ ID NO: 3) disclosed herein are critical for maintaining function. Thus, Lysines (Lys) may be introduced into a type V CRISPR-associated protein at residues corresponding these positions to improve the function of a type V CRISPR-associated protein that doesn't have a Lysine at positions corresponding to residues 925 or 930.

[0150] Turning to the figures, one sees the following. **FIG. 1** shows the process by which a CRISPR associated protein (CAS protein) of the present disclosure was identified by metagenomic mining. CRISPR sequences are found using metaCRT and gene prediction is carried out using Prodigal. Cas12a candidate proteins are identified using local BLASTp of Cas12a sequences from Uniprot and Genbank. Next, sequences of 800 amino acids to 1500 amino acids are identified. Cas 12a candidates beside Cas1 are identified, where Cas1 candidates are identified from local BLASTp analysis of Cas1 sequences from Uniprot. Phylogenies were built using MEGA7 and sequences were annotated with web BLASTp. Putative Cas12a proteins were identified.

[0151] **FIG. 2** is a cladogram in which SEQ ID NO: 3 (mg Cas12a-2) and SEQ ID NO: 1 (mgCas12a-1) resolve at the top from FnCas12a, which is related to LbCas12a, both of which are related to AsCas12a. **FIG. 2** additionally shows the relationship between many Cas12a proteins. The methods of metagenomics mining disclosed herein are, thus, capable of identifying the

relationship between Cas12a proteins and excavating new Cas12a proteins. **FIG. 2** additionally shows bootstrap values. mgCas12a-2 and mgCas12a-1 showed a bootstrap value of 52.

[0152] FIG. 3 through **FIG. 8** show an alignment of proteins selected from the phylogenetic tree of **FIG. 2** including SEQ ID NO: 1 (mgCas12a-1, CDY), SEQ ID NO: 3 (mgCas12a-2, CDZ), SEQ ID NO: 9 (AsCas12a, WP_021), SEQ ID NO: 10 (LbCas12a, WP_035), and SEQ ID NO: 11 (FnCas12a, WP_003). Residues that are absolutely conserved are back shaded. Predictive structures including alpha helices and beta sheets are shown at the top of the alignment. Unidentified positions are shown with dots. **FIG. 7** shows a wedge indicating a critical residue in mgCas12a-1 (position 925) and mgCas12a-2 (position 930), which has been substituted from a K residue to a Q residue.

[0153] FIG. 9A shows a chart of characteristics of Cas12a proteins including AsCas12a, which is from the species *Acidaminococcus sp.* and has a nucleotide length of 3,921, an amino acid length of 1,307, a PAM sequence of 5'-TTTN-3', a sequence identity (in comparison to AsCas12a) of 100%, and a critical residue identity of 100% (33). The chart also shows the LbCas12a protein, which is from the species *Lachnospiraceae sp.* and has a nucleotide length of 3,684, an amino acid length of 1,228, has a PAM sequence of 5'-TTTN-3', a sequence identity (in comparison to AsCas12a) of 33.41%, and a critical residue identity of 79% (26). The chart also shows the FnCas12a protein, which is from the species *Francisella tularensis subsp. novicida* and has a nucleotide length of 3,900, an amino acid length of 1,300, a PAM sequence of 5'-TTTN-3', a sequence identity (in comparison to AsCas12a) of 34.45%, and a critical residue identity of 79% (26). The chart also shows the mgCas12a-1 (SEQ ID NO: 1) protein, which is from metagenome CDYX01038443.1 and has a nucleotide length of 3,789, an amino acid length of 1,263, a speculative PAM sequence of 5'-TTTN-3', a sequence identity (in comparison to AsCas12a) of 32.65%, and a critical residue identity of 88% (29). The chart also shows the mgCas12a-2 (SEQ ID NO: 3) protein, which is from metagenome CDZH01035208.1 and has a nucleotide length of 3,825, an amino acid length of 1,275, a speculative PAM sequence of 5'-TTTN-3', a sequence identity (in comparison to AsCas12a) of 32.65%, and a critical residue identity of 88% (29). **FIG. 9B** shows a chart of various Cas12 orthologs and the percent sequence identity between different Cas12a orthologs. The top row, from left to right, includes Cas12a, As, Lb, Fn, mg-1, and mg-2. The second row, from left to right, includes As followed by 5 blank cells. The third row, from left to right, includes Lb, 33.4, followed by 4 blank cells. The fourth row, from left to right, includes Fn, 34.2, 40.4, followed by 3 blank cells. The fifth row, from left to right, includes mg-1, 32.4, 35.3, 36.6, followed by 2 blank cells. The last row, from left to right, includes mg-2, 30.7, 34.9, 35.9, 52.5, followed by 1 blank cell.

[0154] FIG. 10 shows gel electrophoresis of target dsDNA after exposure to mgCas12a-1 (SEQ ID NO: 1) or mgCas12a-2 (SEQ ID NO: 3) complexed with crRNA in no buffer, NEBuffer 1.1, NEBuffer 2.1, and NEBuffer 3.1 for a 10 minute reaction run at 37°C. Cleaved products are indicated by an arrow. The target dsDNA (LsXTb12) resolves ~2.2 kbp. Cleavage was found to occur in the presence of 3 pmol mgCas12a-1 protein, FnCas12a crRNA #1 at a ratio of 1:1.25, and 300 ng of purified dsDNA in NEBuffer 1.1. Cleavage also occurred in the presence of 6 pmol mgCas12a-1 protein, FnCas12a crRNA #1 at a ratio of 1:1.25, and 300 ng of purified dsDNA in NEBuffer 1.1. Cleavage was also observed in the above conditions in NEBuffer 2.1 and NEBuffer 3.1, however, cleavage was strongest in the NEBuffer 1.1. For mgCas12a-2, cleavage was primarily seen in the presence of 3 pmol mgCas12a-2, FnCas12a crRNA #1 at a ratio of 1:1.25, and 300 ng of purified dsDNA in NEBuffer 1.1 and 6 pmol mgCas12a-1 protein, FnCas12a crRNA #1 at a ratio of 1:1.25, and 300 ng of purified dsDNA in NEBuffer 1.1. In both cases, cleaved products were detected at 1.6 kbp and 0.6 kbp.

[0155] FIG. 11 shows gel electrophoresis of target dsDNA after exposure to mgCas12a-1 (SEQ ID NO: 1) or mgCas12a-2 (SEQ ID NO: 3) complexed with crRNA in no buffer, NEBuffer 1.1, NEBuffer 2.1, and NEBuffer 3.1 for a 10 minute reaction run at 37°C. Cleaved products are indicated by an arrow. The target dsDNA (LsXTb12) resolves ~2.2 kbp. Cleavage was found to occur in the presence of 3 pmol mgCas12a-1 protein, FnCas12a crRNA #1 at a ratio of 1:1.25, and 300 ng of purified dsDNA in NEBuffer 1.1. For mgCas12a-2, cleavage was primarily seen in the presence of 3 pmol mgCas12a-2, FnCas12a crRNA #1 at a ratio of 1:1.25, and 300 ng of purified dsDNA in NEBuffer 1.1 and 6 pmol mgCas12a-1 protein, FnCas12a crRNA #1 at a ratio of 1:1.25, and 300 ng of purified dsDNA in NEBuffer 1.1. In both cases, cleaved products were detected at 1.8 kbp and 0.4 kbp. While **FIG. 11** and **FIG. 10** have similar reaction conditions, **FIG. 10** and **FIG. 11** use two different crRNAs (crRNA #1 for **FIG. 10** and crRNA #2 for **FIG. 11**). Thus, the resulting cleaved products are of different sizes in **FIG. 11** as compared to **FIG. 10**.

[0156] FIG. 12 shows gel electrophoresis of target dsDNA after exposure to Cas12a proteins including mgCas12a-1, mgCas12a-2, AsCas12a, FnCas12a, and LbCas12a complexed with crRNA in NEBuffer 1.1 for a 10 minute reaction run at 37°C. Cleaved products are indicated by an arrow. The target dsDNA (LsXTb12) resolves ~2.2 kbp. For mgCas12a-1, mgCas12a-2, AsCas12a, FnCas12a, and LbCas12a, cleavage occurred in the presence of 6 pmol of mgCas12a-1 protein, 7.5 pmol FnCas12a crRNA #1, and 300 ng of purified dsDNA. Cleaved products were detected at 1.6 kbp and 0.6 kbp. Additionally, during testing of mgCas12a-1, mgCas12a-2, AsCas12a, FnCas12a, or LbCas12a in the presence of 6 pmol of mgCas12a-1 protein, 7.5 pmol FnCas12a crRNA #2, and

300 ng of purified dsDNA, cleavage as best detected for mgCas12a-1. Cleaved products were detected at 1.8 kbp and 0.4 kbp.

[0157] **FIG. 13** shows the target double stranded DNA LsXTb12 tested in the gel electrophoresis experiments of **FIG. 10** to **FIG. 12** and the corresponding binding regions of crRNA #1 and crRNA #2, including in exon 1 and exon 2, respectively. **FIG. 13** shows a solid top bar indicated to be “part of LsXTb12” and is shown from its 5’ end to its 3’ end. Indicated below the target dsDNA LsXTb12 to the left is an arrow running from left to right showing the 5’ UTR and Exon 1. XYLT crRNA #1 (413 to 439) is shown above the solid top bar of the target dsDNA and is shown as binding to a region above Exon 1 in the target dsDNA. Indicated below the target dsDNA LsXTb12 to the right is two arrows running from left to right including Exon 2 followed by part of Exon 3. XYLT crRNA #2 (1593 to 1619) is shown above the solid top bar of the target dsDNA and is shown as binding to a region above Exon 2 in the target dsDNA.

[0158] **FIG. 14** illustrates gel electrophoresis of a target DNA after incubation with a Cas12a nuclease of the present disclosure. The left most lane shows a DNA ladder and the Mock A and Mock B lane shows controls. The Mock A lane shows a reaction lacking gRNA and protein. Mock B indicates a reaction lacking protein. As expected, the target DNA remains uncut in the Mock A and Mock B lanes and resolves at a size of 1430 base pairs (bp). Cas12a proteins tested for cleavage of the target DNA includes FnCas12, AsCas12, LbCas12, He-MgCas12a-1 (humanized and engineered mgCas12a-1), and He-MgCas12a-2 (humanized and engineered mgCas12a-2). He-MgCas12a-1 and He-MgCas12a-2 both cleaved the DNA template (1,430 bp) into two pieces of 750 bp and 680 bp. He-MgCas12a-1 cleaved all the DNA template within 1 hour of incubation as evidenced by the lack of a band at the intact DNA template size. He-MgCas12a-2 cleaved all the DNA template within 4 hours of incubation as evidenced by the lack of a band at the intact DNA template size. Although FnCas12, AsCas12, and LbCas12 cleaved the target DNA, uncleaved DNA template remained even after 4 hours as evidenced by a band at the intact DNA template size.

[0159] **FIG. 15A-B** shows two graphs illustrating genome editing efficiencies of FnCas12 versus He-MgCas12a-1 in rice (**FIG. 15A**) and in *N. benthamiana* (**FIG. 15B**). Genome editing efficiencies were measured by next generation sequencing techniques and the y-axis shows percent editing efficiency. Mock indicates a negative control sample. For genome editing in rice, two crRNAs were evaluated and for genome editing in *N. benthamiana*, three crRNAs were evaluated. He-MgCas12a-1 exhibited higher editing efficiency in rice with crRNA 1 and crRNA 2 as compared to FnCas12. He-MgCas12a-1 exhibited higher editing efficiency in *N. benthamiana* with crRNA 2 than FnCas12.

[0160] FIG. 16 illustrates a gel electrophoresis of 300 ng target DNA (LsXTb12) incubated with Cas12a(6 pmol)/crRNA (7.5 pmol) complexes for 1 hour at 37° in 1X NEBuffer. Each lane of the gel electrophoresis shows various conditions. The lanes in the gel from left to right show: 1) the DNA ladder, 2) the purified dsDNA and without Cas12a or crRNA, 3) mgCas12a-1, crRNA and purified dsDNA, 4) he_mgCas12a-1, crRNA and purified dsDNA,, 5) de_mgCas12a-1, crRNA and purified dsDNA, 6) mgCas12a-2, crRNA and purified dsDNA, 7) he_mgCas12a-2, crRNA and purified dsDNA,, 8) de_mgCas12a-2, crRNA, and purified dsDNA, 9) AsCas12a, crRNA and purified dsDNA, 10) FnCas12a, crRNA and purified dsDNA,, 11) LbCas12a, crRNA, and purified dsDNA. Cleaved products were observed in the mgCas12a lanes and the he-mgCas12a lanes as well as the AsCas12a, FnCas12a, and LbCas12a lanes. Cleaved template DNA was seen at 1.8 kB and 0.65 kB.

[0161] FIG. 17A shows a gel electrophoresis of 300 ng of target DNA (LsXTb12) incubated with Cas12a (6 pmol) and crRNA (7.5 pmol) for 12 hour at 37° in 1X NEBuffer. The lanes in the gel from left to right show: 1) purified dsDNA alone, 2) crRNA and purified dsDNA, 3) purified dsDNA only at 37°C, 4) crRNA and purified dsDNA at 37°C, 5) mgCas12a-1, crRNA and purified dsDNA, 6) he_mgCas12a-1, crRNA and purified dsDNA, 7) de_mgCas12a-1, crRNA and purified dsDNA, 8) mgCas12a-2, crRNA and purified dsDNA, 9) he_mgCas12a-2, crRNA and purified dsDNA,, 10) de_mgCas12a-2, crRNA, and purified dsDNA, 11) AsCas12a, crRNA and purified dsDNA, 12) FnCas12a, crRNA and purified dsDNA, and 13) LbCas12a, crRNA, and purified dsDNA. While AsCas12a, FnCas12a, and LbCas12a fully degraded template DNA, as evidenced by a lack in a band at the target DNA size, intact template DNA was observed in the case of the mgCas12a, he_mgCas12a, and de_mgCas12a proteins. Cleaved products of the template DNA were observed for mgCas12a and he_mgCas12a proteins, which resolved at 1.8 kb and 0.65 kb.

[0162] FIG. 17B shows a gel electrophoresis of 300 ng of target DNA (LsXTb12) incubated with Cas12a (6 pmol) and crRNA (7.5 pmol) for 24 hours at 37° in 1X NEBuffer. The lanes in the gel from left to right show: 1) the DNA ladder, 2) the purified dsDNA and without Cas12a or crRNA, 3) mgCas12a-1, crRNA and purified dsDNA, 4) he_mgCas12a-1, crRNA and purified dsDNA, 5) de_mgCas12a-1, crRNA and purified dsDNA, 6) mgCas12a-2, crRNA and purified dsDNA, 7) he_mgCas12a-2, crRNA and purified dsDNA, 8) de_mgCas12a-2, crRNA, and purified dsDNA, 9) AsCas12a, crRNA and purified dsDNA, 10) FnCas12a, crRNA and purified dsDNA,, 11) LbCas12a, crRNA, and purified dsDNA. Template DNA was observed in the mgCas12a lanes after 24 hours. In contrast, no target DNA was observed in the AsCas12a, FnCas12a, and LbCas12a lanes.

[0163] FIG. 18 shows a gel electrophoresis of 300 ng of target plasmid DNA (~10 kbp) incubated with Cas12a (6 pmol) and crRNA (7.5 pmol) for 1 hour at 37° in 1X NEBuffer. The lanes in the gel from left to right show: 1) the DNA ladder, 2) the purified dsDNA and without Cas12a or crRNA, 3) mgCas12a-1, crRNA and purified dsDNA, 4) he_mgCas12a-1, crRNA and purified dsDNA,, 5) de_mgCas12a-1, crRNA and purified dsDNA, 6) mgCas12a-2, crRNA and purified dsDNA, 7) he_mgCas12a-2, crRNA and purified dsDNA,, 8) de_mgCas12a-2, crRNA, and purified dsDNA, 9) AsCas12a, crRNA and purified dsDNA, 10) FnCas12a, crRNA and purified dsDNA,, 11) LbCas12a, crRNA, and purified dsDNA. Target plasmid DNA was evidenced by a band at 10 kbp and cleaved products were observed at 6 kbp. All Cas12a proteins, with the exception of the de_mgCas12a proteins cleaved the target plasmid DNA to a linearized piece of DNA at 6kbp in size.

[0164] FIG. 19A shows gel electrophoresis of 300 ng of target plasmid DNA incubated with Cas12a nucleases including mgCas12a-1 (wild-type), he_mgCas12a-1 (humanized and engineered mgCas12a-1), de_mgCas12a-1 (dead and engineered mgCas12a), mg Cas12a-2, he_mgCas12a-2, de_mgCas12a-2, AsCas12a, FnCas12a, and LbCas12a after a reaction time of 12 h in 1X NEBuffer at 37°C. The lanes in the gel from left to right show: 1) purified dsDNA alone, 2) crRNA and purified dsDNA, 3) purified dsDNA only at 37°C, 4) crRNA and purified dsDNA at 37°C, 5) mgCas12a-1, crRNA and purified dsDNA, 6) he_mgCas12a-1, crRNA and purified dsDNA,, 7) de_mgCas12a-1, crRNA and purified dsDNA, 8) mgCas12a-2, crRNA and purified dsDNA, 9) he_mgCas12a-2, crRNA and purified dsDNA,, 10) de_mgCas12a-2, crRNA, and purified dsDNA, 11) AsCas12a, crRNA and purified dsDNA, 12) FnCas12a, crRNA and purified dsDNA,, and 13) LbCas12a, crRNA, and purified dsDNA. AsCa12a, FnCas12a, and LbCas12a degraded all template plasmid DNA within 12 hours as evidenced by the lack of any bands. mgCas12a proteins and he_mgCas12a proteins both exhibited linearized cleaved products and some target plasmid DNA as evidenced by bands at their respective expected sizes.

[0165] FIG. 19B shows gel electrophoresis of 300 ng of target plasmid DNA incubated with Cas12a nucleases including mgCas12a-1 (wild-type), he_mgCas12a-1 (humanized and engineered mgCas12a-1), de_mgCas12a-1 (dead and engineered mgCas12a), mg Cas12a-2, he_mgCas12a-2, de_mgCas12a-2, AsCas12a, FnCas12a, and LbCas12a after a reaction time of 24 h in 1X NEBuffer at 37°C. The lanes in the gel from left to right show: 1) the DNA ladder, 2) the purified dsDNA and without Cas12a or crRNA, 3) mgCas12a-1, crRNA and purified dsDNA, 4) he_mgCas12a-1, crRNA and purified dsDNA,, 5) de_mgCas12a-1, crRNA and purified dsDNA, 6) mgCas12a-2, crRNA and purified dsDNA, 7) he_mgCas12a-2, crRNA and purified dsDNA,, 8) de_mgCas12a-2, crRNA, and purified dsDNA, 9) AsCas12a, crRNA and purified dsDNA, 10) FnCas12a, crRNA

and purified dsDNA,, 11) LbCas12a, crRNA, and purified dsDNA. AsCas12a, FnCas12a, and LbCas12a degraded all template plasmid DNA within 24 hours as evidenced by the lack of any bands. mgCas12a proteins and he_mgCas12a proteins both exhibited linearized cleaved products and as evidenced by bands at the expected size.

[0166] FIG. 20A shows a flowchart for mining novel Cas12a proteins from the metagenome database. Metagenome sequence data can be obtained from Genbank and CRISPR sequences can be found using metaCRT and gene prediction (Prodigal). Cas12a candidate proteins are identified using local BLASTp of Cas12a sequences from Uniprot and Genbank. Such identified Cas12a candidate proteins are then analyzed to generate phylogenetic tree using MEGAX, which information is annotated using web BALSTp.

[0167] FIG. 20B shows a phylogenetic tree of various Cas12a orthologs. The metagenomically mined Cas12a proteins of the present disclosure, including mgCas12a-2 and mgCas12a-1 resolve at the right. The tree additionally shows the relationship between other Cas12a proteins using bootstrap values.

[0168] FIG. 20C shows a schematics of the various domains in AsCas12a, mgCas12a-1, and mgCas12a-2, which includes WED-I, REC1, REC2, WED-II, PI, WED-II, RuvC-I, BH, RuvC-II, Nuc, RuvC-III. Note that AsCas12a includes Cas4, Cas1, Cas2, CRISPR domains while Cas1 is absent in mgCas12a-1, and Cas4, Cas1, Cas2 are absent in mgCas12a-2. Further noted is that mgCas12a-1 comprises mutations of D877A and K925Q, and mgCas12a-2 comprises mutations of D873A and K925Q, in RuvC-I and RuvC-II domains, respectively.

[0169] FIG. 21 shows an unrooted and evolutionary distance-based phylogenetic tree of metagenome-derived Cas12a, including mgCas12a-2 and mgCas12a-1, which resolve at right. Other Cas12 orthologs are also indicated, which clockwise from the bottom left are Lb2Cas12a, LbCas12a, PcCas12a, PdCas12a, MbCas12a, SsCas12a, LiCas12a, PmCas12a, FnCas12a, PbCas12a, PeCas12a, Lb3Cas12a, BpCas12a, CEST01022924.1 4, CESD01057036.1 3, CMtCas12a, AsCas12a, mgCas12a-2, mgCas12a-1, CDYR01026036.1 2, CDYL01005663.1 6, EeCas12a, CDYK01004246.1 121, CDYK01036676.1 3, CDYL01025564.1 3, CEAM01003869.1 48, and CDZY01023362.1 31.

[0170] FIG. 22A shows at the top left a table. The first row, from left to right, shows no incubation and 2 hour incubation at 37°C. The second row, from left to right, shows 4 blank cells followed by Fn, WT mg-1, and WT mg-2. The third row indicates the presence or absence of Cas12a in the reaction. The fourth row indicates the presence or absence of crRNA in the reaction. The bottom row indicates the presence or absence of linear dsDNA. The gel immediately below shows each reaction in the table, where the linear dsDNA targets HsCCR5 and is indicated as “S” for substrate.

Bands appearing by the arrow indicated as S shows uncleaved linear dsDNA. Cleaved products are also indicated by arrows and bands appearing by the arrows indicating the position of cleaved products show cleaved pieces of the target linear dsDNA. The gel immediately below shows each reaction in the table, where the linear dsDNA targets HsDNMT1 and is indicated as “S” for substrate. Bands appearing by the arrow indicated as S shows uncleaved linear dsDNA. Cleaved products are also indicated by arrows and bands appearing by the arrows indicating the position of cleaved products show cleaved pieces of the target linear dsDNA. The gel immediately below shows each reaction in the table, where the linear dsDNA targets HsEMX1 and is indicated as “S” for substrate. Bands appearing by the arrow indicated as S shows uncleaved linear dsDNA. Cleaved products are also indicated by arrows and bands appearing by the arrows indicating the position of cleaved products show cleaved pieces of the target linear dsDNA.

[0171] FIG. 22B shows at the top left a table. The first row, from left to right, shows no incubation and 2 hour incubation at 37°C. The second row, from left to right, shows 4 blank cells followed by Fn, WT mg-1, and WT mg-2. The third row indicates the presence or absence of Cas12a in the reaction. The fourth row indicates the presence or absence of crRNA in the reaction. The bottom row indicates the presence or absence of circular dsDNA. The gel immediately below shows each reaction in the table, where the circular dsDNA targets HsCCR5 and is indicated as “S” for substrate. Bands appearing by the arrow indicated as S shows uncleaved linear dsDNA. Linearized products are also indicated by arrows and bands appearing by the arrows indicating the position of linearized products show cleaved pieces of the target circular dsDNA. The gel immediately below shows each reaction in the table, where the circular dsDNA targets HsDNMT1 and is indicated as “S” for substrate. Bands appearing by the arrow indicated as S shows uncleaved circular dsDNA. Cleaved products are also indicated by arrows and bands appearing by the arrows indicating the position of cleaved products show cleaved pieces of the target circular dsDNA. The gel immediately below shows each reaction in the table, where the circular dsDNA targets HsEMX1 and is indicated as “S” for substrate. Bands appearing by the arrow indicated as S shows uncleaved circular dsDNA. Cleaved products are also indicated by arrows and bands appearing by the arrows indicating the position of cleaved products show cleaved pieces of the circular linear dsDNA.

[0172] FIG. 23A shows a table of reaction conditions. The first row shows, from left to right, 0, 1 m, 10 m, 30 m, 1 h, 6 h, and 12 h. The second row indicates the presence or absence of d_mgCas12a, The third row indicates the presence or absence of WT mgCas12. The fourth row indicates the presence or absence of crRNA. The fifth row indicates the presence or absence of linear dsDNA. The gel immediately below shows each reaction in the table, where the mgCas12a-1 is complexed with a crRNA having a 5' handle from AsCas12a (the crRNA, from 5' to 3' is

UAAUUUCUACUCUUGUAGAU). Linear dsDNA is indicated by an “S” for substrate and bands appearing at this position show uncleaved target linear dsDNA. Cleaved products are indicated by “P” for product and bands appearing at this position show cleaved segments of the target linear dsDNA. The gel immediately below shows each reaction in the table, where the mgCas12a-1 is complexed with a crRNA having a 5’ handle from FnCas12a (the crRNA, from 5’ to 3’ is AAUUUCUACUGUUGUAGAU). Linear dsDNA is indicated by an “S” for substrate and bands appearing at this position show uncleaved target linear dsDNA. Cleaved products are indicated by “P” for product and bands appearing at this position show cleaved segments of the target linear dsDNA. The gel immediately below shows each reaction in the table, where the mgCas12a-1 is complexed with a crRNA having a 5’ handle from LbCas12a (the crRNA, from 5’ to 3’ is AAUUUCUACUAAGUGUAGAU). Linear dsDNA is indicated by an “S” for substrate and bands appearing at this position show uncleaved target linear dsDNA. Cleaved products are indicated by “P” for product and bands appearing at this position show cleaved segments of the target linear dsDNA.

[0173] FIG. 23B shows a table of reaction conditions. The first row shows, from left to right, 0, 1 m, 10 m, 30 m, 1 h, 6 h, and 12 h. The second row indicates the presence or absence of d_mgCas12a, The third row indicates the presence or absence of WT mgCas12. The fourth row indicates the presence or absence of crRNA. The fifth row indicates the presence or absence of linear dsDNA. The gel immediately below shows each reaction in the table, where the mgCas12a-2 is complexed with a crRNA having a 5’ handle from AsCas12a (the crRNA, from 5’ to 3’ is UAAUUUCUACUCUUGUAGAU). Linear dsDNA is indicated by an “S” for substrate and bands appearing at this position show uncleaved target linear dsDNA. Cleaved products are indicated by “P” for product and bands appearing at this position show cleaved segments of the target linear dsDNA. The gel immediately below shows each reaction in the table, where the mgCas12a-2 is complexed with a crRNA having a 5’ handle from FnCas12a (the crRNA, from 5’ to 3’ is AAUUUCUACUGUUGUAGAU). Linear dsDNA is indicated by an “S” for substrate and bands appearing at this position show uncleaved target linear dsDNA. Cleaved products are indicated by “P” for product and bands appearing at this position show cleaved segments of the target linear dsDNA. The gel immediately below shows each reaction in the table, where the mgCas12a-2 is complexed with a crRNA having a 5’ handle from LbCas12a (the crRNA, from 5’ to 3’ is AAUUUCUACUAAGUGUAGAU). Linear dsDNA is indicated by an “S” for substrate and bands appearing at this position show uncleaved target linear dsDNA. Cleaved products are indicated by “P” for product and bands appearing at this position show cleaved segments of the target linear dsDNA.

[0174] **FIG. 24** shows a table at top, which from left to right summarizes reaction conditions. The top row from left to right indicates time points of 0 for two columns, 12 hour incubation at 37°C for nine columns, and 24 hour incubation at 37°C for nine columns. The second row indicates the Cas12a in each reaction and shows, from left to right, 4 blank cells, followed by WT mg-1, d_mg-1, WT mg-2, d_mg-2, As, Fn, Lb, 2 blank cells, followed by WT mg-1, d_mg-1, WT mg-2, d_mg-2, As, Fn, and Lb. The third row indicates the presence or absence of Cas12a. The fourth row indicates the presence or absence of crRNA. The fifth row indicates the presence or absence of linear dsDNA. The gel immediately below shows each reaction in the table, where each tested Cas12a is complexed with a crRNA targeted to HsCCR5 linear dsDNA. Linear dsDNA is indicated by an “S” for substrate and bands appearing at this position show uncleaved target linear dsDNA. Cleaved products are indicated by “P” for product and bands appearing at this position show cleaved segments of the target linear dsDNA. The gel immediately below shows each reaction in the table, where each tested Cas12a is complexed with a crRNA targeted to HsDNMT1 linear dsDNA. Linear dsDNA is indicated by an “S” for substrate and bands appearing at this position show uncleaved target linear dsDNA. Cleaved products are indicated by “P” for product and bands appearing at this position show cleaved segments of the target linear dsDNA. The gel immediately below shows each reaction in the table, where each tested Cas12a is complexed with a crRNA targeted to HsEMX1 linear dsDNA. Linear dsDNA is indicated by an “S” for substrate and bands appearing at this position show uncleaved target linear dsDNA. Cleaved products are indicated by “P” for product and bands appearing at this position show cleaved segments of the target linear dsDNA.

[0175] **FIG. 25A** shows a table at top summarizing reaction conditions for each reaction. The first row shows, from left to right, 0, 1m, 10m, 30m, 1h, 2h, 6h, 12h, 24h, and 48h for the last two columns. The second row indicates the presence or absence of Cas12a in each reaction. The third row indicates the presence or absence of linear dsDNA in each reaction. The gel immediately below shows each reaction in the table, where FnCas12a is complexed with a crRNA targeted to linear dsDNA. Linear dsDNA is indicated by an “S” for substrate and bands appearing at this position show uncleaved target linear dsDNA. The gel immediately below shows each reaction in the table, where WT mgCas12a-1 is complexed with a crRNA targeted to linear dsDNA. Linear dsDNA is indicated by an “S” for substrate and bands appearing at this position show uncleaved target linear dsDNA. The gel immediately below shows each reaction in the table, where WT mgCas12a-2 is complexed with a crRNA targeted to linear dsDNA. Linear dsDNA is indicated by an “S” for substrate and bands appearing at this position show uncleaved target linear dsDNA.

[0176] **FIG. 25B** shows a graph of incubation period versus cleaved target dsDNA. The x-axis shows incubation period, which from left to right is, dsDNA only, 1m, 10m, 30m, 1h, 2h, 6h, 12h, 24h, and 48h. The y-axis shows cleaved target dsDNA which ranges from 0.00 to 1.00 in increments of 0.20. The solid line shows data for FnCas12a. The dashed line shows data for WT mgCas12a-1. The dotted line shows data for WT mgCas12a-2.

[0177] **FIG. 26A** shows a table at top summarizing reaction conditions for each reaction. The first two rows show the divalent cations which, from left to right, DW, Control, CaCl₂ at 10 and 100, CoCl₂ at 10 and 100, CuSO₄ at 10 and 100, FeCl₂ at 10 and 100, MnSO₄ at 10 and 100, NiSO₄ at 10 and 100, and ZnSO₄ at 10 and 100. The third row shows the presence or absence of Cas12a in each reaction. The fourth row shows the presence or absence of crRNA in each reaction. The fifth row shows the presence or absence of linear dsDNA in each reaction. The gel immediately below shows each reaction in the table, where FnCas12a is complexed with a crRNA targeted to linear dsDNA. Linear dsDNA is indicated by an “S” for substrate and bands appearing at this position show uncleaved target linear dsDNA. Cleaved products are indicated by “P” for product and bands appearing at this position show cleaved segments of the target linear dsDNA. The gel immediately below shows each reaction in the table, where WT mgCas12a-1 is complexed with a crRNA targeted to linear dsDNA. Linear dsDNA is indicated by an “S” for substrate and bands appearing at this position show uncleaved target linear dsDNA. Cleaved products are indicated by “P” for product and bands appearing at this position show cleaved segments of the target linear dsDNA. The gel immediately below shows each reaction in the table, where WT mgCas12a-2 is complexed with a crRNA targeted to linear dsDNA. Linear dsDNA is indicated by an “S” for substrate and bands appearing at this position show uncleaved target linear dsDNA. Cleaved products are indicated by “P” for product and bands appearing at this position show cleaved segments of the target linear dsDNA.

[0178] **FIG. 26B** shows a graph of cleaved target dsDNA versus various reaction conditions with divalent cations for FnCas12a complexed with crRNA targeting linear dsDNA. The x-axis shows reaction conditions, which from left to right are, Ctrl, 10 mM CaCl₂, 100 mM CaCl₂, 10 mM CoCl₂, 100 mM CoCl₂, 10 mM CuSO₄, 100 mM CuSO₄, 10 mM FeCl₂, 100 mM FeCl₂, 10 mM MnSO₄, 100 mM MnSO₄, 10 mM NiSO₄, 100 mM NiSO₄, 10 mM ZnSO₄, and 100 mM ZnSO₄. The y-axis shows cleaved target dsDNA ranging from -0.20 to 1.20 in increments of 0.20.

[0179] **FIG. 26C** shows a graph of cleaved target dsDNA versus various reaction conditions with divalent cations for WT mgCas12a-1 complexed with crRNA targeting linear dsDNA. The x-axis shows reaction conditions, which from left to right are, Ctrl, 10 mM CaCl₂, 100 mM CaCl₂, 10 mM CoCl₂, 100 mM CoCl₂, 10 mM CuSO₄, 100 mM CuSO₄, 10 mM FeCl₂, 100 mM FeCl₂, 10 mM

MnSO₄, 100 mM MnSO₄, 10 mM NiSO₄, 100 mM NiSO₄, 10 mM ZnSO₄, and 100 mM ZnSO₄.

The y-axis shows cleaved target dsDNA ranging from -0.20 to 0.70 in increments of 0.10.

[0180] FIG. 26D shows a graph of cleaved target dsDNA versus various reaction conditions with divalent cations for WT mgCas12a-2 complexed with crRNA targeting linear dsDNA. The x-axis shows reaction conditions, which from left to right are, Ctrl, 10 mM CaCl₂, 100 mM CaCl₂, 10 mM CoCl₂, 100 mM CoCl₂, 10 mM CuSO₄, 100 mM CuSO₄, 10 mM FeCl₂, 100 mM FeCl₂, 10 mM MnSO₄, 100 mM MnSO₄, 10 mM NiSO₄, 100 mM NiSO₄, 10 mM ZnSO₄, and 100 mM ZnSO₄. The y-axis shows cleaved target dsDNA ranging from -1.20 to 1.00 in increments of 0.20.

[0181] FIG. 27 shows a graph of reaction conditions versus indels (%). The x-axis shows, from left to right, Mock at 48 h targeting CCR5, mgCas12a-1 at 48 h targeting CCR5, mgCas12a-2 at 48 h targeting CCR5, Mock at 72 h targeting CCR5, mgCas12a-1 at 72 h targeting CCR5, mgCas12a-2 at 72 h targeting CCR5, Mock at 48 h targeting DNMT1, mgCas12a-1 at 48 h targeting DNMT1, mgCas12a-2 at 48 h targeting DNMT1, Mock at 72 h targeting DNMT1, mgCas12a-1 at 72 h targeting DNMT1, and mgCas12a-2 at 72 h targeting DNMT1. The y-axis shows indels (%) ranging from 0.0 to 0.6 in increments of 0.1.

[0182] FIG. 28 shows a graph of Cas12a tested versus indel frequency (%). The x-axis shows the Cas12a and crRNA tested, which from left to right is crRNA2, mgCas12a-1-crRNA2, mgCas12a-2-crRNA2, FnCpf1-crRNA2, crRNA4, mgCas12a-1-crRNA4, mgCas12a-2-crRNA4, and FnCpf1-crRNA4. The y-axis shows indel frequency (%) ranging from 0.0 to 2.0 in increments of 0.2.

[0183] A “functional domain” of a protein RNA guided endonuclease of the present disclosure may be a putative transposase DNA binding domain, such as residue 829 through residue 991 of SEQ ID NO: 1, or residue 825 through residue 996 of SEQ ID NO: 3.

[0184] “Sequence identity” as used herein may describe the number, the fraction, or the percentage of nucleobases or amino acid residues that share identity, or are common, between two sequences being compared. Percent (%) sequence identity with respect to a reference polypeptide sequence is the percentage of amino acid residues in a first sequence (*e.g.*, a candidate sequence) that are identical with the amino acid residues in a second sequence (*e.g.*, the reference polypeptide sequence), after aligning the sequences, optionally introducing gaps to achieve the maximum percent sequence identity, and, optionally, not accounting for conservative substitutions as part of the sequence identity. Similarly, percent (%) sequence identity with respect to a reference nucleobase sequence is the percentage of nucleobases in a first sequence (*e.g.*, a candidate sequence) that are identical with the nucleobases in a second sequence (*e.g.*, the nucleobase reference sequence), after aligning the sequences, optionally introducing gaps to achieve the maximum percent sequence identity, and, optionally, not accounting for conservative substitutions

as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are known for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN or Megalign (DNASTAR) software. Appropriate parameters for aligning sequences are able to be determined, including algorithms needed to achieve maximal alignment over the full length of the sequences being compared. For purposes herein, however, percent (%) amino acid sequence identity values are generated using the sequence comparison computer program ALIGN-2. The ALIGN-2 sequence comparison computer program was authored by Genentech, Inc., and the source code has been filed with user documentation in the U.S. Copyright Office, Washington D.C., 20559, where it is registered under U.S. Copyright Registration No. TXU510087. The ALIGN-2 program is publicly available from Genentech, Inc., South San Francisco, Calif., or may be compiled from the source code. The ALIGN-2 program should be compiled for use on a UNIX operating system, including digital UNIX V4.0D. All sequence comparison parameters are set by the ALIGN-2 program and do not vary.

[0185] “Consecutive residues” as used herein may describe nucleobases or amino acid residues that are immediately adjacent to each other in a given sequence.

[0186] “Complementarity” refers to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick or other non-traditional types. A percent complementarity indicates the percentage of residues in a nucleic acid molecule which can form hydrogen bonds (*e.g.*, Watson-Crick base pairing) with a second nucleic acid sequence (*e.g.*, 5, 6, 7, 8, 9, 10 out of 10 being 50%, 60%, 70%, 80%, 90%, and 100% complementary, respectively). “Perfectly complementary” means that all the contiguous residues of a nucleic acid sequence will hydrogen bond with the same number of contiguous residues in a second nucleic acid sequence. “Substantially complementary” as used herein refers to a degree of complementarity that is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 92%, 95%, 97%, 98%, 99%, or 100% over a region of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, or more nucleotides, or refers to two nucleic acids that hybridize under stringent conditions. Sequence identity, such as for the purpose of assessing percent complementarity, may be measured by any suitable alignment algorithm, including but not limited to the Needleman-Wunsch algorithm (see *e.g.* the EMBOSS Needle aligner available at www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html, optionally with default settings), the BLAST algorithm (see *e.g.* the BLAST alignment tool available at blast.ncbi.nlm.nih.gov/Blast.cgi, optionally with default settings), or the Smith-Waterman algorithm (see *e.g.* the EMBOSS Water aligner available at www.ebi.ac.uk/Tools/psa/emboss_water/nucleotide.html, optionally with default settings). Optimal

alignment may be assessed using any suitable parameters of a chosen algorithm, including default parameters.

[0187] Nucleic acid and amino acid sequence homology is determined according to any suitable method known in the art, including but not limited to those described herein.

[0188] For example, alignments and searches for similar sequences can be performed using the U.S. National Center for Biotechnology Information (NCBI, Bethesda, MD) program, MegaBLAST. Use of this program with options for percent identity set at, for example, 70% for amino acid sequences, or set at, for example, 90% for nucleotide sequences, will identify those sequences with 70%, or 90%, or greater sequence identity to the query sequence. Other software known in the art is also available for aligning and/or searching for similar sequences, *e.g.*, sequences at least 70% or 90% identical to an information string containing a secretion signal sequence herein. For example, sequence alignments for comparison to identify sequences at least 70% or 90% identical to a query sequence is often performed by use of, *e.g.*, the GAP, BESTFIT, BLAST, FASTA, and TFASTA programs available in the GCG Sequence Analysis Software Package (available from the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705), with the default parameters as specified therein, plus a parameter for the extent of sequence identity set at the desired percentage. Also, for example, the CLUSTAL program (available in the PC/Gene software package from Intelligenetics, Mountain View, Cal.) may be used.

[0189] These and other sequence alignment methods may be conducted by manual alignment, by visual inspection, or by manual or automatic application of a sequence alignment algorithm, such as any of those embodied by the above-described programs. Various useful algorithms include, *e.g.*: the similarity search method described in W. R. Pearson & D. J. Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444-48 (April 1988); the local homology method described in T. F. Smith & M. S. Waterman, in *Adv. Appl. Math.* 2:482-89 (1981) and in *J. Molec. Biol.* 147:195-97 (1981); the homology alignment method described in S. B. Needleman & C. D. Wunsch, *J. Molec. Biol.* 48(3):443-53 (March 1970); and the various methods described, *e.g.*, by W. R. Pearson, in *Genomics* 11(3):635-50 (November 1991); by W. R. Pearson, in *Methods Molec. Biol.* 24:307-31 and 25:365-89 (1994); and by D. G. Higgins & P. M. Sharp, in *Comp. Appl'ns in Biosci.* 5:151-53 (1989) and in *Gene* 73(1):237-44 (15 Dec. 1988).

[0190] GAP Version 10, which uses the algorithm of Needleman and Wunsch (1970) *supra*, can be used to determine sequence identity or similarity using the following parameters: percent (%) identity and percent (%) similarity for a nucleotide sequence using GAP Weight of 50 and Length Weight of 3, and the nwsgapdna.cmp scoring matrix; % identity or % similarity for an amino acid

sequence using GAP weight of 8 and length weight of 2, and the BLOSUM62 scoring program. Equivalent or similar programs may also be used as will be understood by one of skill in the art. For example, a sequence comparison program can be used that, for any two sequences in question, generates an alignment having identical nucleotide residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by GAP Version 10. In some embodiments, the sequence comparison is performed across the entirety of the query or the subject sequence, or both.

Numbered Embodiments

[0191] The following embodiments recite non-limiting permutations of combinations of features disclosed herein. Other permutations of combinations of features are also contemplated. In particular, each of these numbered embodiments is contemplated as depending from or relating to every previous or subsequent numbered embodiment, independent of their order as listed.

1. A composition comprising: a protein having at least 80% sequence identity with residue 829 through residue 991 of SEQ ID NO: 1, wherein the protein comprises a Lysine at position 925; and a guide RNA comprising a sequence that is reverse complementary to a eukaryotic nucleic acid sequence comprising from 6 to 60 bases.
2. A composition comprising: a protein having at least 80% sequence identity with residue 825 through residue 996 of SEQ ID NO: 3, wherein the protein comprises a Lysine at position 930; and a guide RNA comprising a sequence that is reverse complementary to a eukaryotic nucleic acid sequence comprising from 6 to 60 bases.
3. A composition comprising: a protein having at least 80% sequence identity with SEQ ID NO: 1, wherein the protein comprises a Lysine at position 925; and a guide RNA comprising a sequence that is reverse complementary to a eukaryotic nucleic acid sequence comprising from 6 to 60 bases.
4. A composition comprising: a protein having at least 80% sequence identity with SEQ ID NO: 3, wherein the protein comprises a Lysine at position 930; and a guide RNA comprising a sequence that is reverse complementary to a eukaryotic nucleic acid sequence comprising from 6 to 60 bases.
5. The composition of any one of embodiments 1-4, wherein the eukaryotic nucleic acid sequence is a human nucleic acid sequence.
6. The composition of any one of embodiments 1-4, wherein the eukaryotic nucleic acid sequence is a plant nucleic acid sequence.
7. The composition of any one of embodiments 1 or 3, wherein a nucleotide sequence encoding for SEQ ID NO: 1 comprises at least 80% sequence identity to SEQ ID NO: 2.
8. The composition of any one of embodiments 2 or 4, wherein a nucleotide sequence encoding for SEQ ID NO: 3 comprises at least 80% sequence identity to SEQ ID NO: 4.
9. The composition of any one of embodiments 1-8, wherein the protein is from the Eubacteriaceae family.
10. The composition of any one of embodiments 1-9, wherein the protein comprises a nuclease.
11. The composition of embodiment 10, wherein the nuclease

comprises a type V CRISPR-associated protein. 12. The composition of embodiment 11, the type V CRISPR-associated protein comprises a Cas12a protein. 13. The composition of embodiment 12, wherein the Cas12a protein is metagenomically mined. 14. The composition of any one of embodiment 5-13, wherein the human nucleic acid sequence is implicated in cancer. 15. The composition of any one of embodiments 1-14, wherein the composition comprises a pH of from 7 to 7.9. 16. The composition of any one of embodiments 1-15, wherein the composition comprises a pH of 7. 17. The composition of any one of embodiments 1-16, wherein the composition is formulated in a buffer. 18. The composition of embodiment 17, wherein the buffer comprises Bis-Tris Propane-HCl. 19. The composition of any one of embodiments 17-18, wherein the buffer comprises MgCl₂. 20. The composition of any one of embodiments 17-19, wherein the buffer comprises bovine serum albumin. 21. The composition of any one of embodiments 17-20, wherein the buffer comprises from 0.1 to 50 mM Bis-Tris Propane-HCl. 22. The composition of any one of embodiments 17-21, wherein the buffer comprises from 0.1 to 50 mM MgCl₂. 23. The composition of any one of embodiments 17-22, wherein the buffer comprises from 1 to 500 µg/ml bovine serum albumin. 24. The composition of any one of embodiments 17-23, wherein the buffer comprises 10 mM Bis-Tris Propane-HCl. 25. The composition of any one of embodiments 17-24, wherein the buffer comprises 10 mM MgCl₂. 26. The composition of any one of embodiments 17-25, wherein the buffer comprises 100 µg/ml of bovine serum albumin. 27. The composition of any one of embodiments 1-26, wherein the protein comprises a purification tag. 28. The composition of embodiment 27, wherein the purification tag comprises at least one tag selected from the group consisting of a His tag, a FLAG tag, an AU1 epitope tag, an AU5 epitope tag, a bacteriophage T7 tag, a bacteriophage V5 epitope tag, a Bluetongue virus tag (B-tag), a Glu-Glu tag (EE-tag), an HSV epitope tag, a KT3 epitope tag, a Myc epitope tag, a PDZ ligand tag, a polyarginine tag, a polyaspartate tag, a polycysteine tag, a polyphenylalanine tag, a protein C tag, an S1-tag, an S-tag, a Step-tag, and a VSV-G tag. 29. The composition of any one of embodiments 1-28, wherein the guide RNA comprises an A-rich protospacer adjacent motif (PAM) sequence, a G-rich PAM sequence, a T-rich PAM sequence, or a C-rich PAM sequence. 30. The composition of any one of embodiments 1-29, wherein the guide RNA comprises a T-rich PAM sequence. 31. The composition of any one of embodiments 29-30, wherein the PAM sequence comprises 3 T nucleobases, 2 T nucleobases, or 1 T nucleobase. 32. The composition of any one of embodiments 29-31, wherein the PAM sequence comprises TTTN. 33. The composition of any one of embodiments 5-32, wherein the human nucleic acid sequence comprises a region in KRAS, HER2/neu, PD-1, TCR, p53, CCR5, DNMT1, EMX1, and LKB1. 34. The composition of any one of embodiments 14-33, wherein the cancer comprises a bladder cancer, a bone cancer, a blood

cancer, a breast cancer, a black colored tumor, a thyroid cancer, a parathyroid cancer, a bone marrow cancer, a laryngopharyngeal cancer, a laryngeal cancer, a lung cancer, an esophagus cancer, a pancreatic cancer, a colorectal cancer, a gastric cancer, a tongue cancer, a skin cancer, a brain tumor, a uterine cancer, a head or neck cancer, a gallbladder cancer, an oral cancer, a central nervous system tumor, or a liver cancer. 35. The composition of any one of embodiments 1-34, wherein the guide RNA sequence comprises from 1 to 100 nucleotides. 36. The composition of any one of embodiments 1-35, wherein the composition exhibits at least 2-fold than AsCas12a, FnCas12a, or LbCas12a. 37. The composition of any one of embodiments 1-36, wherein the guide RNA comprises a crRNA and a tracrRNA. 38. The composition of embodiment 37, wherein the crRNA comprises a 5' repeat recognition sequence of AAUU. 39. The composition of any one of embodiments 1-38, wherein protein exhibits cleavage activity in the presence of CaCl₂, CoCl₂, FeCl₂, MnSO₄, or any combination thereof. 40. A method of gene editing, wherein the method comprises contacting a cell with the composition of any one of embodiments 1-39; binding the guide RNA to the eukaryotic nucleic acid sequence; and cleaving the eukaryotic nucleic acid sequence. 41. A method of gene editing, wherein the method comprises providing a composition comprising a protein having at least 80% sequence identity with residue 829 through residue 991 of SEQ ID NO: 1, wherein the protein comprises a Lysine at position 925; and a guide RNA comprising a sequence that is reverse complementary to a nucleic acid sequence comprising from 6 to 60 bases; contacting a cell with the composition; binding the guide RNA to a eukaryotic nucleic acid sequence; and cleaving the nucleic acid sequence. 42. A method of gene editing, wherein the method comprises providing a composition comprising a protein having at least 80% sequence identity with residue 825 through residue 996 of SEQ ID NO: 3, wherein the protein comprises a Lysine at position 930; and a guide RNA comprising a sequence that is reverse complementary to a nucleic acid sequence comprising from 6 to 60 bases; contacting a cell with the composition; binding the guide RNA to a eukaryotic nucleic acid sequence; and cleaving the nucleic acid sequence. 43. A method of gene editing, wherein the method comprises providing a composition comprising a protein having at least 80% sequence identity with SEQ ID NO: 1, wherein the protein comprises a Lysine at position 925; and a guide RNA comprising a sequence that is reverse complementary to a nucleic acid sequence comprising from 6 to 60 bases; contacting a cell with the composition; binding the guide RNA to a eukaryotic nucleic acid sequence; and cleaving the nucleic acid sequence. 44. A method of gene editing, wherein the method comprises providing a composition comprising a protein having at least 80% sequence identity with SEQ ID NO: 3, wherein the protein comprises a Lysine at position 930; and a guide RNA comprising a sequence that is reverse complementary to a nucleic acid sequence comprising from 6 to 60 bases; contacting

a cell with the composition; binding the guide RNA to a eukaryotic nucleic acid sequence; and cleaving the nucleic acid sequence. 45. The method of any one of embodiments 41-44, wherein the eukaryotic nucleic acid sequence is a human nucleic acid sequence. 46. The method of any one of embodiments 41-44, wherein the eukaryotic nucleic acid sequence is a plant nucleic acid sequence. 47. The method of any one of embodiments 41 or 43, wherein a nucleotide sequence encoding for SEQ ID NO: 1 comprises at least 80% sequence identity to SEQ ID NO: 2. 48. The composition of any one of embodiments 42 or 44, wherein a nucleotide sequence encoding for SEQ ID NO: 3 comprises at least 80% sequence identity to SEQ ID NO: 4. 49. The method of any one of embodiments 36-48, wherein the cell comprises a cancer cell. 50. The method of any one of embodiments 36-49, wherein the contacting the cell with the composition comprises administering the composition to a subject in need thereof. 51. The method of embodiment 50, wherein the administering comprises intravenous, subcutaneous, intramuscular, oral, or mucosal administration. 52. The method of any one of embodiments 36-51, wherein the contacting the cell with the composition comprises administering the composition to the cell *ex vivo*. 53. The method of any one of embodiments 41-52, wherein the protein is from the Eubacteriaceae family. 54. The method of any one of embodiments 41-53, wherein the protein comprises a nuclease. 55. The method of embodiment 54, wherein the nuclease comprises a type V CRISPR-associated protein. 56. The method of embodiment 55, the type V CRISPR-associated protein comprises a Cas12a protein. 57. The method of embodiment 56, wherein the Cas12a protein is metagenomically mined. 58. The method of any one of embodiment 45-57, wherein the human nucleic acid sequence is implicated in cancer. 59. The method of any one of embodiments 41-58, wherein the method comprises cleaving the nucleic acid sequence at a pH of from 7 to 7.9. 60. The method of any one of embodiments 41-59, wherein the method comprises cleaving the nucleic acid sequence at a pH of 7. 61. The method of any one of embodiments 41-60, wherein the composition is formulated in a buffer. 62. The method of embodiment 61, wherein the buffer comprises Bis-Tris Propane-HCl. 63. The method of any one of embodiments 61-62, wherein the buffer comprises MgCl₂. 64. The method of any one of embodiments 61-63, wherein the buffer comprises bovine serum albumin. 65. The method of any one of embodiments 61-64, wherein the buffer comprises from 0.1 to 50 mM Bis-Tris Propane-HCl. 66. The method of any one of embodiments 61-65, wherein the buffer comprises from 0.1 to 50 mM MgCl₂. 67. The method of any one of embodiments 61-66, wherein the buffer comprises from 1 to 500 µg/ml bovine serum albumin. 68. The method of any one of embodiments 61-67, wherein the buffer comprises 10 mM Bis-Tris Propane-HCl. 69. The method of any one of embodiments 61-68, wherein the buffer comprises 10 mM MgCl₂. 70. The method of any one of embodiments 61-69, wherein the buffer comprises 100 µg /ml of bovine serum albumin. 71. The

method of any one of embodiments 41-70, wherein the protein comprises a purification tag. 72. The method of embodiment 71, wherein the purification tag comprises at least one tag selected from the group consisting of a His tag, a FLAG tag, an AU1 epitope tag, an AU5 epitope tag, a bacteriophage T7 tag, a bacteriophage V5 epitope tag, a Bluetongue virus tag (B-tag), a Glu-Glu tag (EE-tag), an HSV epitope tag, a KT3 epitope tag, a Myc epitope tag, a PDZ ligand tag, a polyarginine tag, a polyaspartate tag, a polycysteine tag, a polyphenylalanine tag, a protein C tag, an S1-tag, an S-tag, a Step-tag, and a VSV-G tag. 73. The composition of any one of embodiments 41-72, wherein the guide RNA comprises an A-rich protospacer adjacent motif (PAM) sequence, a G-rich PAM sequence, a T-rich PAM sequence, or a C-rich PAM sequence. 74. The method of any one of embodiments 41-73, wherein the guide RNA comprises a T-rich PAM sequence. 75. The composition of any one of embodiments 73-74, wherein the PAM sequence comprises 3 T nucleobases, 2 T nucleobases, or 1 T nucleobase. 76. The method of any one of embodiments 73-75, wherein the PAM sequence comprises TTTN. 77. The method of any one of embodiments 45-76, wherein the human nucleic acid sequence comprises a region in KRAS, HER2/neu, PD-1, TCR, p53, CCR5, DNMT1, EMX1, and LKB1. 78. The method of any one of embodiments 49-77, wherein the cancer cell comprises a bladder cancer cell, a bone cancer cell, a blood cancer cell, a breast cancer cell, a black colored tumor cell, a thyroid cancer cell, a parathyroid cancer cell, a bone marrow cancer cell, a laryngopharyngeal cancer cell, a laryngeal cancer cell, a lung cancer cell, an esophagus cancer cell, a pancreatic cancer cell, a colorectal cancer cell, a gastric cancer cell, a tongue cancer cell, a skin cancer cell, a brain tumor cell, a uterine cancer cell, a head or neck cancer cell, a gallbladder cancer cell, an oral cancer cell, a central nervous system tumor cell, or a liver cancer cell. 79. The method of any one of embodiments 45-78, wherein the composition exhibits at least 2-fold than AsCas12a, FnCas12a, or LbCas12a. 80. The method of any one of embodiments 45-79, wherein the guide RNA comprises a crRNA and a tracrRNA. 81. The method of embodiment 80, wherein the crRNA comprises a 5' repeat recognition sequence of AAUU. 82. The method of any one of embodiments 45-81, wherein protein exhibits cleavage activity in the presence of CaCl₂, CoCl₂, FeCl₂, MnSO₄, or any combination thereof. 83. The method of any one of embodiments 41-82, wherein the guide RNA sequence comprises from 1 to 100 nucleotides. 84. A method of improving cleaving efficiency of a type V CRISPR-associated protein, the method comprising providing the type V CRISPR-associated protein; identifying a residue at position 925 or 930; and mutating the residue at position 925 or 930 to Lysine, thereby improving cleaving efficiency of the type V CRISPR-associated protein.

EXAMPLES

[0192] These examples are provided for illustrative purposes only and not to necessarily limit the scope of the inventive subject matter provided herein.

EXAMPLE 1

Discovery of Cas12a Proteins by Excavation of the Metagenome

[0193] This example illustrates discovery of Cas12a proteins by excavation of the metagenome.

FIG. 1 shows a schematic of the Cas12a excavation process from the metagenome. Metagenome base sequences were downloaded from the NCBI Genbank BLAST database and were established as a local BLASTp database. Next, 16 Cas12a protein sequences and various CRISPR associated protein 1 (Cas1) amino acid sequences were downloaded from the Uniprot database. MetaCRT was used to identify CRISPR repeat and spacer sequence in the metagenome base sequences.

Metagenome sequence having CRISPR sequences were extracted and genes were predicted using the Prodigal program.

[0194] A taxonomic hierarchy was built based on the predicted genes within a 10 kilobase scope of a CRISPR sequence and the Cas12a amino acid sequence was used to predict homologs for Cas12a proteins from the predicted genes. Using the Cas1 gene, it was predicted whether the taxonomic hierarchy of Cas12a homolog has Cas1 homolog, and Cas12a genes that are within 800 amino acids (aa) to 1500 aa scope of Cas1 were selected.

[0195] Non-broken Cas12a proteins aligned using the MAFFT (multiple alignment using fast fourier transform) program to align sequences, run neighbor joining (NJ) methods, and dendrograms were constructed with 100x bootstrap using MEGA7. A maximum-likelihood 1000x bootstrap dendrogram using MEGA7 was constructed using the Cas12a amino acid sequence discovered herein and by selecting genes that form a monophyletic group with known Cas12a genes, to identify any evolutionary relationship.

[0196] **FIGS. 20A-C** show schematics of the mgCas12a proteins of the present disclosure. **FIG. 20A** shows a condensed schematic relative to **FIG. 1**, showing the pipeline for mining Cas12a proteins from metagenome data. **FIG. 20B** shows a phylogenetic tree of metagenome-derived Cas12a proteins of the disclosure and other Cas12a orthologs. Bootstrap values are displayed at each node. Sequences used in the present disclosure include FnCas12a, LbCas12a, AsCas12a, mgCas12a-1, and mgCas12a-2. **FIG. 20C** shows schematics of functionally-characterized novel Cas12a's and AsCas12a (Yamano *et al.* 2016). Protein domains were predicted using structure-based alignments with AsCas12a amino acid sequence. Absence of Cas elements in mgCas12a-1 and mgCas12a-2 are shown by dotted lines. Site-directed mutagenesis residues are marked with a black wedge. Sequence identifiers including contig title and GenBank accession number are at the

right of each schematic. Acronyms in the schematics include: Cas (CRISPR-associated genes), WED (wedge domain), REC (recognition domain), PI (PAM interaction domain), RuvC (RuvC nuclease domain), BH (bridge helix domain), and Nuc (Nuclease domain). **FIG. 2** shows a dendrogram of Cas12a, demonstrating that the metagenomics mining methods disclosed herein were successful. **FIG. 21** shows an unrooted and evolutionary distance-based phylogenetic tree of metagenome-derived Cas12a of the present disclosure and other orthologs. Orthologs used in this paper include FnCas12a, LbCas12a, AsCas12a, mgCas12a-1, and mgCas12a-2.

EXAMPLE 2

Synthesis of Gene and Guide RNA

[0197] This example illustrates synthesis of gene and guide RNA (gRNA). Structure-based alignment of novel Cas12a proteins, AsCas12a proteins, and LbCas12a proteins was performed with the ESPript program. Novel Cas12a proteins that displayed homology to AsCas12a were selected and were substituted with an amino acid residue yielding critical functional loss, as with AsCas12a amino acid residues at the same position. Sequences that exhibited critical functional loss included a K925Q mutant of SEQ ID NO: 1 (mgCas12a-1) and K930Q mutant of SEQ ID NO: 3 (mgCas12a-2).

[0198] Codon usage for humans, Arabidopsis, and colon bacillus was considered and codon optimization was performed. The base sequence of SEQ ID NO: 1 (mgCas12a-1) and SEQ ID NO: 3 (mgCas12a-2) that have been human codon optimized are shown in SEQ ID NO: 7 and SEQ ID NO: 8, respectively. **FIG. 3 – FIG. 8** shows an alignment of Cas12a proteins of the present disclosure, including SEQ ID NO: 1 (mgCas12a-1), SEQ ID NO: 3 (mgCas12a-2), SEQ ID NO: 9 (AsCas12a), SEQ ID NO: 10 (LbCas12a), and SEQ ID NO: 11 (FnCas12a). **FIG. 9A** shows a chart of characteristics of Cas12a proteins of the present disclosure, including Cas12a proteins discovered by metagenomics mining (*e.g.*, SEQ ID NO: 1 (mgCas12a-1) and SEQ ID NO: 3 (mgCas12a-2), AsCas12a, LbCas12a, and FnCas12a). **FIG. 9B** shows a chart of amino acid sequence identities (%) between Cas12a orthologs. AsCas12 has less than 40% sequence identity to all other orthologs in the table. LbCas12a and FnCas12a have less than 40% sequence identity to mgCas12a-1 and mgCas12a-2. LbCas12a has between 40% and 50% sequence identity to FnCas12a. mgCas12a-1 has greater than 50% sequence identity to mgCas12a-2.

[0199] T4 ligation on pET28a-KanR-6xHis-BPNLS vectors was performed with a gene cloned to the pUC57 vector. Restriction cloning was also performed. Cloned vectors were transformed to colon bacillus DH5a and the Rosetta strain. The 5'-handle sequence of crRNA was extracted from the metagenome CRISPR repeat sequence, the RNA structure was modeled and synthesized with

DNA oligonucleotides, sequences were transcribed, and concentration was confirmed using FLUOstar Omega and MEGAshortscript T7 RNA transcriptase.

EXAMPLE 3

Protein Expression and Refinement

[0200] This example illustrates protein expression and refinement. Using Rosetta (DE3), 5 mL of colon bacillus was cultivated overnight. Inoculation was carried out using 500 mL of Terrific Broth (TB) and 100 mg/ml kanamycin. Cultures were cultivated in 37°C culture medium until an OD₆₀₀ of 0.6 was measured and additionally assayed for protein expression at 16 to 18 hours in 22°C after processing with 0.4 µM IPTG (Isopropyl β-D-1-thiogalactopyranoside). After centrifuging, the cells were dissolved in 10 mL of buffer (20 mM HEPES pH 7.5, 100 mM KCl, 20 mM imidazole, 10% glycerol and EDTA-free protease inhibitor cocktail) and cells were redispersed by sonication. Cells were centrifuged three times for 20 minutes at 6000 rpm and filtered using a 0.22 micron filter.

[0201] Using nickel column (HisTrap FF 5 ml) and 300 mM imidazole buffer, the protein was cleaned, eluted, and purified by affinity-chromatography. Protein size was confirmed by SDS-PAGE electrophoresis and was dialyzed overnight using 20 mM HEPES pH 7.5, 100 mM KCl, 1 mM DTT, 10% glycerol. Based on the size of the protein, selective filtering and condensation (Amicon Ultra Centrifugal Filter 100,000 MWCO) was performed. Using the Bradford Protein Assay, protein concentration was measured and protein was stored at -80°C.

EXAMPLE 4

In Vitro Cleavage Analysis

[0202] This example illustrates in vitro cleavage analysis. Xylosyltransferase of *Lactuca sativa* was amplified by PCR, the PAM (protospacer adjacent motif) sequence was predicted, and guide RNA was designed. RNP (ribonucleoprotein) compounds of mgCas12a-1 (SEQ ID NO: 1) and mgCas12a-2 (SEQ ID NO: 3) were brought to room temperature for 20 minutes and protein and RNA were at a molar ratio of 1:1.25. After processing the refined xylosyltransferase PCR product, RNPs were resuspended in various buffers including NEBuffer 1.1 (1X buffer components, 10 mM Bis-Tris-Propane-HCl, 10 mM MgCl₂ and 100 µg/ml BSA), NEBuffer 2.1 (1X buffer components, 50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂ and 100 µg/ml BSA) and NEBuffer 3.1 (1X buffer components, 100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl₂ and 100 µg/ml BSA) and in vitro cleavage analysis was performed at 37°C. NEBuffer 1.1, NEBuffer 2.1 and NEBuffer 3.1 had pH values of 7.0, pH 7.9 and pH 7.9 at 25°C, respectively. Reactions were run until completion and

stopped after 10 minutes of incubation at 65°C. The reaction products were confirmed via 1.5% agarose gel electrophoresis. **FIGS. 10-12** show gel electrophoresis results.

[0203] As described in **FIGS. 10-12**, when mgCas12a-1 and crRNA was formulated in the NEBuffer 1.1 buffer, target dsDNA was cut. Moreover, if mgCas12a-2 and crRNA was formulated in the NEBuffer 1.1, the target dsDNA was cut. Thus, mgCas12a-1 and mgCas12a-2 were active at a pH of 7.0.

[0204] **FIG. 16** shows an *in vitro* cleavage assay of Cas12a nucleases including mgCas12a-1 (wild-type), he_mgCas12a-1 (humanized and engineered mgCas12a-1), de_mgCas12a-1 (dead and engineered mgCas12a), mg Cas12a-2, he_mgCas12a-2, de_mgCas12a-2, AsCas12a, FnCas12a, and LbCas12a. The concentration of each RNP was 300 nM, the curly brace indicates the cleaved templates in the gel electrophoresis, which resolved at 1.8 kb and 0.65 kb. The samples of RNPs were incubated for 1 h at 37°C with 1X NEBuffer. mgCas12a and de_mgCas12a were also humanized. Materials and methods used in the *in vitro* cleavage assay included a protein to gRNA molar ratio of 1:1.25. In a 20 µL reaction, 300 nM (911.4 ng) of protein (FnCas12a-BPNLS, 158.82 kDa) was used. In a 20 µL reaction, 375 nM (102.5 ng) of crRNA (LsXTb12) was used. Reactions were run for 1 hour. 300 ng of template DNA was used and incubation was carried out for 1 h at 37°C. All Cas12a nucleases cleaved the template DNA (2.45 kB) into two pieces of 1.8 kB and 0.65 kB. mgCas12a-1, he_mgCas12a-1, mgCas12a-2, he_mgCas12a-2, AsCas12a, FnCas12a and LbCas12a cut the template DNA in two, while template DNA mixed with de_mgCas12a-1 and de_mgCas12a-2 resolved at the uncut size.

[0205] **FIG. 17A-B** shows an *in vitro* cleavage assay of Cas12a nucleases including mgCas12a-1 (wild-type), he_mgCas12a-1 (humanized and engineered mgCas12a-1), de_mgCas12a-1 (dead and engineered mgCas12a), mg Cas12a-2, he_mgCas12a-2, de_mgCas12a-2, AsCas12a, FnCas12a, and LbCas12a after reaction times of 12 h (**FIG. 17A**) and 24 h (**FIG. 17B**). The concentration of each RNP was 300 nM, the curly brace indicates the cleaved templates in the gel electrophoresis, which resolved at 1.8 kb and 0.65 kb. The samples of RNPs were incubated for 12 h and 24 h at 37°C with 1X NEBuffer. mgCas12a and de_mgCas12a were also humanized. Materials and methods used in the *in vitro* cleavage assay included a protein to gRNA molar ratio of 1:1.25. In a 20 µL reaction, 300 nM (911.4 ng) of protein (FnCas12a-BPNLS, 158.82 kDa) was used. In a 20 µL reaction, 375 nM (102.5 ng) of crRNA (LsXTb12) was used. Reactions were run for 12 h and 24 h at 37°C and 300 ng of template DNA was used. Samples were incubated for 12 h and 24 h at 37°C. AsCas12a, FnCas12a, and LbCas12a fully degraded template DNA in 12 h, indicating that they have interminable dsDNase activity. mgCas12a-1, he_mgCas12a-1, mgCas12a-2, and he_mgCas12a-2 exhibited less interminable dsDNase activity and template DNA remained after a 24 h reaction.

Samples incubated with de_mgCas12a-1 and de_mgCas12a-2 also lost interminable dsDNase activity.

[0206] FIG. 18 shows an *in vitro* cleavage assay of target plasmid DNA with Cas12a nucleases including mgCas12a-1 (wild-type), he_mgCas12a-1 (humanized and engineered mgCas12a-1), de_mgCas12a-1 (dead and engineered mgCas12a), mg Cas12a-2, he_mgCas12a-2, de_mgCas12a-2, AsCas12a, FnCas12a, and LbCas12a. The concentration of each RNP was 300 nM, an arrow shows cleaved templates, which are 6 kb of the linearized product. The samples of RNPs were incubated for 1 h at 37°C with 1X NEBuffer. mgCas12a and de_mgCas12a were also humanized. Materials and methods used in the *in vitro* cleavage assay included a protein to gRNA molar ratio of 1:1.25. In a 20 µL reaction, 300 nM (911.4 ng) of protein (FnCas12a-BPNLS, 158.82 kDa) was used. In a 20 µL reaction, 375 nM (102.5 ng) of crRNA (LsXTb12) was used. Reactions were run for 1 h at 37°C. 300 ng of template plasmid DNA was used and samples were incubated for 1 h at 37°C. All Cas12a nucleases cleaved template DNA, from 10 kb to its linearized form of 6 kb in size. mgCas12a-1, he_mgCas12a-1, mgCas12a-2, he_mgCas12a-2, AsCas12a, FnCas12a, and LbCas12a cut and linearized the target plasmid DNA, while target plasmid DNA incubated with de_mgCas12a-1 and de_mgCas12a-2 remained at the uncut size.

[0207] FIG. 19A-B shows an *in vitro* cleavage assay of target plasmid DNA with Cas12a nucleases including mgCas12a-1 (wild-type), he_mgCas12a-1 (humanized and engineered mgCas12a-1), de_mgCas12a-1 (dead and engineered mgCas12a), mg Cas12a-2, he_mgCas12a-2, de_mgCas12a-2, AsCas12a, FnCas12a, and LbCas12a after reaction times of 12 h (**FIG. 19A**) and 24 h (**FIG. 19B**). The concentration of each RNP was 300 nM, an arrow shows cleaved templates, which are 6 kb of the linearized product. The samples of RNPs were incubated for 12 h and 24 h at 37°C with 1X NEBuffer. mgCas12a and de_mgCas12a were also humanized. Materials and methods used in the *in vitro* cleavage assay included a protein to gRNA molar ratio of 1:1.25. In a 20 µL reaction, 300 nM (911.4 ng) of protein (FnCas12a-BPNLS, 158.82 kDa) was used. In a 20 µL reaction, 375 nM (102.5 ng) of crRNA (LsXTb12) was used. Reactions were run for 12 h and 24 h at 37°C and the amount of template DNA used was 300 ng. Samples were incubated for 12 h and 24 h at 37°C. AsCa12a, FnCas12a, and LbCas12a degraded all template plasmid DNA within 12 hours, indicating that these nucleases exhibited interminable dsDNase activity. mgCas12a-1, he_mgCas12a-1, mgCas12a-2 and he_mgCas12a-2 exhibited less interminable dsDNase activity and template plasmid DNA remained after a 24 h reaction. mgCas12a-1 and de_mgCas12a-2 also lost interminable dsDNase activity.

EXAMPLE 5**Genome Editing with the Cas12a Protein of SEQ ID NO: 1**

[0208] This example illustrates genome editing with the Cas12a protein of SEQ ID NO: 1. SEQ ID NO: 1 is recombinantly expressed or chemically synthesized. If recombinantly expressed, a Cas12a protein of SEQ ID NO: 1 comprises a purification tag used to purify the protein by affinity chromatography. The Cas12a protein of SEQ ID NO: 1 is coupled with a guide RNA (also referred to as crRNA) having a sequence that is reverse complementary to a nucleic acid sequence of interest. The nucleic acid sequence of interest is double stranded DNA (dsDNA) and is from a human. The guide RNA and the Cas12a protein of SEQ ID NO: 1 are administered to a subject. The subject is a human. The guide RNA guides the Cas12a protein of SEQ ID NO: 1 to the nucleic acid sequence of interest and the Cas12a protein of SEQ ID NO: 1 cleaves the dsDNA.

EXAMPLE 6**Genome Editing with the Cas12a Protein of SEQ ID NO: 3**

[0209] This example illustrates genome editing with the Cas12a protein of SEQ ID NO: 3. A Cas12a protein of SEQ ID NO: 3 is recombinantly expressed or chemically synthesized. If recombinantly expressed, the Cas12a protein of SEQ ID NO: 3 comprises a purification tag used to purify the protein by affinity chromatography. The Cas12a protein of SEQ ID NO: 3 is coupled with a guide RNA (also referred to as crRNA) having a sequence that is reverse complementary to a nucleic acid sequence of interest. The nucleic acid sequence of interest is double stranded DNA (dsDNA) and is from a human. The guide RNA and the Cas12a protein of SEQ ID NO: 3 are administered to a subject. The subject is a human. The guide RNA guides SEQ ID NO: 3 to the nucleic acid sequence of interest and the Cas12a protein of SEQ ID NO: 3 cleaves the dsDNA.

EXAMPLE 7**Engineering Cells with the Cas12a Protein of SEQ ID NO: 1**

[0210] This example illustrates engineering cells ex vivo with the Cas12a protein of SEQ ID NO: 1. A Cas12a protein of SEQ ID NO: 1 is recombinantly expressed or chemically synthesized. If recombinantly expressed, the Cas12a protein of SEQ ID NO: 1 comprises a purification tag used to purify the protein by affinity chromatography. The Cas12a protein of SEQ ID NO: 1 is coupled with a guide RNA (also referred to as crRNA) having a sequence that is reverse complementary to a nucleic acid sequence of interest. The nucleic acid sequence of interest is double stranded DNA (dsDNA) and is from a human. The guide RNA and the Cas12a protein of SEQ ID NO: 1 are administered to a plurality of cells. The guide RNA guides the Cas12a protein of SEQ ID NO: 1 to the nucleic acid sequence of interest and the Cas12a protein of SEQ ID NO: 1 cleaves the dsDNA,

thereby editing the plurality of cells. The cells are edited to knock out an aberrant gene or to introduce a functional gene. The edited cells are administered to a subject in need thereof. The subject is a human and has cancer. The cancer is gastric cancer, colorectal cancer, liver cancer, lung cancer, or breast cancer.

EXAMPLE 8

Engineering Cells with the Cas12a Protein of SEQ ID NO: 3

[0211] This example illustrates engineering cells *ex vivo* with the Cas12a protein of SEQ ID NO: 3. A Cas12a protein of SEQ ID NO: 3 is recombinantly expressed or chemically synthesized. If recombinantly expressed, the Cas12a protein of SEQ ID NO: 3 comprises a purification tag used to purify the protein by affinity chromatography. The Cas12a protein of SEQ ID NO: 3 is coupled with a guide RNA (also referred to as crRNA) having a sequence that is reverse complementary to a nucleic acid sequence of interest. The nucleic acid sequence of interest is double stranded DNA (dsDNA) and is from a human. The guide RNA and the Cas12a protein of SEQ ID NO: 3 are administered to a plurality of cells. The guide RNA guides the Cas12a protein of SEQ ID NO: 3 to the nucleic acid sequence of interest and the Cas12a protein of SEQ ID NO: 3 cleaves the dsDNA, thereby editing the plurality of cells. The cells are edited to knock out an aberrant gene or to introduce a functional gene. The edited cells are administered to a subject in need thereof. The subject is a human and has cancer. The cancer is gastric cancer, colorectal cancer, liver cancer, lung cancer, or breast cancer.

EXAMPLE 9

In Vitro Cleavage Analysis of Cpf1 Proteins

[0212] This example illustrates *in vitro* cleavage analysis of Cpf1 proteins. Proteins and gRNA were co-incubated at a 1:1.2 molar ratio. In the case of 20 μ l reactions, 100 nM (320 ng) of the protein FnCas12a-BPNLS (159.81 kDa) was used. In the case of 20 μ l reactions, 120 nM (80 ng) of the crRNA DHCR7 was used. Reaction times of 1 hour and 4 hours at 37°C were tested and 200 ng of template DNA was used. **FIG. 14** shows the results of an *in vitro* cleavage assay using various nucleases including, FnCas12, AsCas12, LbCas12, He-MgCas12a-1 (humanized and engineered mgCas12a-1), and He-MgCas12a-2 (humanized and engineered mgCas12a-2), 1 hour after incubation of the target with the nucleases and 4 hours after incubation of the target with the nucleases. The concentration of each RNP was 100 nM. Arrows indicate the cleaved fragments of the template, which were resolved at 680 base pairs (bps) and 750 bps. Mock A indicates a sample in which neither gRNA or protein were used in the cleavage assay. Mock B indicates a sample in

which no protein was used in the cleavage assay. He-MgCas12a-1 and He-MgCas12a-2 both cleaved the DNA template (1,430 bp) into two pieces of 750 bp and 680 bp.

[0213] By 1 h, He-MgCas12a-1 had completely cut the template DNA in two while some remaining template DNA at the uncut size was evident when incubated with He-MgCas12a-2. By 4 h, He-MgCas12a-2 had also completely cut the template DNA. The cleavage efficiencies of He-MgCas12a-1 and He-MgCas12a-2 were higher than the other three Cas12 nucleases, as uncut template DNA remained after incubation with the other three Cas12 nucleases (FnCas12, AsCas12, and LbCas12) at 1 hour. Additionally, the FnCas12 lane indicated that both template and cleaved products were degraded over time, while neither the He-MgCas12a-1 nor the He-MgCas12a-2 lanes indicated any further DNA degradation.

EXAMPLE 10

Genome Editing of rice and *N. benthamiana*

[0214] This example illustrates genome editing of rice and *N. benthamiana*. Two crRNAs in rice and three crRNAs in *N. benthamiana* were used to evaluate genome editing efficiencies of He-MgCas12a-1 versus FnCas12. Genome editing efficiency values were measured using amplicon targeted deep sequencing. Plants and other materials used in these assays include 20-30 seeds of lettuce (*Lactuca sativa* var. Chungchima), MS salt with vitamins (M0222, Duchefa, RV Haarlem, Netherlands), razor blades (NO.10, FEATHER SAFETY RAZOR, Osaka, Japan), forceps (Cat. 3-SA, Jonostick by regine Switzerland Standard, China), cell strainer (Cat. 93100, SPL, Korea), 1000 µl wide-bore tip (T-205-WB-C-R-S, Axygen, NY), growth chamber 24°C (HB103M, HanBaek Scientific Co., Korea), pH meter (STARA2115, ThermoFisher scientific, Waltham, MA, USA), and a sterilizer (Cat.BF-60AC, BioFree, Korea). PEG transfection was carried out using 1) an enzyme solution, which included mannitol (M0803, Duchefa, RV Haarlem, Netherlands), KCl (P5405, Sigma-Aldrich, USA), MES (M1503, Duchefa, RV Haarlem, Netherlands), CaCl₂ (C3881, Sigma-Aldrich, Japan), BSA (A9056, Sigma-Aldrich, USA), cellulase R-10 (Yakult Pharmaceutical Inc., Tokyo, Japan), and macerozyme R-10 (Yakult Pharmaceutical Inc., Tokyo, Japan), 2) a PEG solution, which included NaCl (7548-4405, Daejung chemicals and metals, Korea), KCl (P5405, Sigma-Aldrich, USA), CaCl₂(C3881, Sigma-Aldrich, Japan), and MES (M1503, Duchefa, RV Haarlem, Netherlands), and 3) a MMG solution, which included mannitol (M0803, Duchefa, RV Haarlem, Netherlands), MgCl₂ (M0533, Duchefa, RV Haarlem, Netherlands), and MES (M1503, Duchefa, RV Haarlem, Netherlands).

[0215] Plant transformation and regeneration was carried out as follows. Plants and reagents for protoplast transformation included first sterilizing lettuce seeds with a 2% sodium hypochlorite (Clorox) for 10 min, washing seeds 5 times with sterile dH₂O, and planting the sterile seeds on ½

MS media. Lettuce leaves were harvested 5 days after germination for protoplast preparation. 40 mL of enzyme solution was made with 0.4 M mannitol, 20 mM KCl, 20 mM MES (pH 5.7), 1.5 % Cellulase R-10 (Yakult), and 0.3% Macerozyme R-10 (Yakult). Incubations were carried out at 55°C for 10 min, 10 mM CaCl₂ and 0.1 %BSA was added, and enzyme solution was filtered through a 0.45 µm syringe filter.

[0216] Protoplast preparation included cutting ten to fifteen leaves from rice or *N.benthamiana* plantlets with a razor, piling two or three leaves on a droplet of sterile water, and slicing piled leaves together. 20 mL of the enzyme solution was poured into a 90 mm diameter plate and fifteen sliced leaves were transferred into the 20 mL enzyme solution. The solution was covered with aluminum foil. The 90 mm plate was placed on a gyratory shaker at 50 rev/min and the plate was incubated for four to five hours. The enzyme solution with protoplasts were poured into a round tube and the same volume of W5 solution was added to the 20 mL enzyme solution. W5 solution included 154 mM NaCl, 125 mM CaCl₂, 5 mM KCl, and 2 mM MES (pH 5.7). 40 mL of enzyme solution containing protoplasts were flown through a 100 µm cell strainer into a 50 mL round tube. The cell strainer was removed and the 50 mL tube was centrifuged at 100 g (or 80 g in Hanil) for 5 min. The supernatant was removed using a 20 mL long pipette. 1 mL of MMG solution was added. 5 mL of MMG solution was made by mixing 2.5 mL of 0.8 M mannitol, 0.25 mL of 300 mM MgCl₂, and 0.1 mL of 200 mM MES (pH 5.7). 10 mL of MMG solution was made by mixing 5 mL of 0.8 M mannitol, 0.5 mL of 300 mM MgCl₂, and 0.2 mL of 200 mM MES (pH 5.7). Protoplasts were counted with a hemacytometer, cell numbers were adjusted up to 2×10^6 cells/mL by adding MMG solution. 200 µl containing 2×10^5 /mL of protoplasts were aliquoted into 1.5 mL tubes.

[0217] Protoplasts were transformed with CRISPR RNPs. A 20 µl transformation reaction was set up in a 1.5 mL tube as shown below in **TABLE 3**.

TABLE 3: Transformation Reaction

RNP	2×10^5/mL Protoplasts
sgRNA	5 µg
Cas9 protein	10 µg
Plus reagent™	2 µl
Lipofectamine™ 3000	2 µl
NEB Buffer 3.1	2 µl
dH ₂ O up to	20 µl

[0218] Both Lipofectamine™ 3000 and Plus reagent™ transfection reagents were utilized for RNP deliver with PEG 4000. RNP can be replaced with Cpf1 or other Cas proteins. GFP-Cas9 was employed to help trace Cas9 localization instead of Cas9.

[0219] The RNP transformation mixture was incubated for 10 min at room temperature. 200 µl of the protoplast solution was aliquoted with a 1,000 µl wide bore tip into a clean 1.5 mL tube. The RNP mixture was added to the 200 µl protoplast solution and mixed gently. The same volume (200 µl) of 40% PEG solution (shown below in **TABLE 4**) into the RNP-protoplast solution.

TABLE 4: PEG Solution

40% PEG Solution Ingredients	5 ml	10 ml
0.8 M Mannitol	1.25 ml	2.5 ml
1M CaCl ₂	0.5 ml	1 ml
PEG 4000	2 g	4 g
dH ₂ O up to	5 ml	10 ml

[0220] The RNP-protoplast-PEG solution was gently pipetted five to ten times. The RNP-protoplast-PEG solution was placed at room temperature for 10 min. 800 µl of W5 solution was added to the RNP-protoplast-PEG solution and inverted four to five times. The tubes were centrifuged at 100 g for 1 min in a large tabletop centrifuge and the supernatant was discarded. 200 µl of W5 solution was added to samples and the samples were incubated for 4 hours at 28°C. Protoplasts were harvested and the genomic DNA was extracted. The target DNA region was amplified and amplicon targeted sequencing was performed.

[0221] The amplicon setup was carried out as follows. Working on ice, primers were added using a multichannel pipettes by adding 4 µl from the vertical i5(S5XX) primer strip to the columns of the plate and 4 µl from the horizontal i7(N7XX) primer strip to the rows of the plate. 10 µl of polymerase was added from the 8-well strip into each well, 2 µl of 0.5 ng/µl DNA was transferred into the corresponding wells. Plates were sealed, vortexed, and spun in a plate centrifuge for 2 mins at 2000 g. PCR was carried out using the cycling conditions outlined in **TABLE 5**. Either a qPCR machine or a standard PCR machine can be used to when using the KAPA HiFi mix. qPCR had the advantage of monitoring how each sample was amplified in real time, since each contained SYBR green.

TABLE 5: PCR Cycling Conditions

	Temperature	Time	Cycles
Denaturation	98	2 mins	1X
Denaturation	98	30	
Annealing	60	30	10-12 cycles
Extension	72	30	
Final	72	5 mins	1X
Hold	4		

[0222] The second round PCR products were cleaned as follows. 30 µl of H₂O was added to each well of the PCR plate to bring the total volume per well to 50 µl, 50 µl of AmpureXP beads was added to each well of a 96-well round bottom plate, and 50 µl of PCR product was transferred to the 96-well plate containing 50 µl AmpureXP mix and the solution was pipetted up and down 10x to mix. The samples were incubated for at least 10 minutes on the bench. The plate was placed on a 96-well plate magnet for 5 mins until the liquid appears clear. The supernatant was discarded by pipetting and aspirating. Samples were washed by adding 190 µl of 70-80% ethanol to each sample and left for 30 seconds. The ethanol was discarded by pipetting and aspiration. Washing samples with ethanol and discarding ethanol, as described above, was repeated for a total of two washes. Plates were removed from the magnet and allowed to air dry for 2-3 minutes and it was ensured that no ethanol was detected. The plate was taken off the magnet and the beads were resuspended in 22 µl of H₂O and mixed thoroughly by pipetting up and down 10 times. The plates were allowed to incubate for at least 10 mins on the bench. The plate was placed back on the magnet for 5 minutes and it was ensured that the liquid appeared clear. Finally, 20 µl of the supernatant was transferred to a new 96-well PCR plate.

[0223] PCR products were pooled and the library was quantified as follows. 10 µl of each second round PCR product was transferred from the plate into a single microcentrifuge tube. The concentration of DNA was determined using the Qubit. The DNA concentration was adjusted to 2 nM.

[0224] Sequencing was carried out using the MiSeq loading protocol as per Illumina instructions. Primer sequences are shown below in **TABLE 6**.

TABLE 6: Primer Sequences

	Target Specific Primer Sequence	Adapter Primer Sequence
Rice DWF5 crRNA #1		
>NGS_OsDwarf5_F2	ATTCCAGGGAATGGA ACTAT (SEQ ID NO: 31)	TCGTCGGCAGCGTCAGATGTGTATAA GAGACAG ATTCCAGGGAATGGA ACTAT (SEQ ID NO: 32)
>NGS_OsDwarf5_R2	TATTGGATAGCAACC AAAGC (SEQ ID NO: 33)	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAG TATTGGATAGCAACC AAAGC (SEQ ID NO: 34)
Rice DWF5 crRNA #2		
>1273 NGS OsDwarf5 F3	GGTGAGCTTATTTAT TAGGCTT (SEQ ID NO: 35)	TCGTCGGCAGCGTCAGATGTGTATAA GAGACAG GGTGAGCTTATTTAT TAGGCTT (SEQ ID NO: 36)

	Target Specific Primer Sequence	Adapter Primer Sequence
>1274 NGS OsDwarf5 R3	GGTGAAGAATGTCATCGCTAAT (SEQ ID NO: 37)	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG GGTGAAGAATGTCATCGCTAAT (SEQ ID NO: 38)
Tobacco XTb12 crRNA #1		
>1267 NGS XTb12_1/2 F1	AAATCCCCC AAAACCACTTTT (SEQ ID NO: 39)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG AAATCCCCC AAAACCACTTTT (SEQ ID NO: 40)
>1268 NGS XTb12_1/2 R1	CGGTGTTATCGCCGAATTTCCG (SEQ ID NO: 41)	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG CGGTGTTATCGCCGAATTTCCG (SEQ ID NO: 42)
Tobacco XTb12 crRNA #2		
>1269 NGS NbXTb12_1/2 F2	GGTTTACTCTCAAAGTTGACCT (SEQ ID NO: 43)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG GGTTTACTCTCAAAGTTGACCT (SEQ ID NO: 44)
>1270 NGS NbXTb12_1/2 R2	GGGCAGCTCATCATCTTCATTC (SEQ ID NO: 45)	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG GGGCAGCTCATCATCTTCATTC (SEQ ID NO: 46)
Tobacco XTb12 crRNA #3		
>1271 NGS NbXTb12_1/2 F3	CTCTGTACTAAGTAGTACACAC (SEQ ID NO: 47)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG CTCTGTACTAAGTAGTACACAC (SEQ ID NO: 48)
>1272 NGS NbXTb12_1/2 R3	GCTTGGAATATTGAG AAGTGAT (SEQ ID NO: 49)	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG GCTTGGAATATTGAGAAGTGAT (SEQ ID NO: 50)

[0225] FIG. 15A illustrates that FnCas12a exhibited genome editing efficiencies in rice of 0.5%, 0.3%, and 0.9% in crRNA1-1, crRNA1-2, crRNA2, respectively and He-MgCas12a-1 exhibited genome editing efficiencies in rice of 1.9%, 0.7%, and 10.2% in crRNA1-1, crRNA1-2, crRNA2. FIG. 15B illustrates that FnCas12a exhibited genome editing efficiencies in *N.benthamiana* of 0.8%, 1.4%, and 4.8% in crRNA1, crRNA2, and crRNA3, respectively and He-MgCas12a-1 exhibited genome editing efficiencies in *N.benthamiana* of 0.7%, 3.7%, and 3.4% in crRNA1, crRNA2, and crRNA3, respectively.

EXAMPLE 11

Genome Editing of CCR5 and DNMT1

[0226] This example illustrates genome editing of CCR5 and DNMT1.

[0227] Cell culture. HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, penicillin, and streptomycin.

[0228] crRNA sequences. Nucleotide sequences of synthetic crRNAs were obtained from IDT for CCR5 and DNMT1 and are listed below in **TABLE 7**.

TABLE 7 – crRNA Sequences for Targeting CCR5 and DNMT1

Genes	crRNA Sequence (5'-3')
CCR5	CACCGAAUUUCUACUGUUGUAGAUGGAGUGAAGGGAGAGUUUGUCAAUU UUUUG (SEQ ID NO: 51)
DNMT1	GGUCAAUUUCUACUGUUGUAGAUGCUCAGCAGGCACCUGCCUCUUUU (SEQ ID NO: 52)

[0229] RNP preparation and electroporation. Before transfection of proteins in cells, purified mgCas12a-1 and mgCas12a-2 proteins (100 pmol) were incubated with CCR5 or DNMT1 crRNA (200 pmol) at room temperature for 20 minutes to form the RNP complex. Nucleofection of HEK293T cells was performed using Lonza. Each nucleofection reaction included mixing approximately 2×10^5 cells in 20 μ l of nucleofection reagent with 10 μ l of RNP.

[0230] Genomic DNA extraction. Cells were harvested at 48 and 72 hours after transfection. Genomic DNA extraction was performed using PureLink Genomic DNA kits (Invitrogen) following the manufacture's instruction.

[0231] Deep sequencing analysis of on-target sites. The genomic region flanking the target site for each gene was amplified using a two-step PCR method. First, the genomic DNA from the edited and control samples was isolated and PCR amplified for 35 cycles using Q5 High-fidelity DNA polymerase and adapter primers. Sequences of adapter primers are shown below in **TABLE 8**. The resulting amplicons were prepared using a QIAquick PCR Purification kit. These samples were subjected to eight cycles of PCR using KAPA HotStart DNA Polymerase for indexing, followed by AMPure bead purification. Purified DNA samples were quantified by Qubit 2.0 Fluorometer, size analyzed by BioAnalyzer, and pooled in an equimolar ratio. Sequencing libraries were sequenced with the Illumina MiniSeq. Data was analyzed using the Cas-Analyzer program.

TABLE 8 – Adapter Primer Sequences

Genes	Adapter Primer Sequence (5'-3')
CCR5	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGTATTTCTGTTTCAGATC AC (SEQ ID NO: 53)
	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCCCATCAATTATAGAA

	AGCC (SEQ ID NO: 54)
DNMT1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGCACACAGCAGGCCTT
	TG (SEQ ID NO: 55)
	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCCAATAAGTGGCAGA
	GTGC (SEQ ID NO: 56)

[0232] FIG. 27 shows the resulting editing efficiencies of mgCas12a-1, mgCas12a-2, and mock (negative control). mgCas12a-1 exhibited more efficient editing efficiency when targeting CCR5, as indicated by the higher percent indels at 48h and 72h, in comparison to the negative control (mock) and mgCas12a-2. mgCas12a-2 exhibited percent indels over background when targeting CCR5 at 72 h. Both mgCas12a-1 and mgCas12a-2 exhibited percent indels over background when target DNMT1 at 48 h and 72h, with mgCas12a-2 exhibiting slightly higher % indels than mgCas12a-1.

[0233] TABLE 9 below summarizes targeted deep sequencing data for indels of mgCas12a-1 and mgCas12a-2.

TABLE 9 – Targeted Deep Sequencing Data for Indels of mgCas12a-1 and mgCas12a-2

Samples	Gene	Time	Name	Total sequences	With both indicator sequences	Insertions	Deletions	Indel frequency	Indel frequency (%)
1	CCR5	48 h	Mock	137952	137475	0	187	187 (0.1%)	0.1
2			mgCas12a-1	119684	119250	36	418	454 (0.4%)	0.4
3			mgCas12a-2	112387	112077	8	150	158 (0.1%)	0.1
4		72 h	Mock	139323	138942	8	179	187 (0.1%)	0.1
5			mgCas12a-1	156705	156159	39	738	777 (0.5%)	0.5
6			mgCas12a-2	158717	158392	5	237	242 (0.2%)	0.2
7	DNMT1	48 h	Mock	141182	136856	19	316	335 (0.2%)	0.2
8			mgCas12a-1	122368	120871	70	424	494 (0.4%)	0.4

Samples	Gene	Time	Name	Total sequences	With both indicator sequences	Insertions	Deletions	Indel frequency	Indel frequency (%)
9		72 h	mgCas12a-2	121928	120592	46	509	555 (0.5%)	0.5
10			Mock	98480	96480	0	192	192 (0.2%)	0.2
11			mgCas12a-1	126317	123792	2	511	513 (0.4%)	0.4
12			mgCas12a-2	47398	47000	12	199	211 (0.5%)	0.5

EXAMPLE 12

Genome Editing of *Nicotiana benthamiana*

[0234] This example illustrates genome editing of CCR5 and DNMT1.

[0235] Plant growth condition. All plants were grown under a $150 \text{ E m}^{-2} \text{ s}^{-1}$ LED light under long-day (14-h light/10-h dark photoperiod) conditions at 25°C.

[0236] Protoplast transfection. Tobacco (*Nicotiana benthamiana*) and seeds were sterilized in a 0.4% hypochlorite solution for 1 min, washed three times in distilled water, and sown on a 0.5× Gamborg B5 solid medium supplemented with 2% sucrose. The 4-week-old leaves grown in B5 media were digested with enzymes (1.5% cellulose R10, 0.3% macerozyme R10, 0.5 M Mannitol, 8 mM CaCl_2 , 5 mM MES [pH 5.7], 0.1% BSA) for 4 h at 25°C in darkness. The mixture was filtered before protoplasts were collected by centrifugation at 100g in a round-bottomed tube for 6 min. Re-suspended protoplasts were washed with W5 solution (154 mM NaCl, 125 mM $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 5 mM KCl, 2 mM MES [pH 5.7]) and pelleted by centrifugation at 100g for 6 min. Finally, protoplasts were re-suspended in MMG solution (0.4 M mannitol, 15 mM MgCl_2 , 4 mM MES [pH 5.7]) and counted under a microscope using a hemocytometer. Protoplasts were diluted to a density of 1×10^6 protoplasts/ml of MMG solution and stabilized for at least for 30 min at 4°C before PEG-mediated transfection. 2×10^5 protoplast cells were transfected with Nuclease protein (10 µg of mgCas12a-1, mgCas12a-2, or FnCpf1) pre-mixed with in vitro-transcribed crRNA (22 µg). Prior to transfection, Nuclease proteins were mixed with crRNA in 1× NEB buffer 1 and incubated for 10-20 min at room temperature. A mixture of protoplasts re-suspended in 200 µl MMG solution was gently mixed with 10-20 µl of RNP complex and 210-220 µl of freshly prepared PEG solution (0.2M mannitol, 40%W/V PEG-4000, 100mM CaCl_2) and incubated at 25°C for 15 min. After a 15 min incubation at room temperature, transformation was stopped by

adding 840-880 μ L of W5 solution. Protoplasts were collected by centrifuging for 2 min at 100g at room temperature and washed once with 1 ml of wash buffer by centrifugation for another 2 min at 100g. Protoplasts were resuspended in 500 μ L of W5 solution and incubated for 2 days in a growth chamber at 25°C.

[0237] Targeted deep sequencing. The on-target was amplified from genomic DNA. Indices and sequencing adaptors were added by additional PCR. High-throughput sequencing was performed using Illumina MiniSeq. **TABLE 10** below shows the crRNA target region and sequence.

TABLE 10 – CRISPR RNA (crRNA) target region and sequence

Target Gene	crRNA (primer name)	crRNA sequence (PAM site)
NbFucT14_1 NbFucT14_2	NbFTa14_1/2-2	TTTGGATAATTTGTACTCTTGTCGATGT (SEQ ID NO: 57)
	NbFTa14_1/2-4	TTTAGTCCACAAACAGCTAAGCCCACAT (SEQ ID NO: 58)

[0238] TABLE 11 below shows primers used for target region PCR analysis.

TABLE 11 – Primers For Target-Region PCR Analyses

Target gene	Primer name	Sequence	Size (bp)
NbFucT14_1	NGS NbFTa14_1_F	TGAGCTGAAGATGGATTATG (SEQ ID NO: 59)	216
	NGS NbFTa14_1_R	TCATGCTTAAGATAAAAGAG (SEQ ID NO: 60)	
NbFucT14_2	NGS NbFTa14_2_F	TCATGAGCTTAAGATGGATC (SEQ ID NO: 61)	217
	NGS NbFTa14_2_R	GTTTAAGCTAAAAGAACTAC (SEQ ID NO: 62)	

[0239] TABLE 12 below shows Cas12a editing efficiency, complexed with two crRNAs, for two FucT14 targets, as measured by MiniSeq.

TABLE 12 – Cas12a Editing Efficiency

Target gene	crRNA	Nuclease	Total Sequences	With both indicator sequences	More than minimum frequency	Insertions	Deletions	Indel frequency
<i>FucT14-1</i>	2	none	161551	161421	160896	4	180	184 (0.1%)
		mgCas12a-1	124361	124255	123844	3	168	171 (0.1%)
		mgCas12a-2	99154	99053	98734	0	131	131 (0.1%)
		FnCpf1	50060	50022	49808	0	63	63 (0.1%)
	4	none	161551	161411	160899	4	178	182 (0.1%)

Target gene	crRNA	Nuclease	Total Sequences	With both indicator sequences	More than minimum frequency	Insertions	Deletions	Indel frequency
		mgCas12a-1	106782	106706	106330	0	1877	1877 (1.8%)
		mgCas12a-2	126665	126544	126057	79	885	964 (0.8%)
		FnCpf1	64554	64501	64272	15	470	485 (0.8%)
<i>FucT14-2</i>	2	none	49459	49422	49192	2	49	51 (0.1%)
		mgCas12a-1	81191	81101	80738	0	90	90 (0.1%)
		mgCas12a-2	83694	83614	83286	0	99	99 (0.1%)
		FnCpf1	108803	108682	108260	0	112	112 (0.1%)
	4	none	49459	49427	49199	2	49	51 (0.1%)
		mgCas12a-1	54918	54854	54532	6	689	695 (1.3%)
		mgCas12a-2	127825	127691	127213	2	143	145 (0.1%)
		FnCpf1	64265	64168	63882	0	162	162 (0.3%)

EXAMPLE 13

Cas12a Cleavage of Linear dsDNA and Circular dsDNA

[0240] This example describes Cas12a cleavage of linear dsDNA and circular dsDNA. FnCas12a, mgCas12a-1, and mgCas12a-2 were complexed with crRNA targeting HsCCR5, HsDNMT1, and HsEMX1 (Hs standing for human). Cas12a nucleases were tested for their ability to target and cleave linear dsDNA and circular DNA. Purified PCR product was used to obtain target linear dsDNA and purified plasmid DNA was used to obtain circular dsDNA. Conditions tested include no incubation and 2 hours incubation at 37°C. The appropriate negative controls were also tested, as summarized in the tables.

[0241] **FIG. 22A-B** show sequence-specific cleavage of dsDNA by crRNA guided-mgCas12a proteins of the present disclosure. **FIG. 22A** shows sequence-specific cleavage of linear dsDNA by crRNA guided-Cas12a proteins including FnCas12a, WT mgCas12a-1 and WT mgCas12a-2. The substrate is indicated by an arrow and the letter S and cleaved products in the gel are also shown with arrows. Purified PCR product was used for linear dsDNA. The numbers below gel image indicate substrate DNA band intensity. **FIG. 22B** shows sequence-specific cleavage of circular dsDNA by crRNA guided-Cas12a proteins including FnCas12a, WT mgCas12a-1 and WT mgCas12a-2. The substrate is indicated by an arrow and the letter S and linearized product (from cleavage) in the gel are also shown with arrows. Purified plasmid DNA was used for circular dsDNA. The numbers below gel image indicate substrate DNA band intensity.

EXAMPLE 14

Cas12a Cleavage of Target DNA Using Different 5' Handles

[0242] This example describes Cas12a cleavage of target DNA using different 5' handles. mgCas12a-1 and mgCas12a-2 were complexed with crRNA to target a specific nucleic acid. The crRNA for guiding mgCas12a-1 and mgCas12a-2 included the nucleotides at the 5' end prior to the step-loop region of the crRNA for AsCas12a, FnCas12a, and LbCas12a. Cleavage activity was monitored by running gels on the reaction of the Cas12a/crRNA complex with the target nucleic acid at various time points including 0 hr, 1 min, 10 min, 30 min, 1 h, 6 h, and 12 h.

[0243] **FIG. 23A-B** shows that the mgCas12a proteins of the present disclosure can utilize three different types of Cas12a handles. **FIG. 23A** shows cleavage of target linear dsDNA by WT mgCas12a-1 complexed with a crRNA having a 5' handle from AsCas12a, FnCas12a, and LbCas12a. **FIG. 23B** shows cleavage of target linear dsDNA by WT mgCas12a-2 complexed with a crRNA having a 5' handle from AsCas12a, FnCas12a, and LbCas12a. As seen in these gels, both mgCas12a proteins can utilize three types of 5' handles: AsCas12a, FnCas12a and LbCas12a, to sequence-specifically cleave dsDNA. The substrate ("S") indicates the target nucleic acid and the "P" indicates cleaved products. The numbers below gel image indicate substrate DNA band intensity. Controls included a catalytically inactive mgCas12a ("d_mgCas12a"), no Cas12a, or no crRNA. mgCas12a-1 and mgCas12a-2 exhibited cleavage of the target nucleic acid with all three crRNAs having the various 5' nucleotides found in crRNAs for AsCas12a, FnCas12a, and LbCas12a.

EXAMPLE 15

Random dsDNase Activity of Cas12a-RNPs

[0244] This example describes random dsDNase activity of Cas12a_RNPs. mgCas12a-1, d_mgCas12a_1 (deactivated), mgCas12a_2, d_mgCas12a_2 (deactivated), AsCas12a, FnCas12a, and LbCas12a were complexed with crRNAs to target a linear dsDNA in HsCCR5, HsDNMT1, and HsEMX1. Cleavage was monitored over time.

[0245] As shown in **FIG. 24**, and **FIG. 25A**, seven different Cas12a-RNPs were incubated with target dsDNA for 12 or 24 hours. The target substrate dsDNA is indicated with an "S" for substrate and cleaved products are indicated with a "P" for products. The dsDNA substrate and resulting cleaved products were almost entirely degraded, which may be due to random DNase activity of some Cas12a orthologs, including FnCas12a and LbCas12a, after incubation of the reaction for 12 hours and 24 hours. Both substrate and cleaved products were detected in the 12 hour reaction with the metagenomically mined Cas12a proteins of the present disclosure, including WT mg-1 (SEQ ID NO: 1) and WT mg-2 (SEQ ID NO: 2). **FIG. 25B** shows a graph of time versus dsDNase activity

of each Cas12a, demonstrating that target substrate dsDNA remains at later time points for mgCas12a-1 (SEQ ID NO: 1), indicating that mgCas12a-1 exhibits lower random DNase activity. [0246] **FIG. 25A-B** show that Cas12a exhibits random dsDNase activity of target linear dsDNA (human DNMT1). **FIG. 25A** shows FnCas12a, WT mgCas12a-1, and mgCas12a-2, complexed with crRNA and incubated with linear dsDNA for different time periods. The numbers below gel image indicate substrate DNA band intensity. The band corresponding to the substrate linear dsDNA essentially disappeared over time for the FnCas12a, and was very faint for mgCas12a-2. Substrate bands were observed at the same later time points for mgCas12a-1. **FIG. 25B** shows a graph of time versus dsDNase activity of each Cas12a, demonstrating that the substrate target dsDNA remains at later time points for mgCas12a-1.

EXAMPLE 16

Csa12a Activity in the Presence of Divalent Cations

[0247] This example describes Cas12a activity in the presence of divalent cations. Cas12a cleavage activity of linear dsDNA was tested for FnCas12a, WT mgCas12a-1, and WT mgCas12a-2. Cleavage of the target linear dsDNA was evaluated by running a gel for each of the divalent cations tested including CaCl₂, CoCl₂, CuSO₄, FeCl₂, MnSO₄, NiSO₄, and ZnSO₄. S indicates bands corresponding to the target linear dsDNA target substrate and P corresponds to cleavage products of the reactions.

[0248] **FIG. 26A-D** shows the activity of each Cas12a-RNP in the presence of different divalent cation. **FIG. 26A** shows the results from Cas12a-RNP cleavage of target, linear dsDNA in the presence of seven different divalent cations were given to each Cas12a-RNP. Abbreviations in the figure include DW for distilled water and Ctrl for 1X NEBuffer 1.1 (control). **FIG. 26B** shows sequence-specific dsDNA cleavage of FnCas12a-RNP under presence of different divalent cations. **FIG. 26C** shows sequence-specific dsDNA cleavage of each WT mgCas12a-1RNP under presence of different divalent cations. **FIG. 26D** shows sequence-specific dsDNA cleavage of each WT mgCas12a-2-RNP under presence of different divalent cations.

[0249] Cleaved products from targeting linear dsDNA was observed for all Cas12a orthologs tested. Linearized products from cleavage of circular dsDNA were clearly observed for mgCas12a-1 and mgCas12a-2. A light band was also observed for linearized products from cleavage of circular dsDNA by FnCas12a. However, the absence of a band at the circular dsDNA substrate (S) in all Cas12a orthologs tested indicates cleavage of the Cas12a by all orthologs, including FnCas12a.

[0250] While preferred embodiments of the present invention have been shown and described herein, it will be apparent to those skilled in the art that such embodiments are provided by way of

example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

CLAIMS

WHAT IS CLAIMED IS:

1. A composition comprising:
 - at least one of i) - iv):
 - i) a polypeptide having at least 80% sequence identity with residue 829 through residue 991 of SEQ ID NO: 1, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 925 of SEQ ID NO: 1;
 - ii) a polypeptide having at least 80% sequence identity with residue 825 through residue 996 of SEQ ID NO: 3, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 930 of SEQ ID NO: 3;
 - iii) a polypeptide having at least 80% sequence identity with SEQ ID NO: 1, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 925 of SEQ ID NO: 1; or
 - iv) a polypeptide having at least 80% sequence identity with SEQ ID NO: 3, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 930 of SEQ ID NO: 3;

and

 - a guide RNA coupled to the at least one of i) – iv), wherein the guide RNA comprises a sequence that is reverse complementary to a eukaryotic nucleic acid sequence comprising from 6 to 60 bases.
2. The composition of claim 1, wherein the eukaryotic nucleic acid sequence is a human nucleic acid sequence.
3. The composition of claim 2, wherein the human nucleic acid sequence comprises a region in KRAS, HER2/neu, PD-1, TCR, p53, CCR5, DNMT1, EMX1, and LKB1.
4. The composition of claim 1, wherein the eukaryotic nucleic acid sequence is a plant nucleic acid sequence.
5. The composition of claim 1, wherein the polypeptide is a type V CRISPR-associated protein.
6. The composition of claim 5, wherein the type V CRISPR-associated protein is a Cas12a protein.
7. The composition of claim 1, wherein the polypeptide further comprises a purification tag.

8. The composition of claim 7, wherein the purification tag comprises at least one tag selected from the group consisting of a His tag, a FLAG tag, an AU1 epitope tag, an AU5 epitope tag, a bacteriophage T7 tag, a bacteriophage V5 epitope tag, a Bluetongue virus tag (B-tag), a Glu-Glu tag (EE-tag), an HSV epitope tag, a KT3 epitope tag, a Myc epitope tag, a PDZ ligand tag, a polyarginine tag, a polyaspartate tag, a polycysteine tag, a polyphenylalanine tag, a protein C tag, an S1-tag, an S-tag, a Step-tag, and a VSV-G tag.
9. The composition of claim 1, wherein the guide RNA comprises an A-rich protospacer adjacent motif (PAM) sequence, a G-rich PAM sequence, a T-rich PAM sequence, or a C-rich PAM sequence.
10. The composition of claim 9, wherein the PAM sequence comprises 3 T nucleobases, 2 T nucleobases, 1 T nucleobase, or TTTN.
11. The composition of claim 1, wherein the guide RNA comprises a crRNA and a tracrRNA.
12. The composition of claim 1, wherein the composition exhibits at least 2-fold increased genome editing efficiency than AsCas12a, FnCas12a, or LbCas12a.
13. The composition of claim 1, further comprising an excipient.
14. The composition of claim 13, wherein the excipient has a pH of from 7 to 8.
15. The composition of claim 13, wherein the excipient is a buffer.
16. The composition of claim 14, wherein the buffer comprises at least one of Bis-Tris Propane-HCl, MgCl₂, or bovine serum albumin.
17. A method of gene editing, wherein the method comprises:
 - providing a composition comprising:
 - at least one of i) - iv):
 - i) a polypeptide having at least 80% sequence identity with residue 829 through residue 991 of SEQ ID NO: 1, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 925 of SEQ ID NO: 1;
 - ii) a polypeptide having at least 80% sequence identity with residue 825 through residue 996 of SEQ ID NO: 3, wherein the polypeptide

comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 930 of SEQ ID NO: 3;

iii) a polypeptide having at least 80% sequence identity with SEQ ID NO: 1, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 925 of SEQ ID NO: 1; or

iv) a polypeptide having at least 80% sequence identity with SEQ ID NO: 3, wherein the polypeptide comprises an amino acid sequence comprising a Lysine aligned to a Lysine at position 930 of SEQ ID NO: 3;

and

a guide RNA coupled to the at least one of i) – iv), wherein the guide RNA

comprises a sequence that is reverse complementary to a eukaryotic nucleic acid sequence comprising from 6 to 60 bases;

contacting a cell with the composition;

binding the guide RNA to a eukaryotic nucleic acid sequence; and

cleaving the nucleic acid sequence.

18. The method of claim 17, wherein the eukaryotic nucleic acid sequence is a human nucleic acid sequence.

19. The method of claim 18, wherein the human nucleic acid sequence comprises a region in KRAS, HER2/neu, PD-1, TCR, p53, CCR5, DNMT1, EMX1, and LKB1.

20. The method of claim 17, wherein the eukaryotic nucleic acid sequence is a plant nucleic acid sequence.

21. The method of claim 17, wherein the polypeptide is a type V CRISPR-associated protein.

22. The method of claim 21, wherein the type V CRISPR-associated protein is a Cas12a protein.

23. The method of claim 17, wherein the polypeptide comprises a purification tag.

24. The method of claim 23, wherein the purification tag comprises at least one tag selected from the group consisting of: a His tag, a FLAG tag, an AU1 epitope tag, an AU5 epitope tag, a bacteriophage T7 tag, a bacteriophage V5 epitope tag, a Bluetongue virus tag (B-tag), a Glu-Glu tag (EE-tag), an HSV epitope tag, a KT3 epitope tag, a Myc epitope tag, a PDZ ligand tag,

a polyarginine tag, a polyaspartate tag, a polycysteine tag, a polyphenylalanine tag, a protein C tag, an S1-tag, an S-tag, a Step-tag, and a VSV-G tag.

25. The method of claim 17, further comprising an excipient.
26. The method of claim 25, wherein the excipient has a pH of from 7 to 8.
27. The method of claim 25, wherein the excipient is a buffer.
28. The method of claim 27, wherein the buffer comprises at least one of Bis-Tris Propane-HCl, $MgCl_2$, or bovine serum albumin.
29. The method of claim 17, wherein the guide RNA comprises an A-rich protospacer adjacent motif (PAM) sequence, a G-rich PAM sequence, a T-rich PAM sequence, or a C-rich PAM sequence.
30. The method of claim 18, wherein the PAM sequence comprises 3 T nucleobases, 2 T nucleobases, 1 T nucleobase, or TTTN.
31. The method of claim 17, wherein the guide RNA comprises a crRNA and a tracrRNA.
32. The method of claim 17, wherein an efficiency of the step of cleaving is at least 2-fold higher than genome editing efficiency of AsCas12a, FnCas12a, or LbCas12a.
33. A method of improving cleaving efficiency of a type V CRISPR-associated protein, the method comprising:
 - providing the type V CRISPR-associated protein;
 - identifying a residue of the type V CRISPR-associated protein that is aligned with at Lysine at position 925 of SEQ ID NO: 1 or at Lysine at position 930 of SEQ ID NO: 3; and
 - mutating the residue to Lysine.

1/29

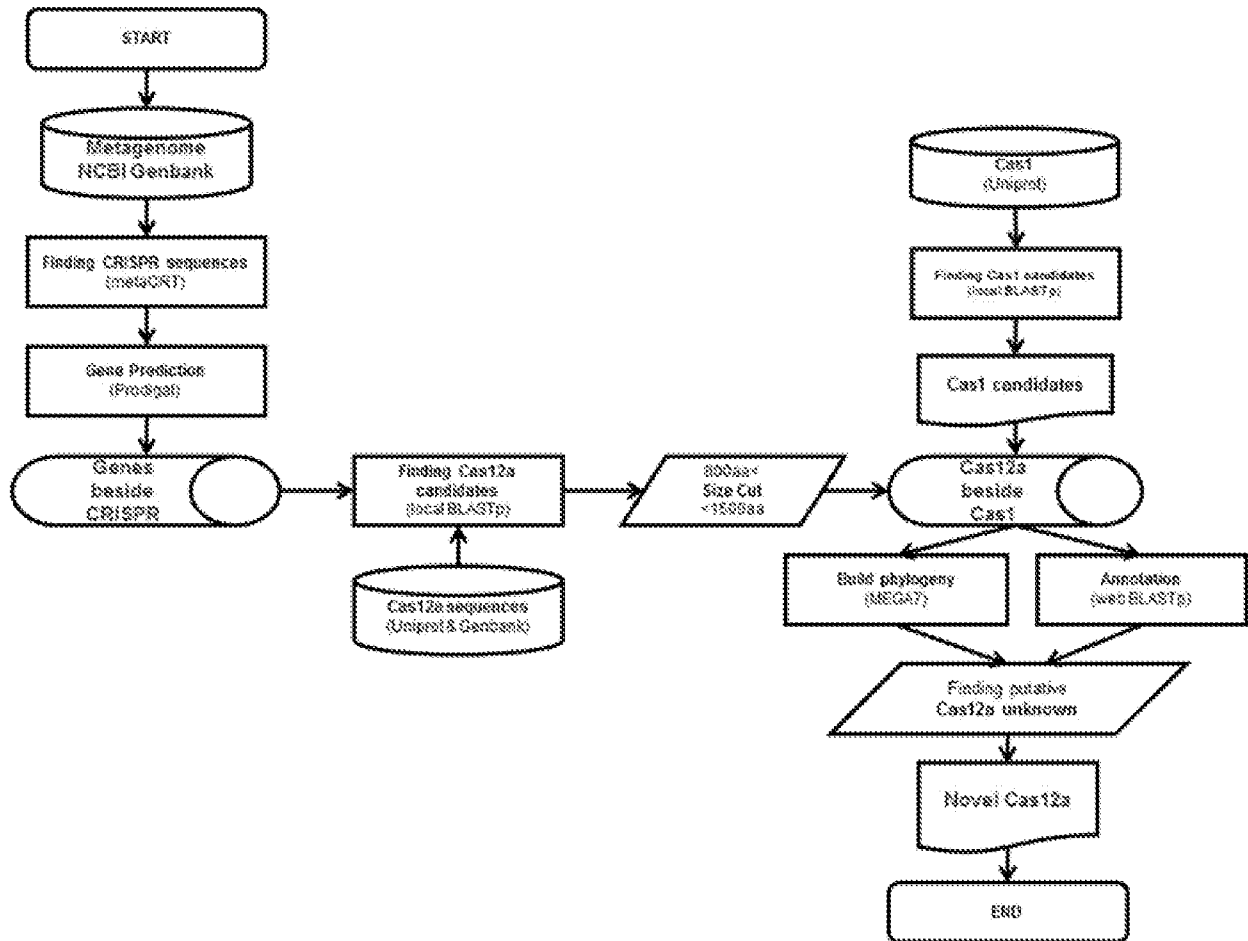


FIG. 1

2/29

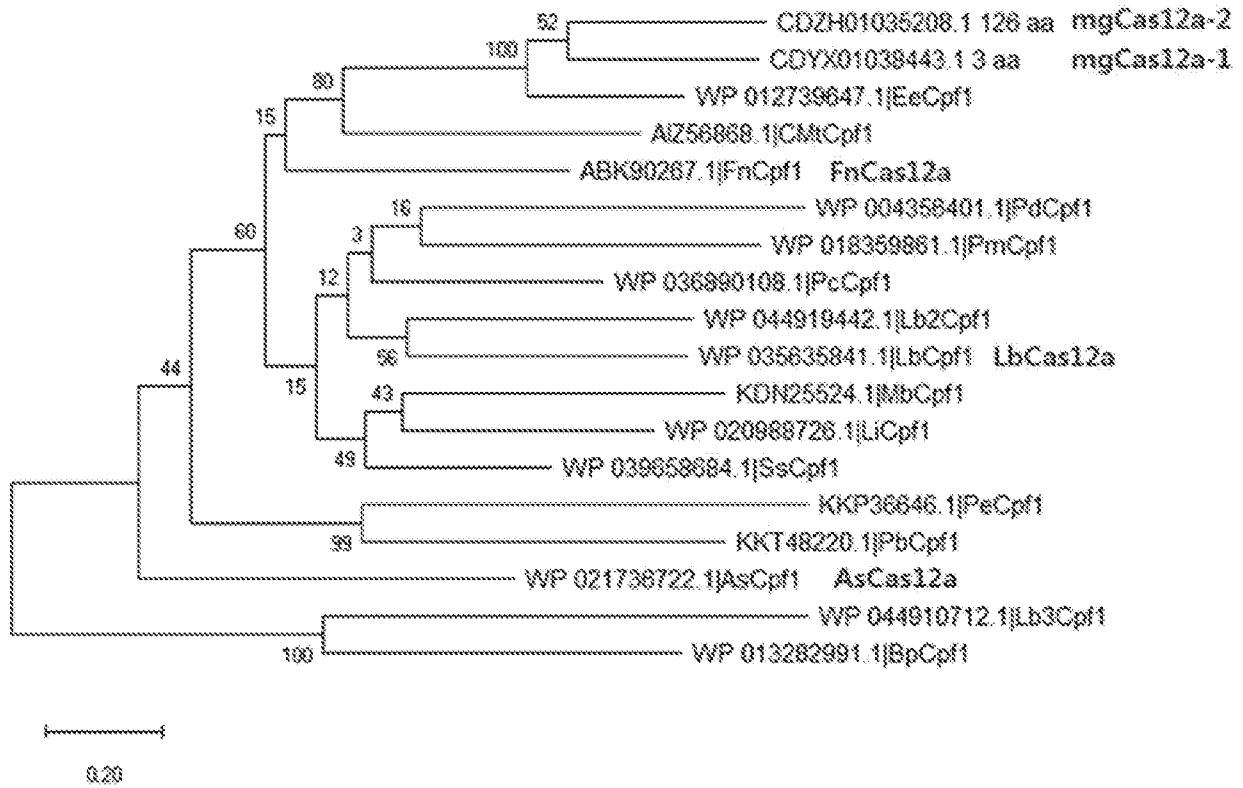


FIG. 2

3/29

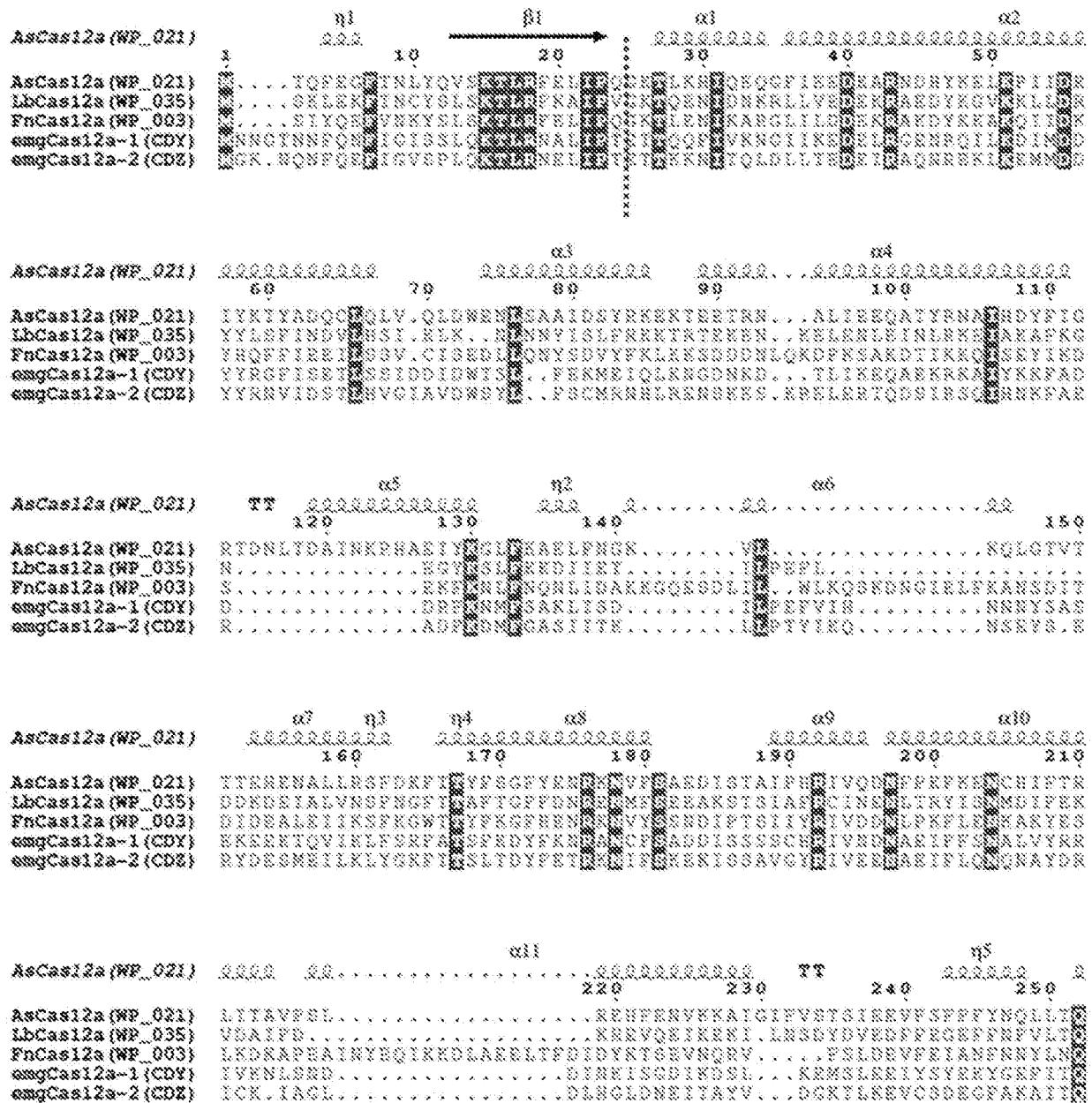


FIG. 3

4/29

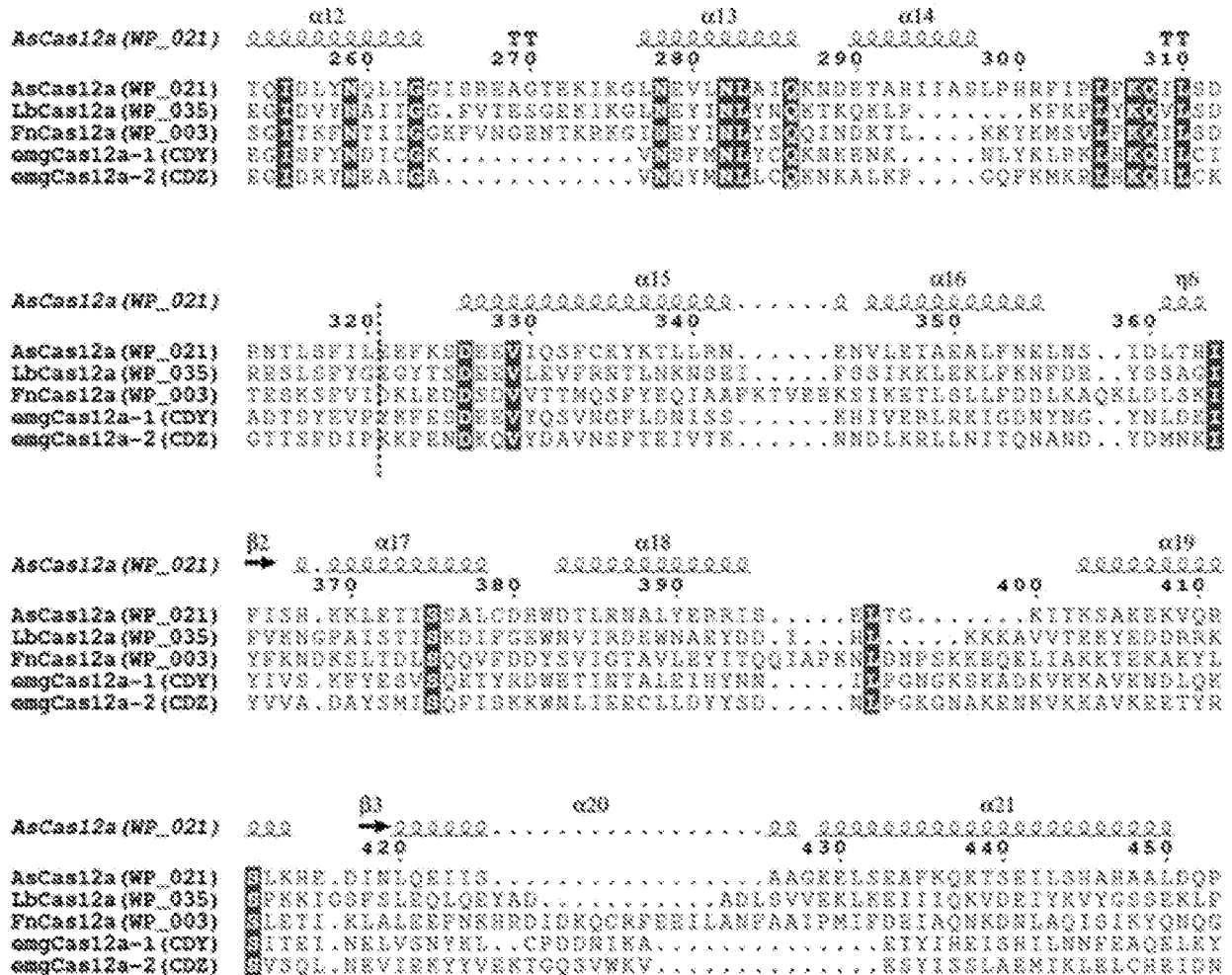


FIG. 4

5/29

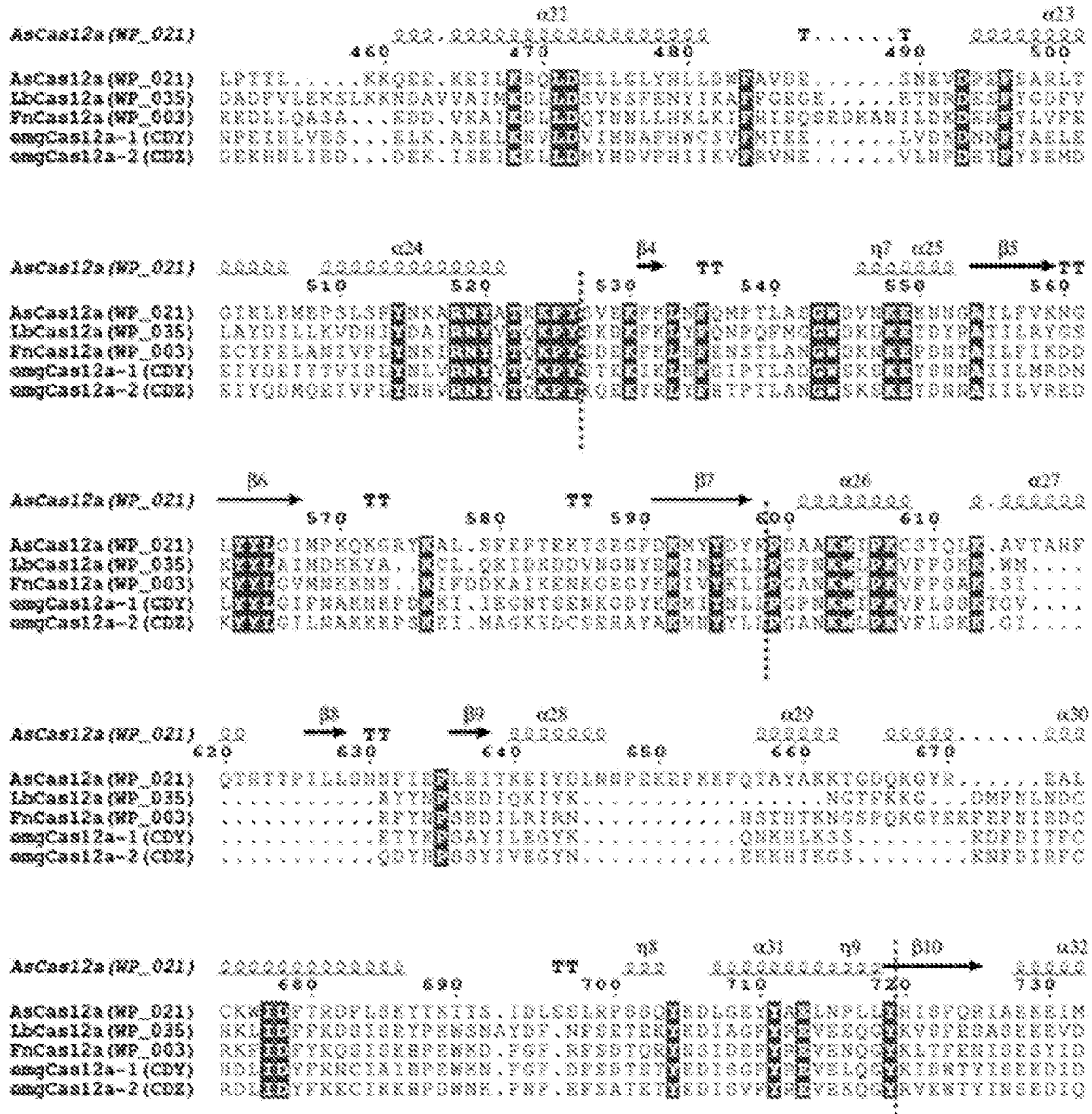


FIG. 5

6/29

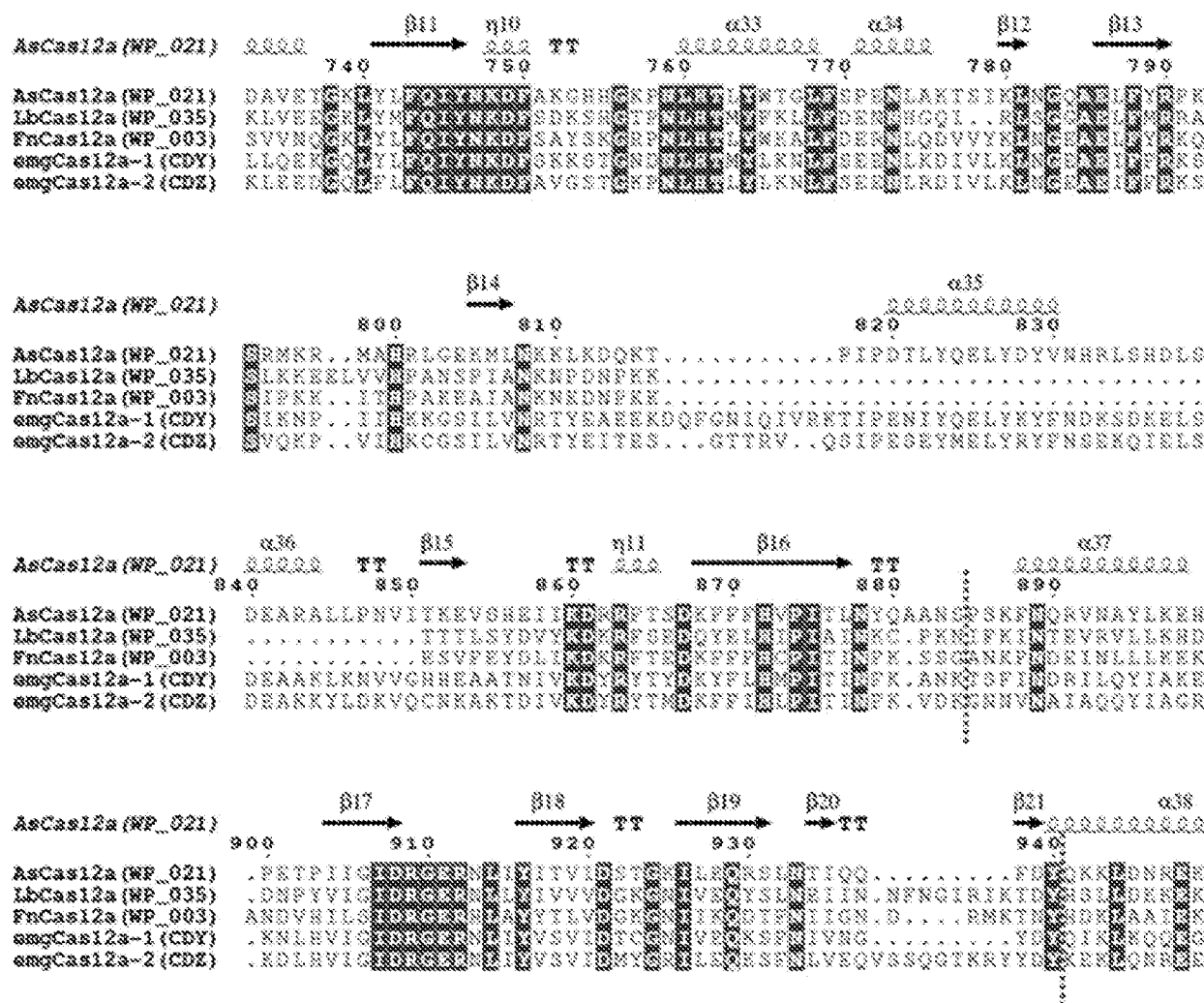


FIG. 6

7/29

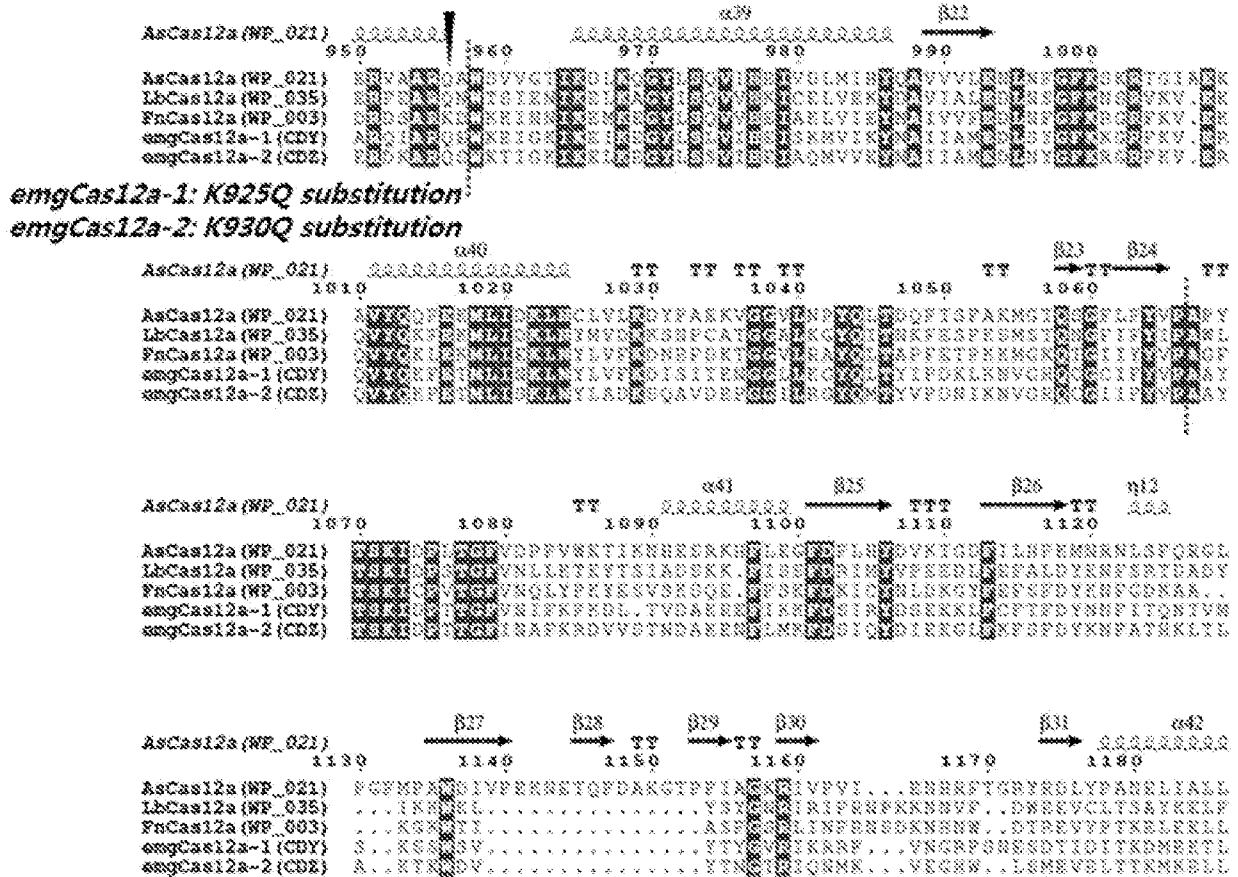


FIG. 7

8/29

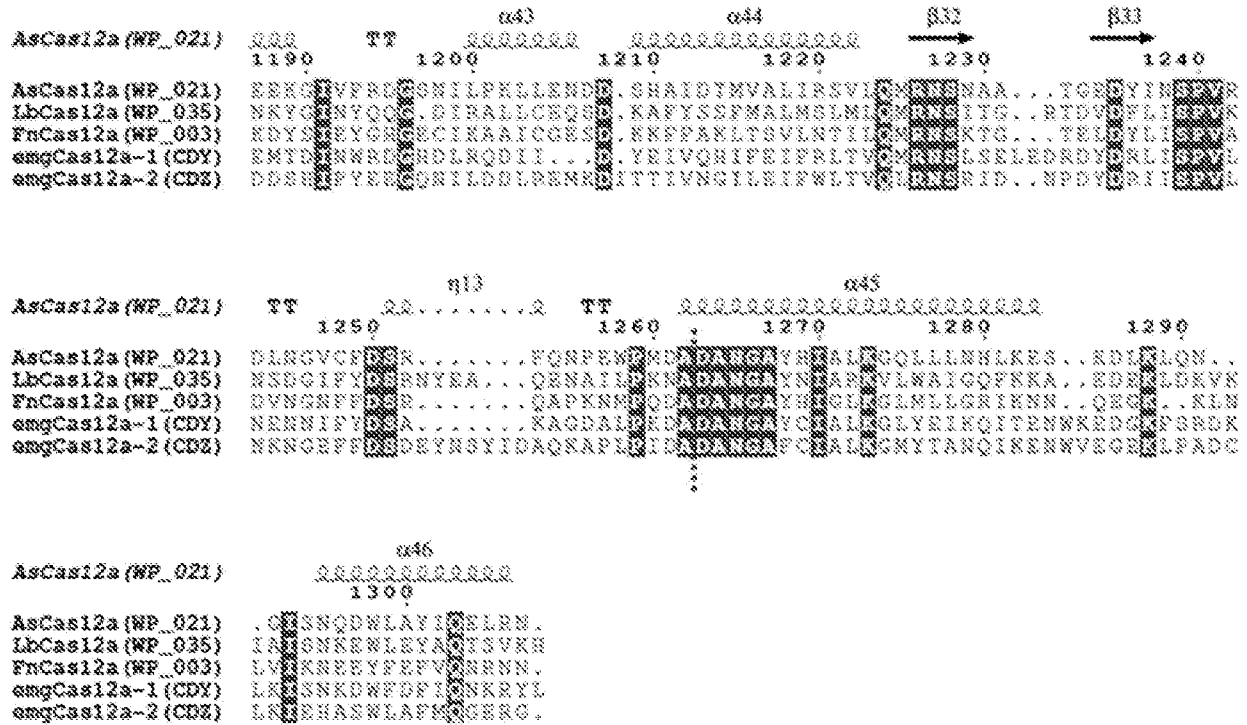


FIG. 8

9/29

	Old			New	
	AsCas12a (standard)	LbCas12a	FnCas12a	mgCas12a-1	mgCas12a-2
Reference	2015, F.Zhang lab	2015, F.Zhang lab	2015, F.Zhang lab	2018, S.H.Choe lab	2018, S.H.Choe lab
Species	<i>Acidimicrococcus</i> sp.	<i>Lochnerotrypanos</i> sp.	<i>Pantodon hudsonii</i> subsp. novus	metagenome CD19011018443.1	metagenome CD19011018443.1
Nucleotide length	3,921	3,884	3,892	3,789	3,825
Amino acid length	1,307	1,295	1,301	1,263	1,275
PAM sequence	5'-TTTN-3'	5'-TTTN-3'	5'-TTTN-3'	5'-TTTN-3' (speculative)	5'-TTTN-3' (speculative)
Sequence identity	100%	95.81%	94.85%	92.65%	91.29%
Critical residue identity (Yamamoto et al. Cell 2016)	100% (33)	79% (28)	79% (28)	88% (79)	88% (28)

FIG. 9A

Cas12a	As	Lb	Fn	mg-1	mg-2
As					
Lb	33.4				
Fn	34.2	40.4			
mg-1	32.4	35.3	36.6		
mg-2	30.7	34.9	35.9	52.2	

FIG. 9B

10/29

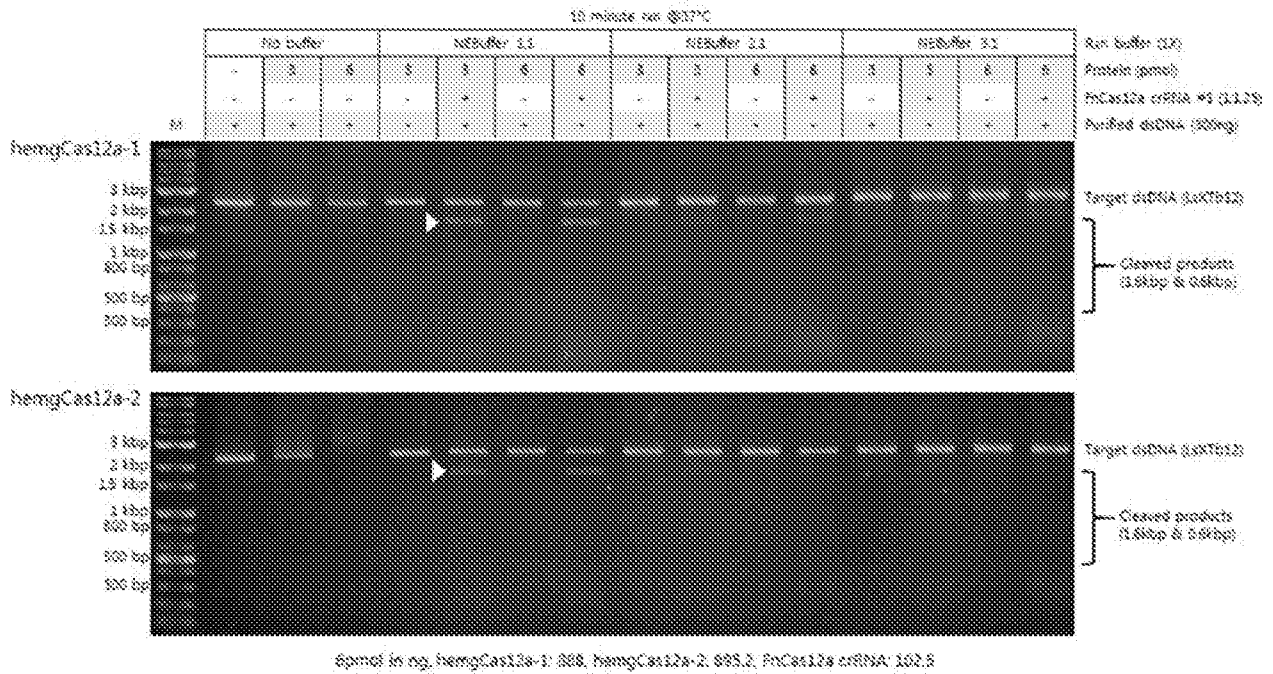


FIG. 10

11/29

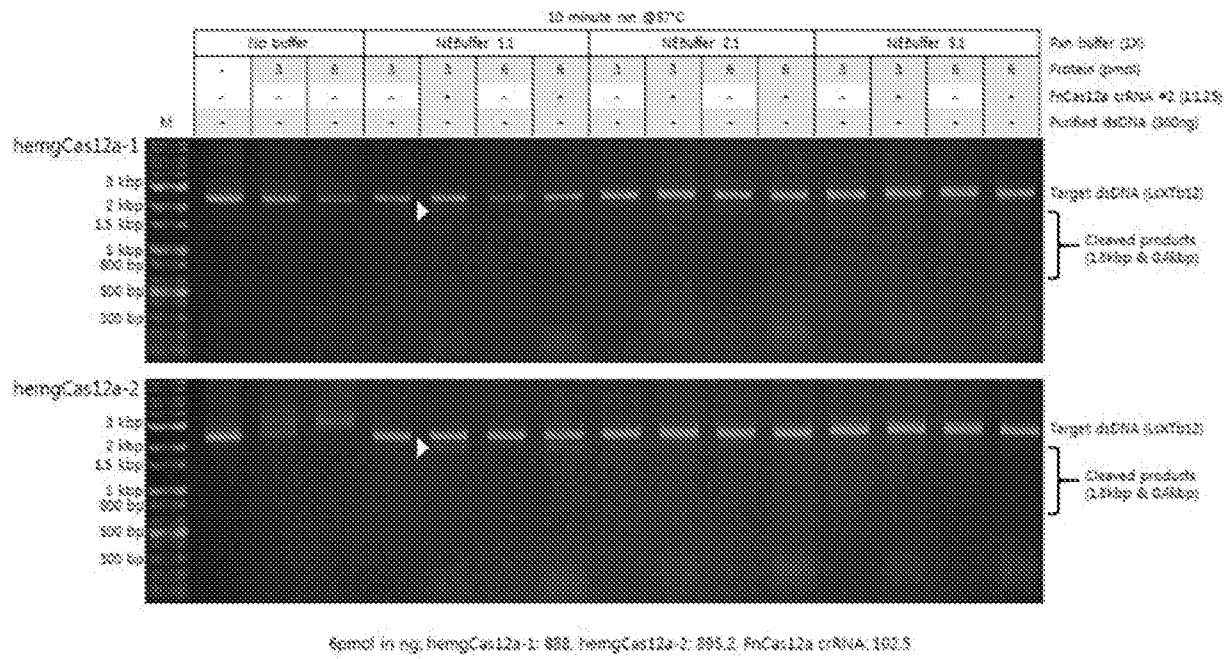


FIG. 11

12/29

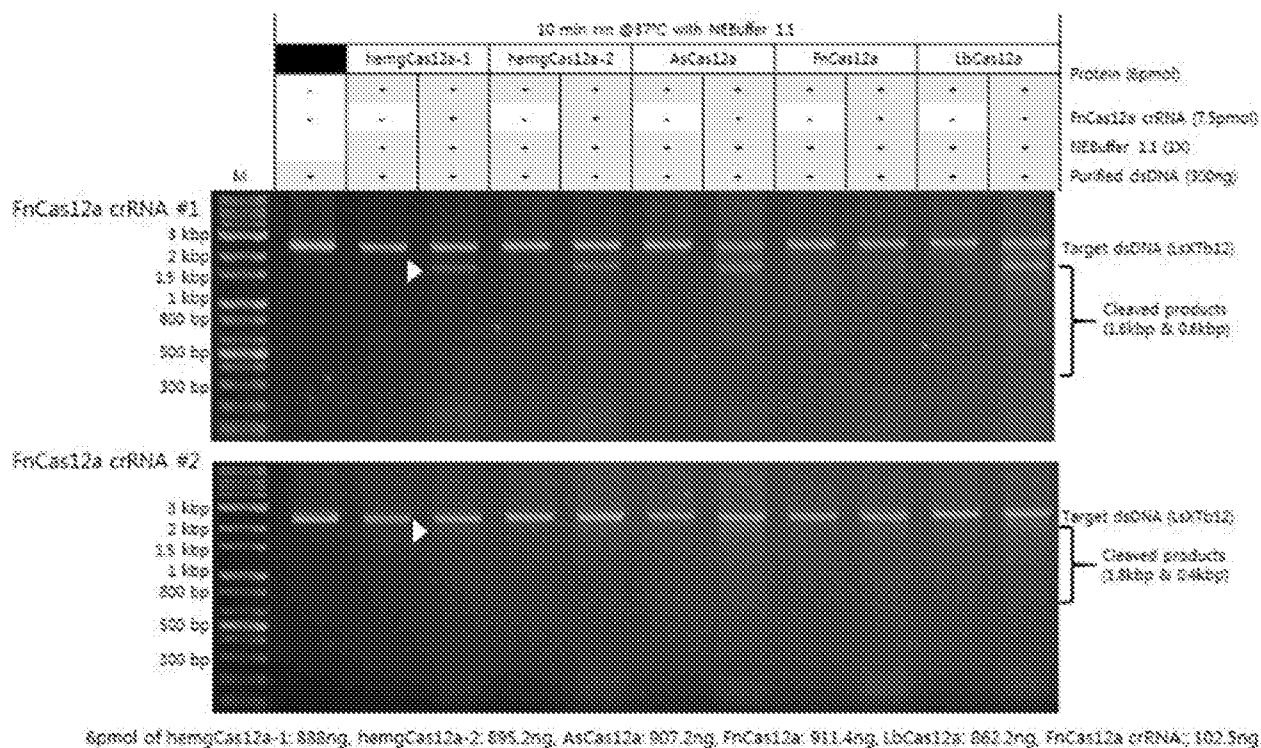


FIG. 12

13/29



FIG. 13

14/29

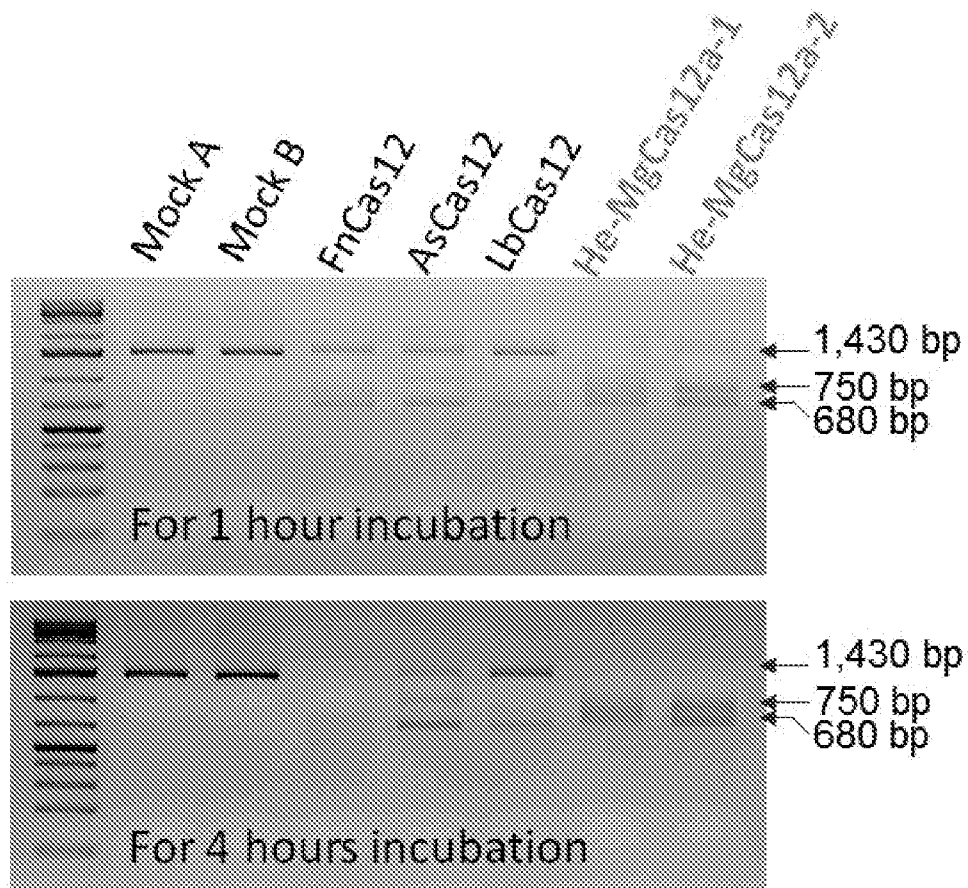


FIG. 14

15/29

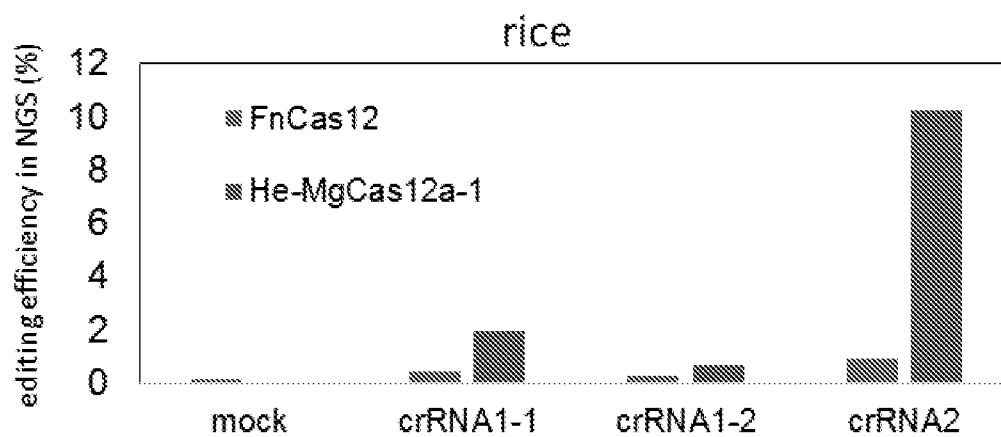


FIG. 15A

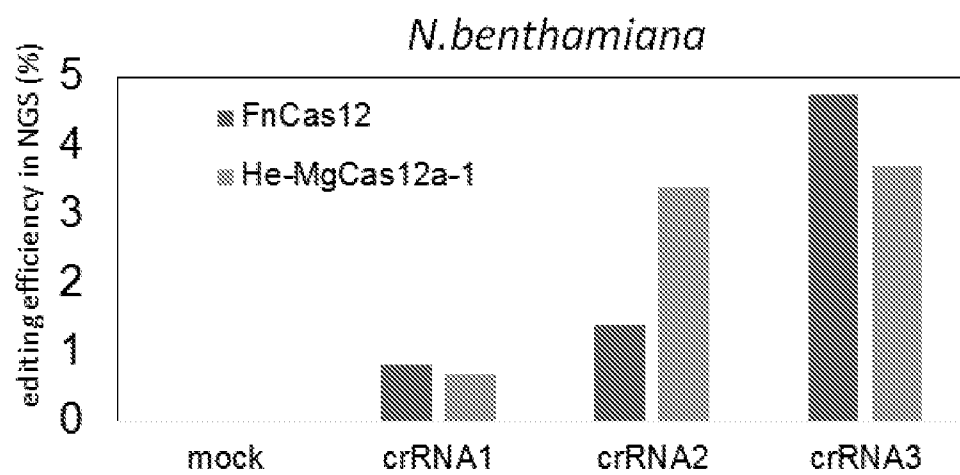


FIG. 15B

16/29

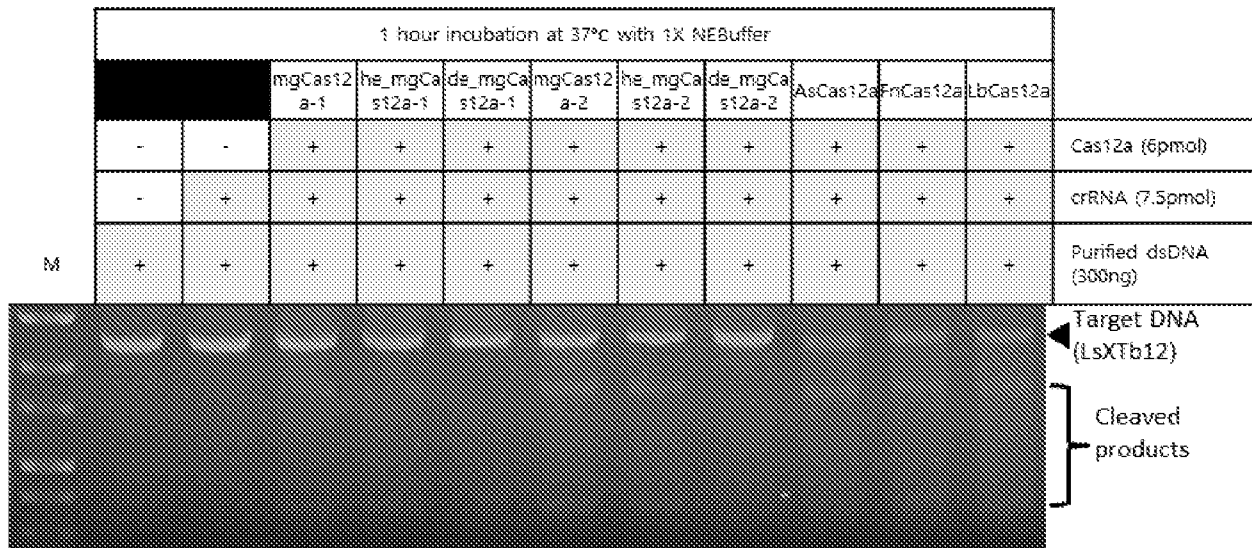


FIG. 16

17/29

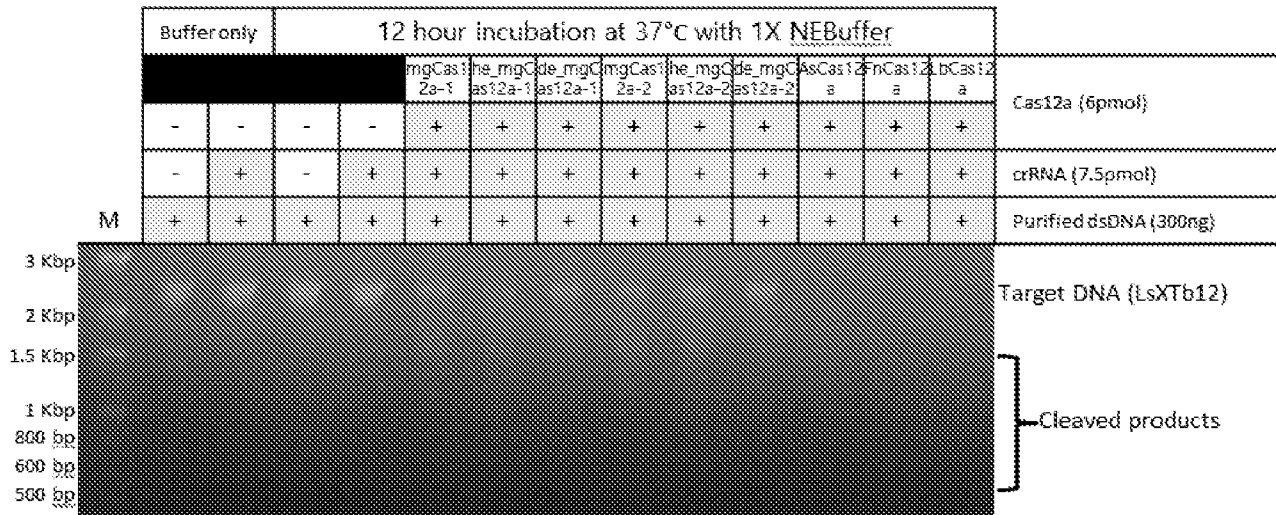


FIG. 17A

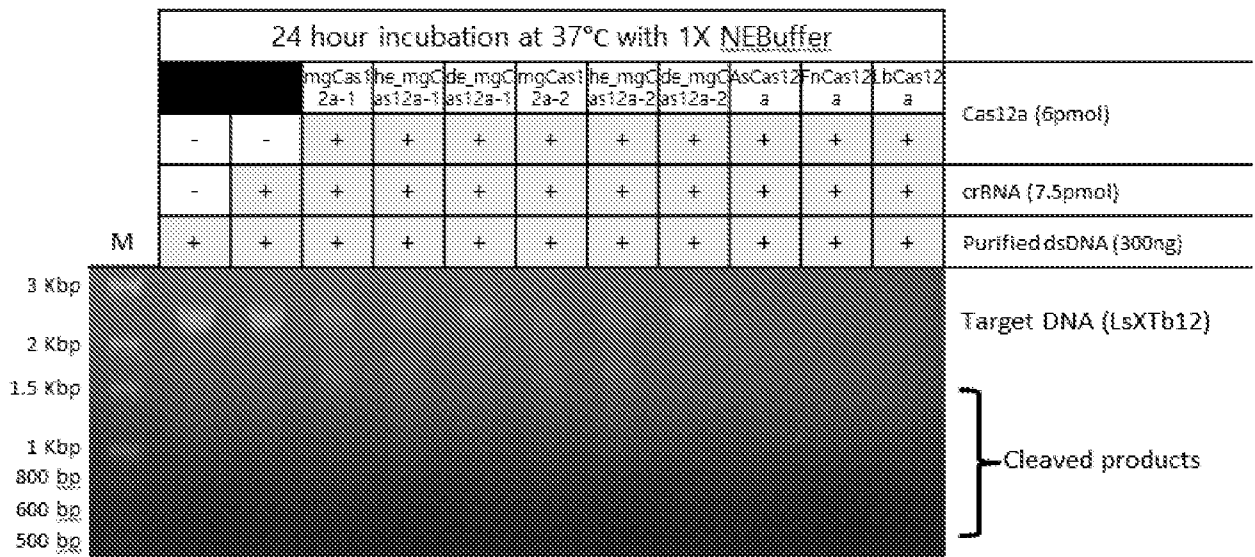


FIG. 17B

18/29

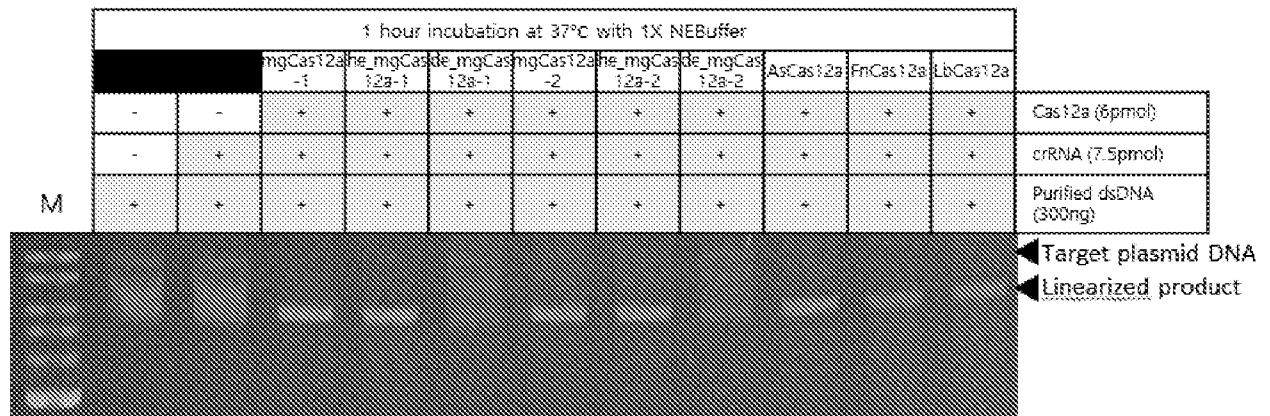


FIG. 18

19/29

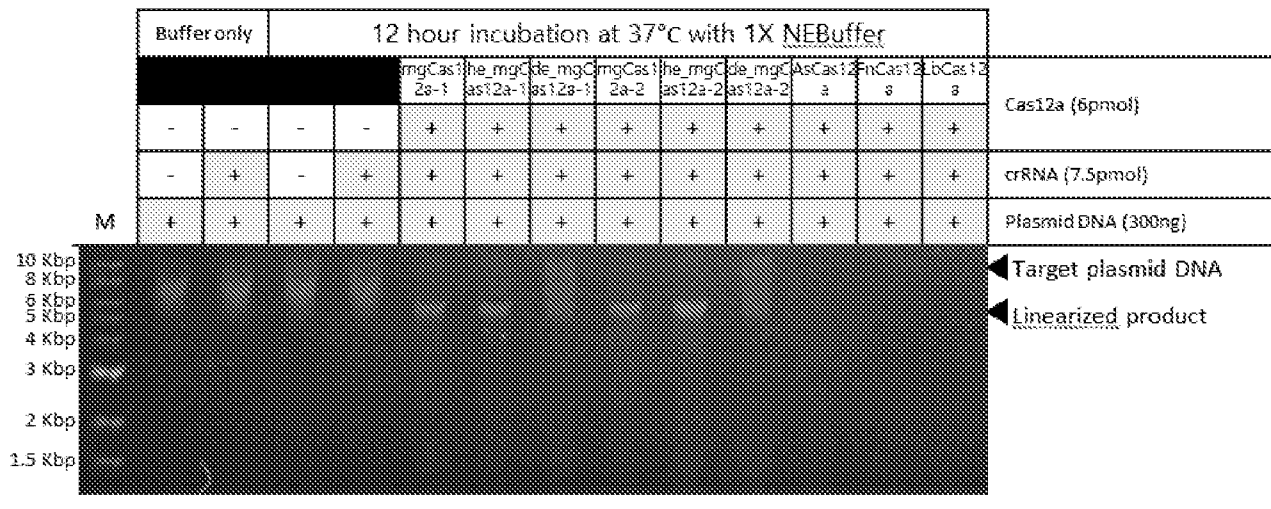


FIG. 19A

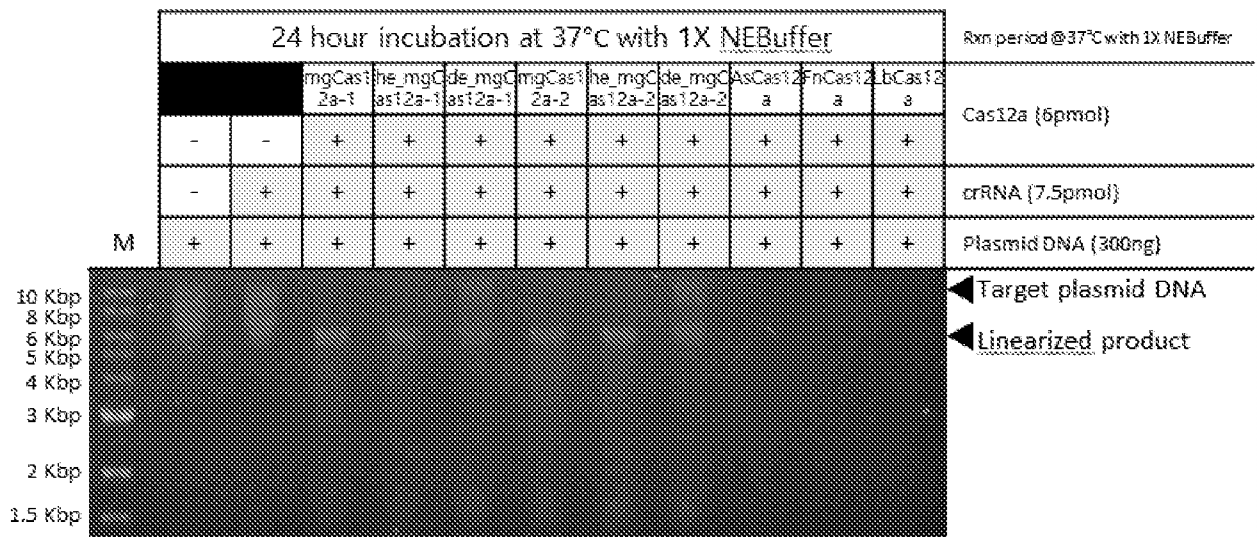


FIG. 19B

20/29

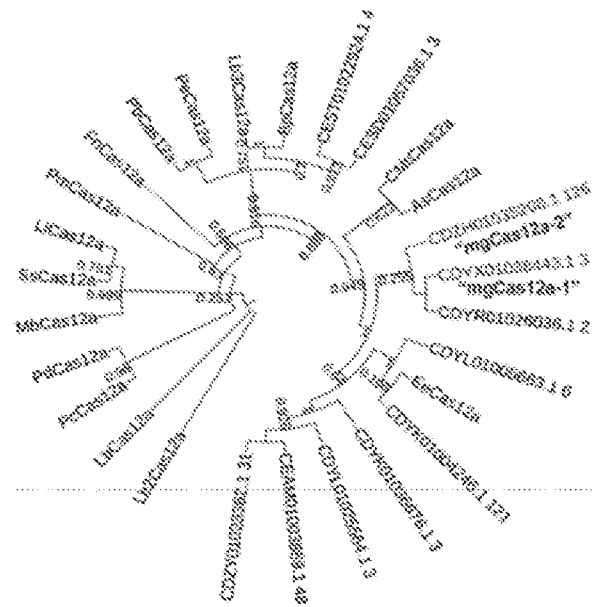
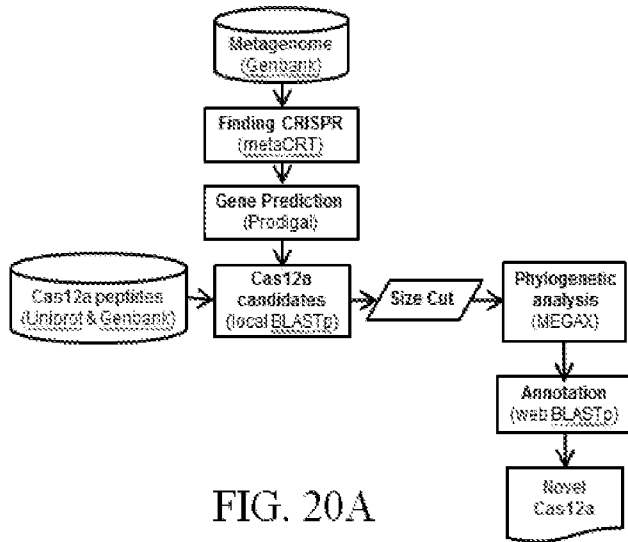
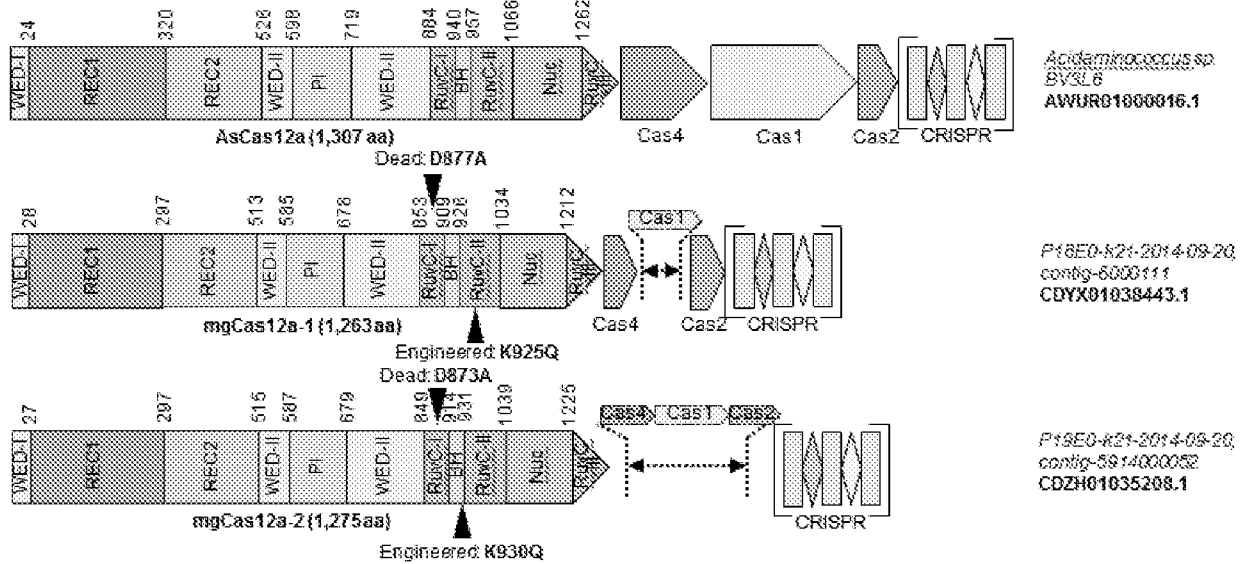


FIG. 20B



21/29

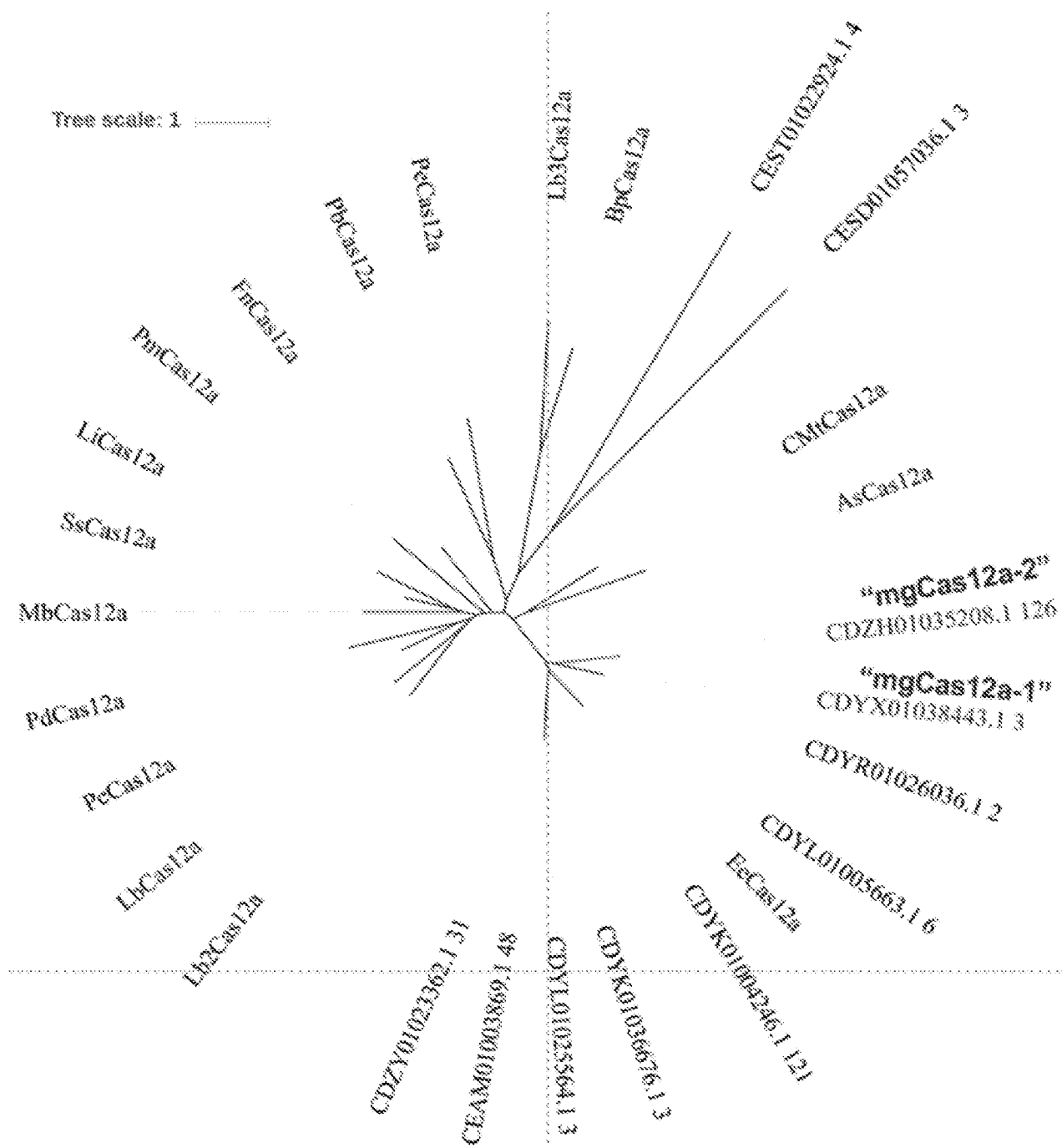


FIG. 21

22/29

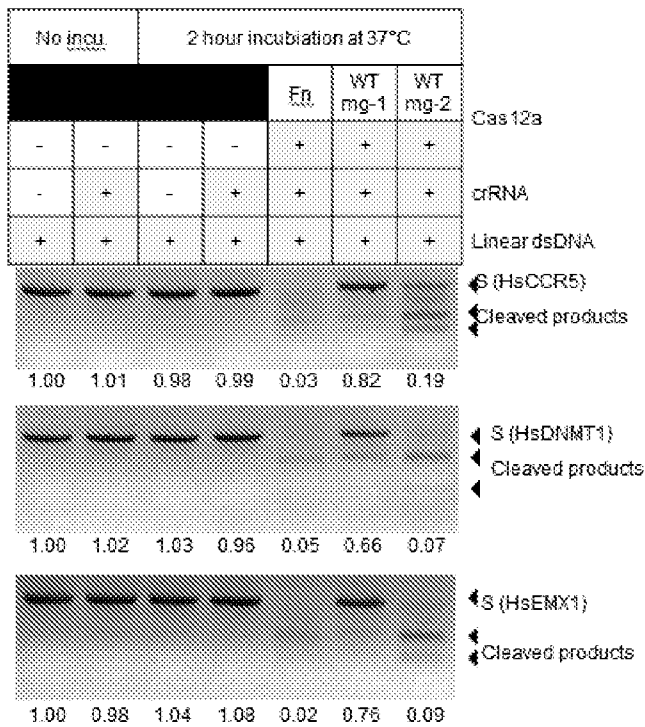


FIG. 22A

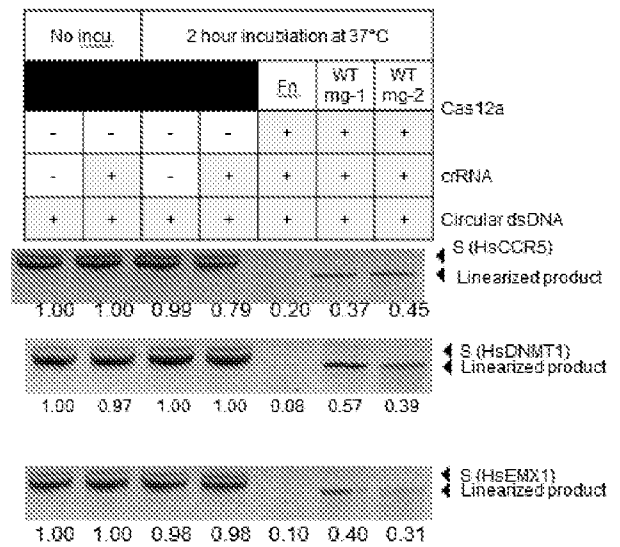


FIG. 22B

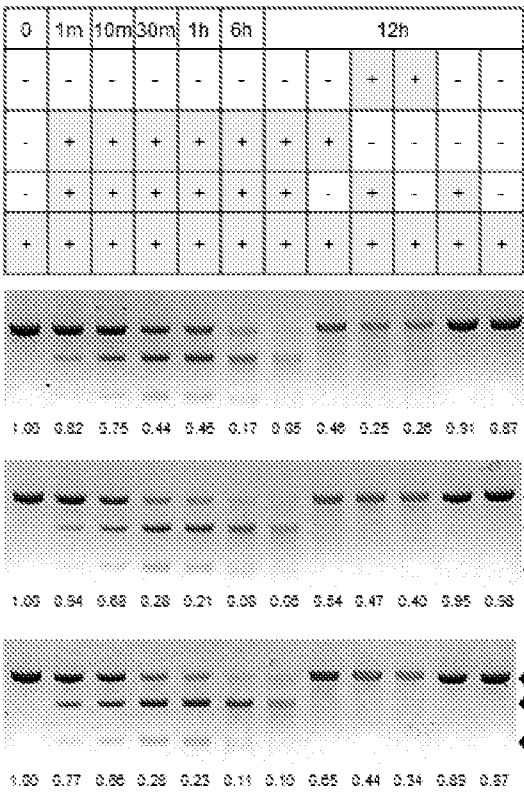


FIG. 23A

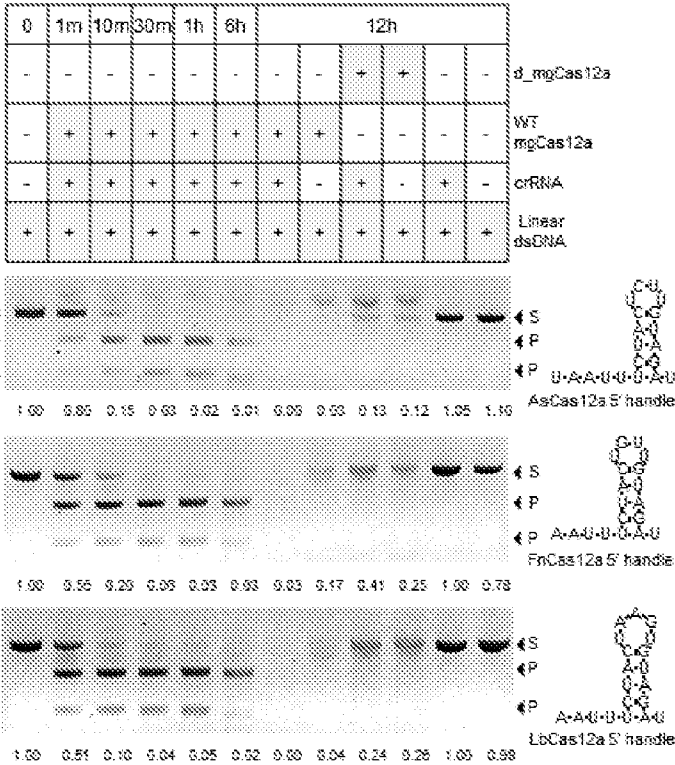


FIG. 23B

24/29

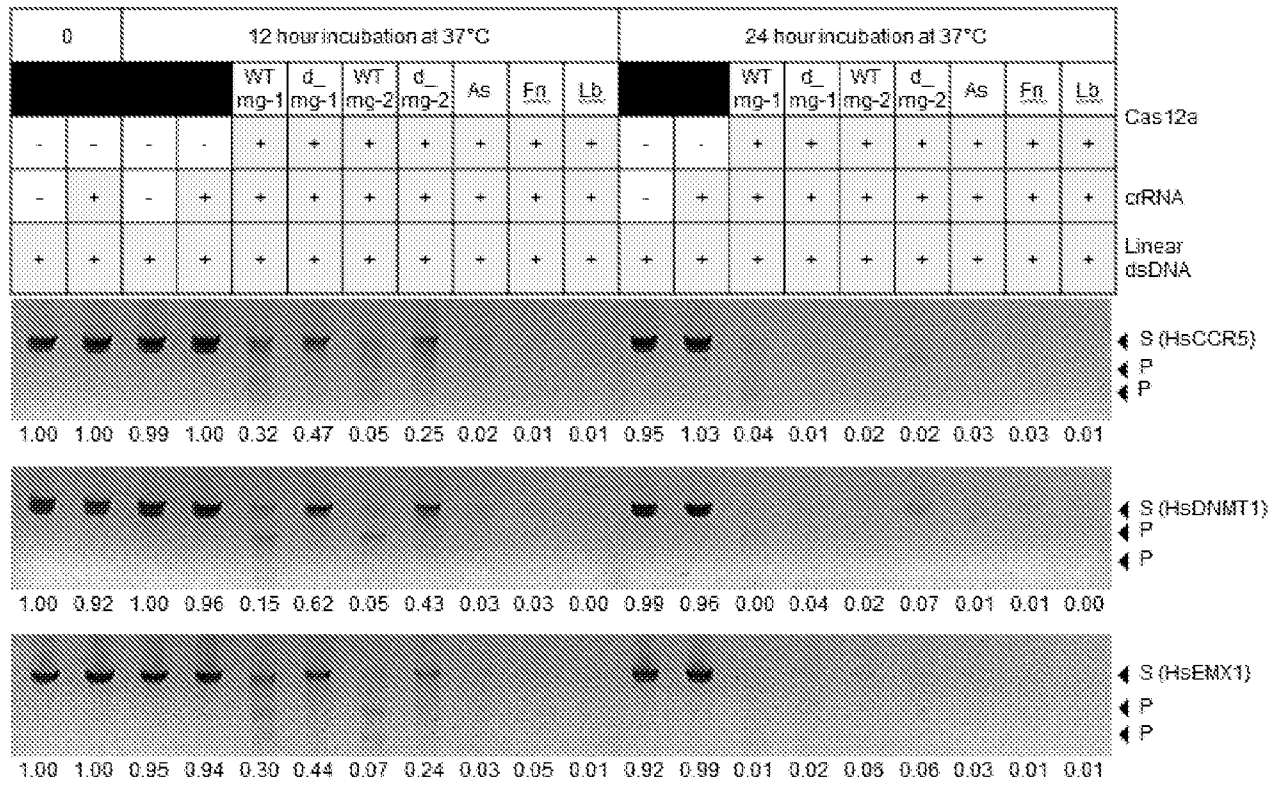


FIG. 24

25/29

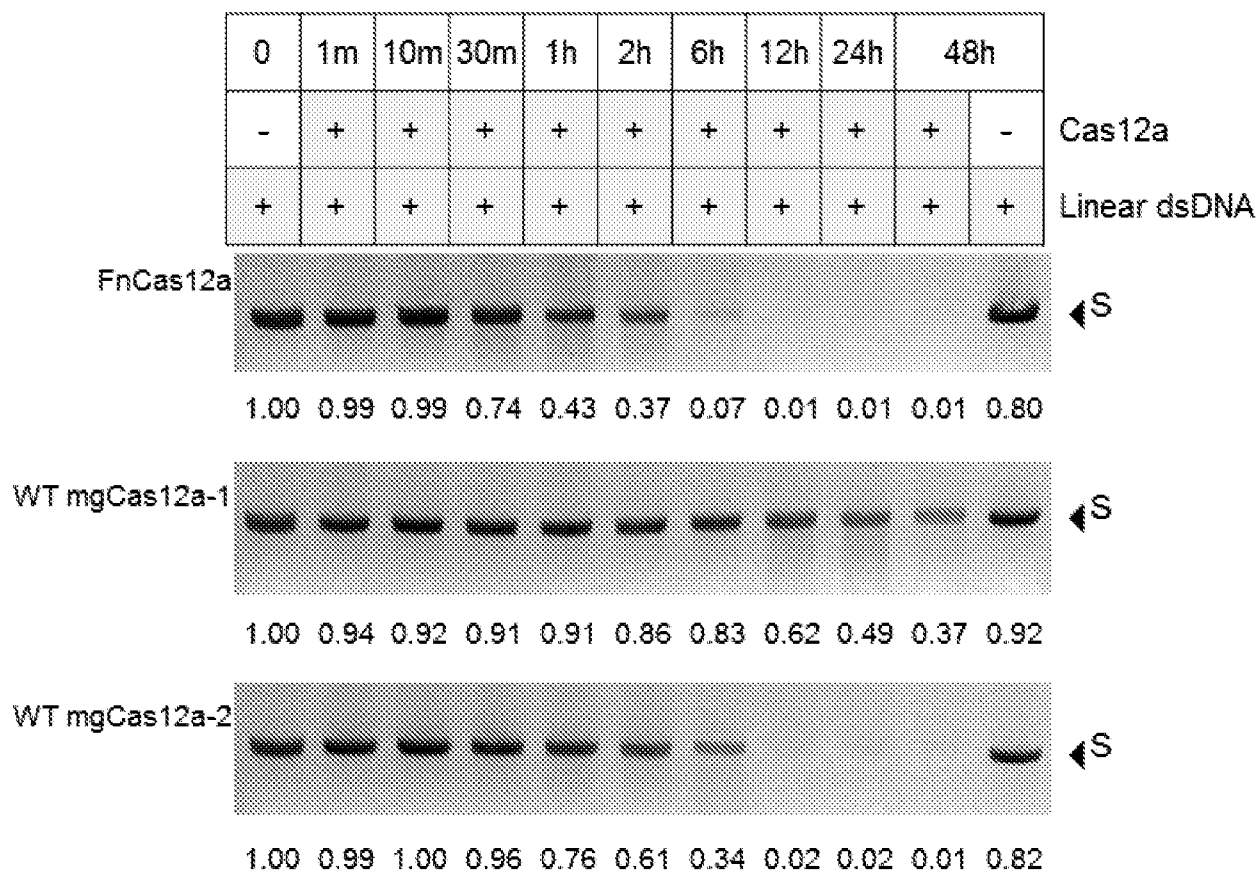


FIG. 25A

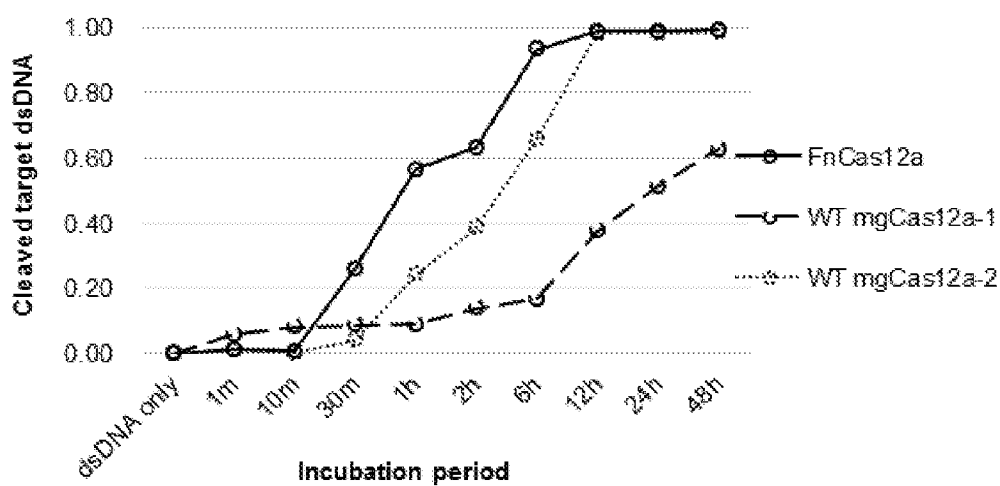


FIG. 25B

26/29

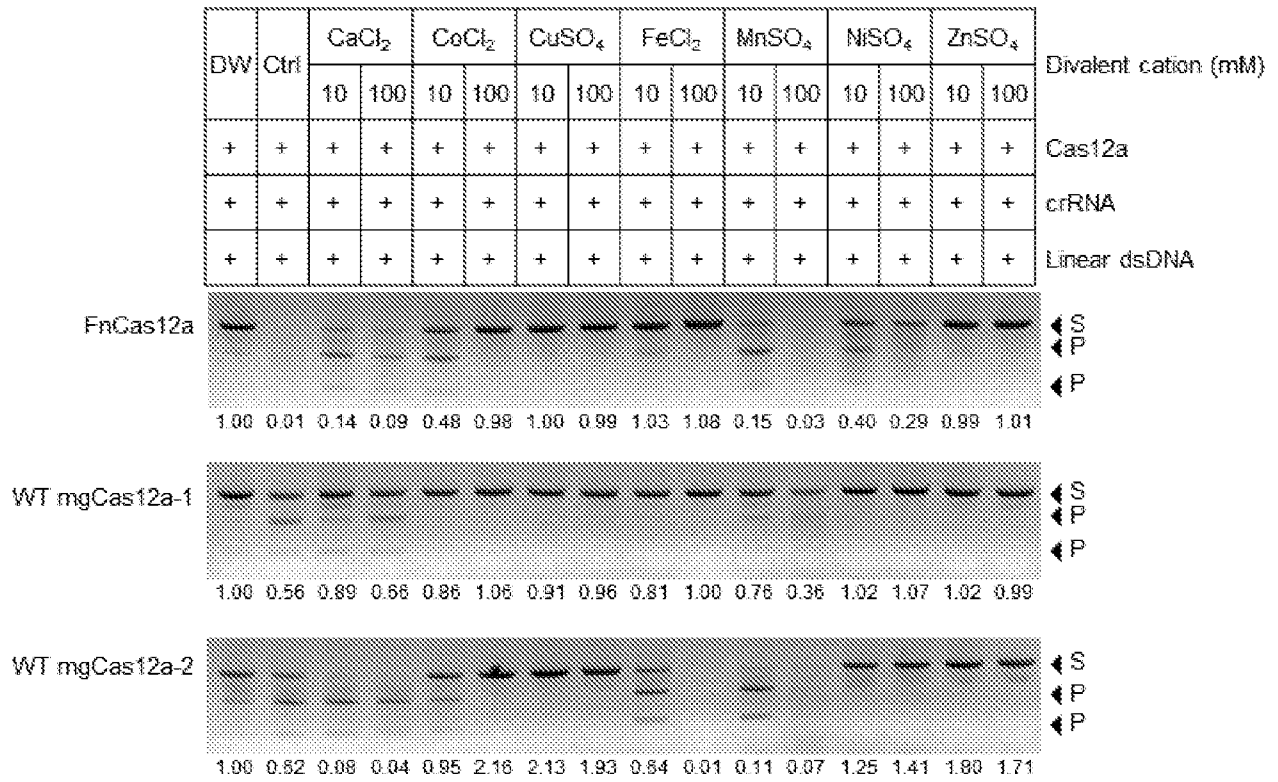
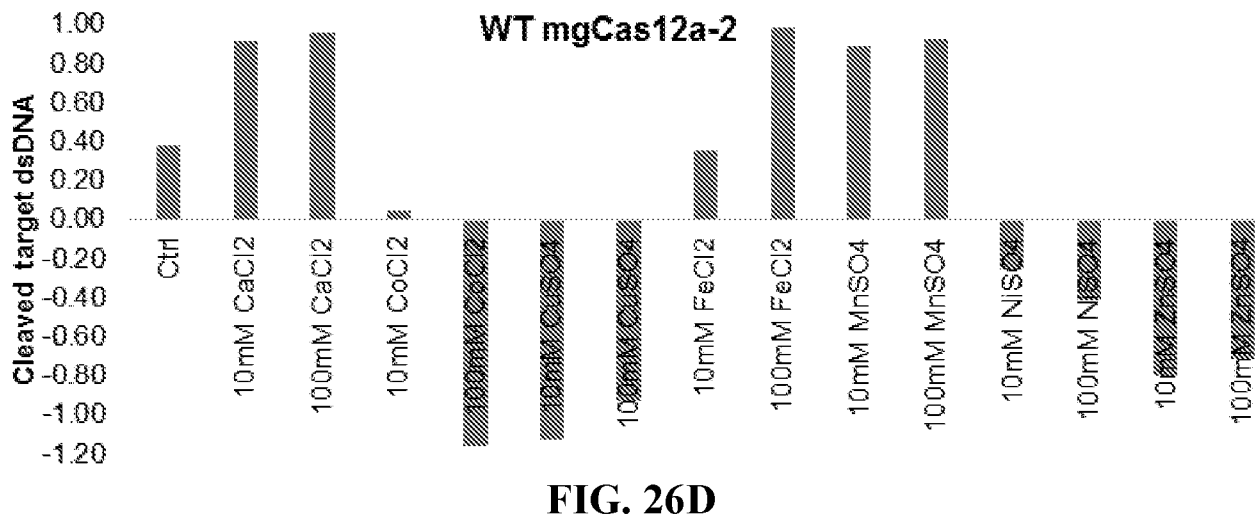
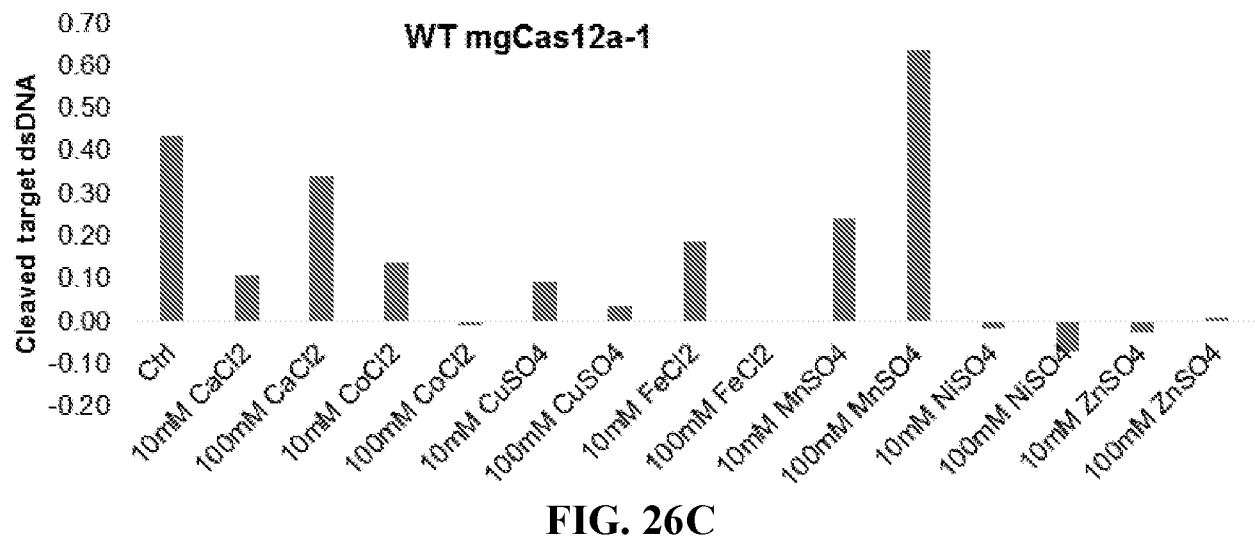
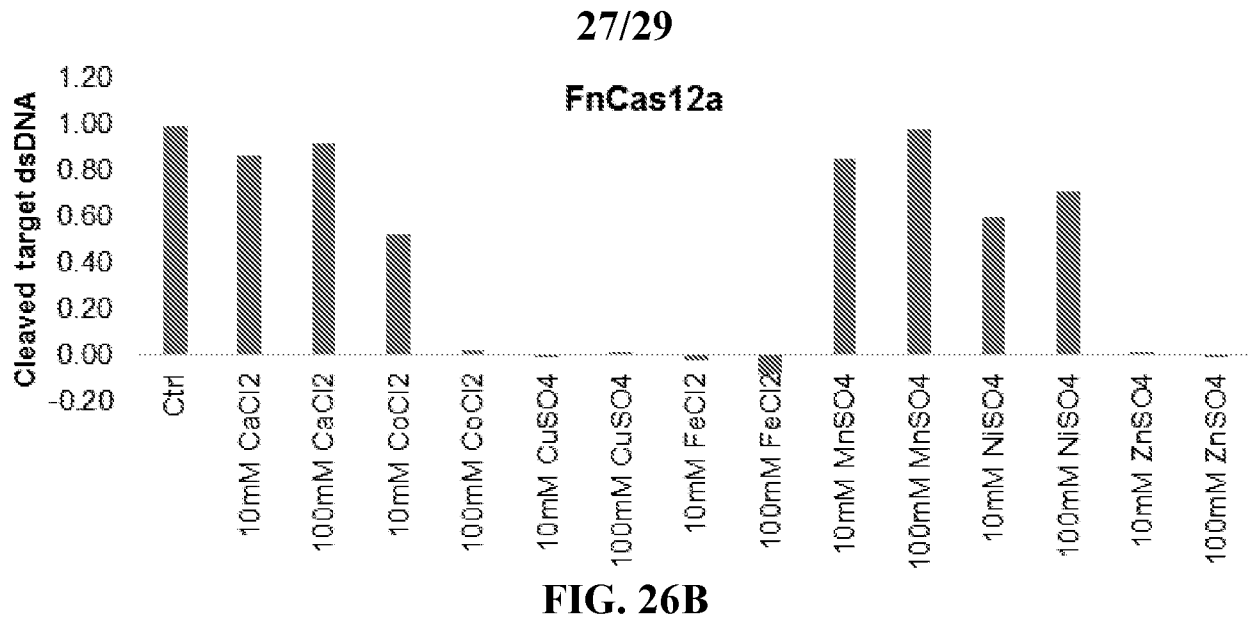


FIG. 26A



28/29

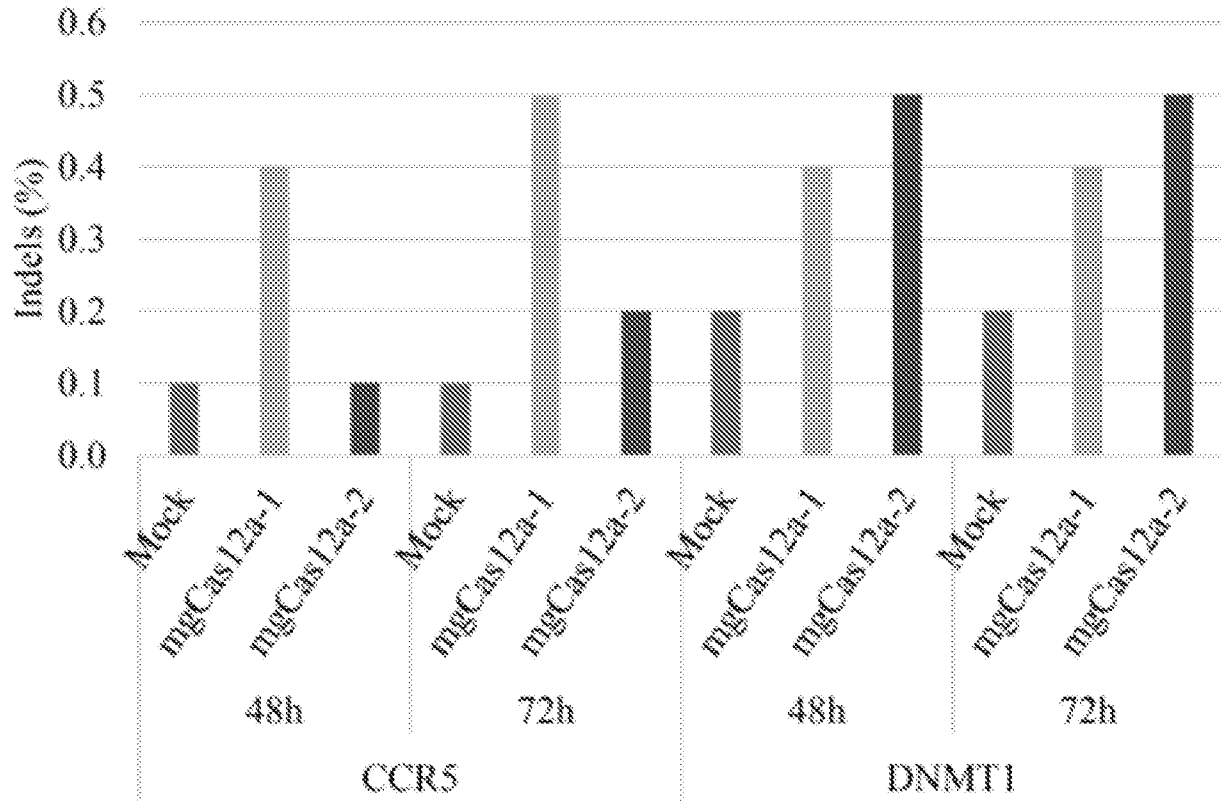


FIG. 27

29/29

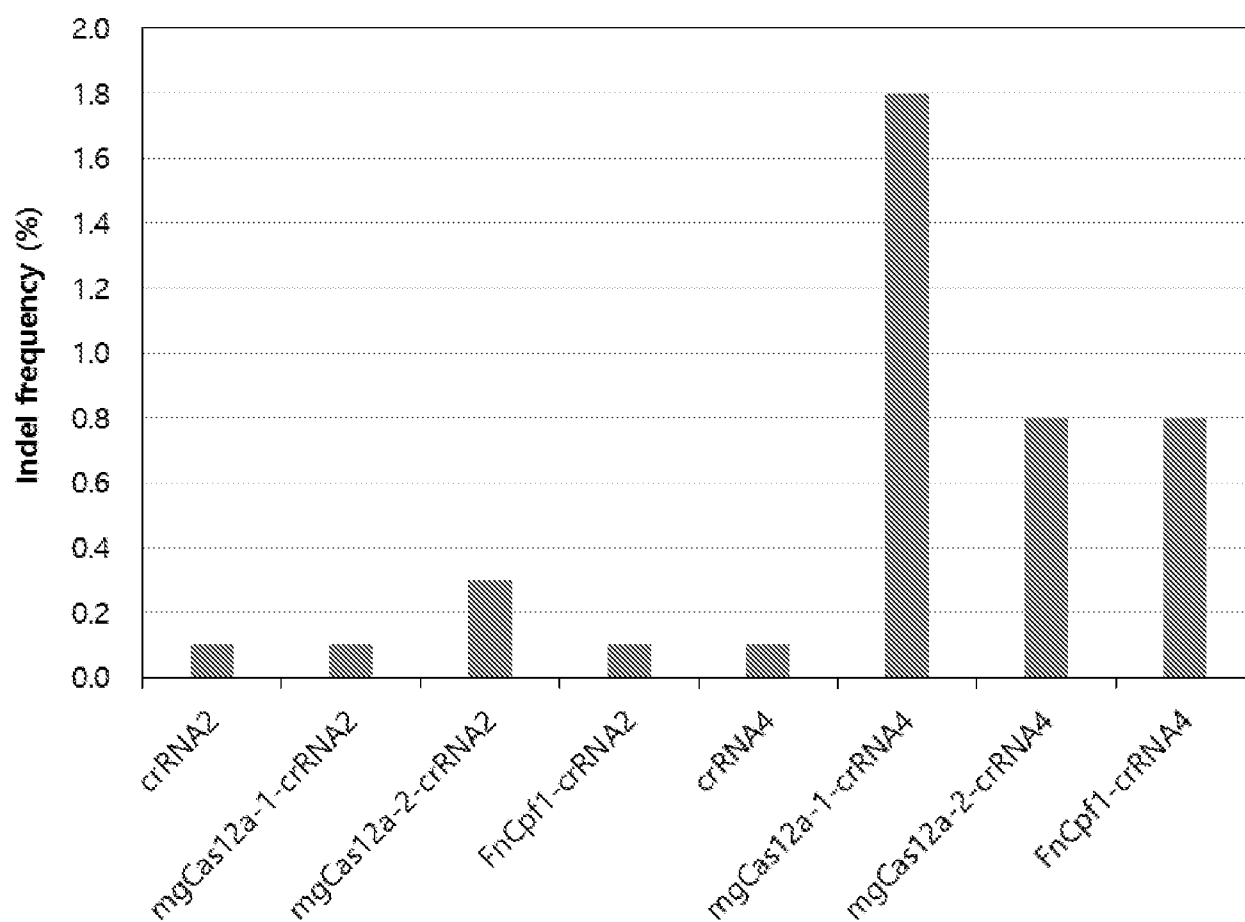


FIG. 28

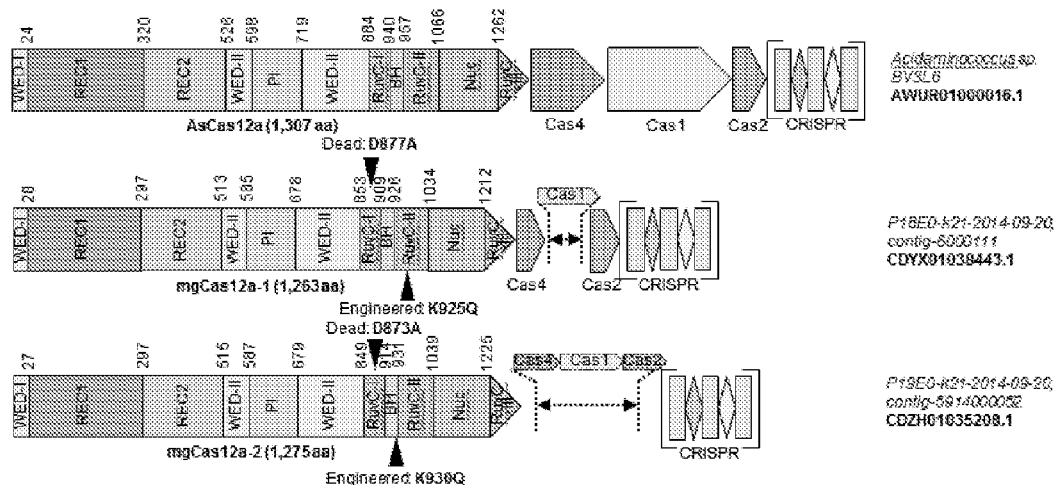


FIG. 20C