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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 September 2024 (26.09.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/196921 A1

(51) International Patent Classification:
C12N 9/22 (2006.01)

(21) International Application Number:
PCT/US2024/020512

(22) International Filing Date:
19 March 2024 (19.03.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/491,118 20 March 2023 (20.03.2023) US
63/585,659 27 September 2023 (27.09.2023) US

(71) Applicant: **PIONEER HI-BRED INTERNATIONAL, INC.** [US/US]; 7100 NW 62ND Avenue, P.O. Box 1014, Johnston, Iowa 50131-1014 (US).

(72) Inventor: **YOUNG, Joshua K.**; 7250 NW 62ND Avenue, P.O. Box 552, Johnston, Iowa 50131-0552 (US).

(74) Agent: **FORTINI, Jenna** et al.; 7250 NW 62ND Avenue, P.O. Box 552, Johnston, Iowa 50131-0552 (US).

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

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

Declarations under Rule 4.17:
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

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

(54) Title: CAS POLYPEPTIDES WITH ALTERED PAM RECOGNITION

(57) Abstract: The present disclosure relates to methods for altering the PAM specificity of Cas-alpha polypeptides. The disclosure also relates to Cas-alpha 10 polypeptides having altered PAM specificity and methods and compositions for use thereof.

WO 2024/196921 A1

CAS POLYPEPTIDES WITH ALTERED PAM RECOGNITION

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to US Provisional Application No. 63/491,118, filed March 20, 2023, and US Provisional Application No. 63/585,659, filed September 27, 2023, which are incorporated by reference herein in their entireties.

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

[0002] The official copy of the sequence listing is submitted electronically as an XML formatted sequence listing with a file named "108587-WO-SEC-1_Sequence_Listing_ST26" created on March 18, 2024 and having a size of 119 kilobytes, and is filed concurrently with the specification. The sequence listing comprised in this XML formatted document is part of the specification and is herein incorporated by reference in its entirety.

FIELD

[0003] The disclosure relates to the field of molecular biology, in particular to compositions of novel polynucleotide-guided Cas polypeptides, and compositions and methods for editing or modifying the genome of a cell.

BACKGROUND

[0004] Recombinant DNA technology has made it possible to insert DNA sequences at targeted genomic locations and/or modify specific endogenous chromosomal sequences. Site-specific integration techniques, which employ site-specific recombination systems, as well as other types of recombination technologies, have been used to generate targeted insertions of genes of interest in a variety of organism. Genome-editing techniques such as designer zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), or homing meganucleases, are available for producing targeted genome perturbations, but these systems tend to have low specificity and employ designed nucleases that need to be redesigned for each target site, which renders them costly and time-consuming to prepare.

[0005] Newer technologies utilizing archaeal or bacterial adaptive immunity systems have been identified, called CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), which

comprise different domains of effector proteins that encompass a variety of activities (DNA recognition, binding, and optionally cleavage).

[0006] The protospacer adjacent motif (PAM) requirement of CRISPR-associated (Cas) polypeptides restricts their targeting range (Shmakov et al, 2015; Zetsche et al, 2015; Burstein et al, 2017; Karvelis et al, 2020; Pausch et al, 2020). This becomes particularly apparent in genome editing applications where the outcome is dependent on the proximity of the desired edit to the cut site (e.g., template-free editing and homology-directed repair) or for approaches that impose additional sequence requirements on target selection (e.g., base editing; Anzalone et al, 2020).

[0007] Disclosed herein are methods for altering the PAM specificity of Cas-alpha polypeptides. Also disclosed are Cas-alpha 10 polypeptides having altered PAM specificity and methods and compositions for use thereof.

SUMMARY

[0008] In a first aspect, the disclosure provides methods for altering protospacer adjacent motif (PAM) specificity of a target Cas-alpha polypeptide, the method comprising: (a) comparing the PAM interacting (PI) domain of an orthologous Cas-alpha polypeptide with the PI domain of the target Cas-alpha polypeptide, wherein the orthologous Cas-alpha polypeptide has different PAM specificity than the target Cas-alpha polypeptide; (b) selecting one or more amino acids and/or one or more polypeptide chains from the PI domain of the orthologous Cas-alpha polypeptide; (c) incorporating the one or more amino acids and/or the one or more polypeptide chains selected from the PI domain of the orthologous Cas-alpha polypeptide into one or more structurally similar positions of the target Cas-alpha polypeptide resulting in a modified target Cas-alpha polypeptide; and (d) determining PAM recognition of the modified target Cas-alpha polypeptide.

[0009] In some examples of the method for altering PAM specificity of a target Cas-alpha polypeptide, the orthologous Cas-alpha polypeptide has different PAM recognition than the target Cas-alpha polypeptide.

[0010] In some examples of the method for altering PAM specificity of a target Cas-alpha polypeptide, the target Cas-alpha polypeptide is a Cas-alpha 10 polypeptide.

[0011] In some examples of the method for altering PAM specificity of a target Cas-alpha polypeptide, the Cas-alpha 10 polypeptide comprises an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 2, and the PI domain comprises amino acids from S63 to I196.

[0012] In some examples of the method for altering PAM specificity of a target Cas-alpha polypeptide, the orthologous Cas-alpha polypeptide is Cas-alpha 1, Cas-alpha 2, Cas-alpha 3, Cas-alpha 4, Cas-alpha 5, Cas-alpha 6, Cas-alpha 7, Cas-alpha 8, or Cas-alpha 11.

[0013] In a second aspect, the disclosure provides a synthetic or non-naturally occurring Cas-alpha 10 polypeptide comprising an amino acid amino acid sequence having at least 90% sequence identity to any one of SEQ ID Nos. 5-16 or 28-87. In some examples of this second aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide has endonuclease activity. In some examples of this second aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide is a deactivated Cas-alpha 10 endonuclease complexed to a deaminase. In some examples of this second aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide is complexed to a heterologous protein domain through a linker, and wherein the heterologous protein domain has methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, or nucleic acid binding activity.

[0014] In a third aspect, the disclosure provides a synthetic or non-naturally occurring Cas-alpha 10 polypeptide comprising a PAM interacting (PI) domain, wherein the PI domain recognizes a PAM sequence comprising 5'-YTC-3', 5'-TTTC-3', 5'-GTTC-3', 5'-GTYC-3', 5'-GTWC-3', 5'-DTHC-3', 5'-GTNC-3', 5'-NNNC-3', 5'-GNNC-3', 5'-HCTC-3', 5'-NTCC-3', 5'-TNC-3', 5'-GTTY-3', 5'-DTTY-3', 5'-GTTT-3', 5'-GTHC-3', 5'-NYTY-3', 5'-TTTY-3', 5'-TTYTY-3', 5'-TYTY-3', 5'-TTNN-3', 5'-NNCC-3', 5'-NCCC-3', 5'-NTNN-3', 5'-NCTY-3', 5'-NTHC-3', 5'-NNNN-3', or 5'-NNCN-3', wherein N = A, C, G, or T, Y = T or C, D = G, A, or T, W = A or T, and H = A, T, or C. In some examples of this third aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide has endonuclease activity. In some examples of this third aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide is a deactivated Cas-alpha 10 endonuclease complexed to a deaminase. In some examples of this third aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide is complexed to a heterologous protein domain through a linker, and wherein the heterologous protein domain has methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, reverse transcriptase activity, RNA aptamer binding activity, or nucleic acid binding activity. In further examples of this third aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide has nickase activity. In

yet further examples of this third aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide is complexed to a reverse transcriptase or operably associated with a reverse transcriptase.

[0015] In a fourth aspect, the disclosure provides a synthetic or non-naturally occurring Cas-alpha 10 polypeptide comprising an amino acid amino acid sequence having at least 90% sequence identity to SEQ ID NO: 2,wherein the amino acid sequence comprises a mutation relative to the amino acid position of SEQ ID NO: 2, wherein the mutation comprises a K85S mutation; a K85A mutation; a N92R mutation; a N92K mutation; a Q125K mutation; a N88H mutation; a N88K mutation; a N88Q mutation; a Q125F mutation; a Y72V mutation; a Y72E mutation; a Y72Q mutation; a Y72T mutation; a Y72C mutation; a Y72A mutation; a Y72S mutation; a Y72P mutation; a Y72G mutation; a Y72D mutation; a Y72L mutation; a K85G mutation; a K85D mutation; a K85N mutation; a N88D mutation; a Q89D mutation; a N92D mutation; a N92C mutation; a N92A mutation; a N92G mutation; a N92F mutation; a N92E mutation; a N92H mutation; a N92I mutation; a N92L mutation; a N92V mutation; a N92W mutation; a N92Q mutation; a N92S mutation; a N92P mutation; a N92Y mutation; a N92T mutation; a N92M mutation; a Q125R mutation; or a Q125P mutation; or one or more of the following combinations of mutations: a combination of a K85Q mutation and a N92L mutation; a combination of a N88H mutation and a Q89G mutation; a combination of a N88K mutation and a Q89G mutation; a combination of a N88Q mutation and a Q89G mutation; a combination of a K85S mutation and a N92L; a combination of a K85S mutation and a N92Q mutation; a combination of a K85S mutation and a N92C mutation; a combination of a K85S mutation and a N92H mutation; a combination of a K85S mutation and a N92A mutation; a combination of a K85S mutation and a N92M mutation; a combination of a K85N mutation and a N92L mutation; a combination of a K85N mutation and a N92H; a combination of a K85N mutation and a N92A mutation; a combination of a K85N mutation and a N92C mutation; a combination of a K85N mutation and a N92M mutation; a combination of a K85N mutation and a N92Q mutation; a combination of a K85N mutation and a N92I mutation; a combination of a K85S mutation, a N88D mutation, and a Q89G mutation; a combination of a K85S mutation, a N88H mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89D mutation, and a Q125R mutation; a combination of a K85S mutation, a N88D mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89G mutation, and a Q125R mutation; a

combination of a K85S mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72S mutation, a K85D mutation, a Q125R mutation, and a N127R mutation; a combination of a Y72A mutation, a K85S mutation, a Q89D mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88H mutation, a Q89G mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88D mutation, a Q89D mutation, a N92W mutation, and a Q125R mutation; or a combination of a K85Q and a N92W mutation. In particular examples of this fourth aspect, the amino acid sequence of the Cas-alpha 10 polypeptide comprises the K85S mutation. In other examples of this fourth aspect, the amino acid sequence of the Cas-alpha 10 polypeptide comprises the combination of the K85N mutation and the N92L mutation. In some examples of this fourth aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide has endonuclease activity. In some examples of this fourth aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide is a deactivated Cas-alpha 10 endonuclease complexed to a deaminase. In some examples of this fourth aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide is complexed to a heterologous protein domain through a linker, and wherein the heterologous protein domain has methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, reverse transcriptase activity, RNA aptamer binding activity, or nucleic acid binding activity. In further examples of this fourth aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide has nickase activity. In yet further examples of this fourth aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide is complexed to a reverse transcriptase or operably associated with a reverse transcriptase.

[0016] In a fifth aspect, the disclosure provides a synthetic composition comprising: (a) a Cas-alpha 10 polypeptide having DNA binding activity, the Cas-alpha 10 polypeptide comprising an amino acid amino acid sequence having at least 90% sequence identity to any one of SEQ ID Nos. 5-16 or 28-87; and (b) at least one guide polynucleotide comprising a region of complementarity to a target polynucleotide, wherein the target polynucleotide is heterologous to the source of the Cas-alpha 10 polypeptide, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide and forms a complex with the at least one guide polynucleotide, and wherein the complex binds to the target polynucleotide. In some examples of the synthetic composition, the Cas-alpha 10 polypeptide has endonuclease activity. In some examples of the synthetic composition, the Cas-alpha 10 polypeptide is a deactivated Cas-alpha 10 endonuclease

complexed to a deaminase. In some examples of the synthetic composition, the Cas-alpha 10 polypeptide is complexed to a heterologous protein domain through a linker, and wherein the heterologous protein domain has methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, reverse transcriptase activity, RNA aptamer binding activity, or nucleic acid binding activity. In further examples of this fifth aspect, the Cas-alpha 10 polypeptide has nickase activity. In yet further examples of this fifth aspect, the Cas-alpha 10 polypeptide is complexed to a reverse transcriptase or operably associated with a reverse transcriptase

[0017] In a sixth aspect, the disclosure provides a synthetic composition comprising: (a) a Cas-alpha 10 polypeptide having DNA binding activity, the Cas-alpha 10 polypeptide comprising a PAM interacting (PI) domain, wherein the PI domain recognizes a PAM sequence on a target polynucleotide, wherein the PAM sequence comprises 5'-YTC-3', 5'-TTTC-3', 5'-GTTC-3', 5'-GTYC-3', 5'-GTWC-3', 5'-DTHC-3', 5'-GTNC-3', 5'-NNNC-3', 5'-GNNC-3', 5'-HCTC-3', 5'-NTCC-3', 5'-TNC-3', 5'-GTTY-3', 5'-DTTY-3', 5'-GTTT-3', 5'-GTHC-3', 5'-NYTY-3', 5'-TTTY-3', 5'-TTYTY-3', 5'-TYTY-3', 5'-TTNN-3', 5'-NNCC-3', 5'-NCCC-3', 5'-NTNN-3', 5'-NCTY-3', 5'-NTHC-3', 5'-NNNN-3', or 5'-NNCN-3', wherein N = A, C, G, or T, Y = T or C, D = G, A, or T, W = A or T, and H = A, T, or C; and (b) at least one guide polynucleotide comprising a region of complementarity to a target polynucleotide, wherein the target polynucleotide is heterologous to the source of the Cas-alpha 10 polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the at least one guide polynucleotide, and wherein the complex binds to the target polynucleotide. In some examples of the synthetic composition, the Cas-alpha 10 polypeptide has endonuclease activity. In some examples of the synthetic composition, the Cas-alpha 10 polypeptide is a deactivated Cas-alpha 10 endonuclease complexed to a deaminase. In some examples of the synthetic composition, the Cas-alpha 10 polypeptide is complexed to a heterologous protein domain through a linker, and wherein the heterologous protein domain has methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, reverse transcriptase activity, RNA aptamer binding activity, or nucleic acid binding activity. In further examples of this sixth aspect, the Cas-alpha 10 polypeptide has nickase activity. In yet further examples of this sixth aspect, the Cas-alpha 10 polypeptide is complexed to a reverse transcriptase or operably associated with a reverse transcriptase.

[0018] In a seventh aspect, the disclosure provides a synthetic composition comprising: (a) a Cas-alpha 10 polypeptide having DNA binding activity, the Cas-alpha 10 polypeptide comprising an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 2, wherein the amino acid sequence comprises a mutation relative to the amino acid position of SEQ ID NO: 2, wherein the mutation comprises a K85S mutation; a K85A mutation; a N92R mutation; a N92K mutation; a Q125K mutation; a N88H mutation; a N88K mutation; a N88Q mutation; a Q125F mutation; a Y72V mutation; a Y72E mutation; a Y72Q mutation; a Y72T mutation; a Y72C mutation; a Y72A mutation; a Y72S mutation; a Y72P mutation; a Y72G mutation; a Y72D mutation; a Y72L mutation; a K85G mutation; a K85D mutation; a K85N mutation; a N88D mutation; a Q89D mutation; a N92D mutation; a N92C mutation; a N92A mutation; a N92G mutation; a N92F mutation; a N92E mutation; a N92H mutation; a N92I mutation; a N92L mutation; a N92V mutation; a N92W mutation; a N92Q mutation; a N92S mutation; a N92P mutation; a N92Y mutation; a N92T mutation; a N92M mutation; a Q125R mutation; or a Q125P mutation; or one or more of the following combinations of mutations: a combination of a K85Q mutation and a N92L mutation; a combination of a N88H mutation and a Q89G mutation; a combination of a N88K mutation and a Q89G mutation; a combination of a N88Q mutation and a Q89G mutation; a combination of a K85S mutation and a N92L; a combination of a K85S mutation and a N92Q mutation; a combination of a K85S mutation and a N92C mutation; a combination of a K85S mutation and a N92H mutation; a combination of a K85S mutation and a N92A mutation; a combination of a K85S mutation and a N92M mutation; a combination of a K85N mutation and a N92L mutation; a combination of a K85N mutation and a N92H; a combination of a K85N mutation and a N92A mutation; a combination of a K85N mutation and a N92C mutation; a combination of a K85N mutation and a N92M mutation; a combination of a K85N mutation and a N92Q mutation; a combination of a K85N mutation and a N92I mutation; a combination of a K85S mutation, a N88D mutation, and a Q89G mutation; a combination of a K85S mutation, a N88H mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89D mutation, and a Q125R mutation; a combination of a K85S mutation, a N88D mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89G mutation, and a Q125R mutation; a combination of a K85S mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72S mutation, a K85D mutation, a Q125R mutation, and a N127R mutation; a combination of a Y72A mutation, a K85S mutation, a Q89D

mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88H mutation, a Q89G mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88D mutation, a Q89D mutation, a N92W mutation, and a Q125R mutation; or a combination of a K85Q and a N92W mutation; and (b) at least one guide polynucleotide comprising a region of complementarity to a target polynucleotide, wherein the target polynucleotide is heterologous to the source of the Cas-alpha 10 polypeptide, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide and forms a complex with the at least one guide polynucleotide, and wherein the complex binds to the target polynucleotide.

[0019] In particular examples of this seventh aspect, the amino acid sequence of the Cas-alpha 10 polypeptide comprises the K85S mutation. In other examples of this seventh aspect, the amino acid sequence of the Cas-alpha 10 polypeptide comprises the combination of the K85N mutation and the N92L mutation. In some examples of the synthetic composition, the Cas-alpha 10 polypeptide has endonuclease activity. In some examples of the synthetic composition, the Cas-alpha 10 polypeptide is a deactivated Cas-alpha 10 endonuclease complexed to a deaminase. In some examples of the synthetic composition, the Cas-alpha 10 polypeptide is complexed to a heterologous protein domain through a linker, and wherein the heterologous protein domain has methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, reverse transcriptase activity, RNA aptamer binding activity, or nucleic acid binding activity. In further examples of this seventh aspect, the Cas-alpha 10 polypeptide has nickase activity. In yet further examples of this seventh aspect, the Cas-alpha 10 polypeptide is complexed to a reverse transcriptase or operably associated with a reverse transcriptase.

[0020] In an eighth aspect, the disclosure provides a method of editing a target polynucleotide in a cell, the method comprising: (a) providing to the cell a Cas-alpha 10 polypeptide having at least 90% sequence identify to any one of SEQ ID Nos. 5-16 or 28-87, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide; (b) providing the cell with at least one guide polynucleotide comprising a region of complementarity to the target polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the guide polynucleotide and the complex binds the target polynucleotide; and (c) introducing at least one nucleotide modification in the target polynucleotide via the complex, wherein the target polynucleotide is heterologous to the Cas-alpha 10 polypeptide.

[0021] In some examples of the eighth aspect, the PAM sequence recognized by the Cas-alpha 10 polypeptide comprises 5'-YTC-3', 5'-TTTC-3', 5'-GTTC-3', 5'-GTYC-3', 5'-GTWC-3', 5'-DTHC-3', 5'-GTNC-3', 5'-NNNC-3', 5'-GNNC-3', 5'-HCTC-3', 5'-NTCC-3', 5'-TNC-3', 5'-GTTY-3', 5'-DTTY-3', 5'-GTTT-3', 5'-GTHC-3', 5'-NYTY-3', 5'-TTTY-3', 5'-TTYTY-3', 5'-TYTY-3', 5'-TTNN-3', 5'-NNCC-3', 5'-NCCC-3', 5'-NTNN-3', 5'-NCTY-3', 5'-NTHC-3', 5'-NNNN-3', or 5'-NNCN-3', wherein N = A, C, G, or T, Y = T or C, D = G, A, or T, W = A or T, and H = A, T, or C.

[0022] In some examples of the eighth aspect, the method further comprises providing the cell with a donor DNA molecular or a polynucleotide modification template.

[0023] In some examples of the eighth aspect, the cell is derived or obtained from an animal, a fungus, or a plant. In some examples of the eighth aspect, the plant is a dicot or a monocot. In some aspects, the plant is maize, soybean, cotton, wheat, canola, oilseed rape, sorghum, rice, rye, barley, millet, oats, sugarcane, turfgrass, switchgrass, alfalfa, sunflower, tobacco, peanut, potato, Arabidopsis, safflower, or tomato.

[0024] In some examples of the eighth aspect, the Cas-alpha 10 polypeptide has endonuclease activity.

[0025] In some examples of the eighth aspect, the at least one guide polynucleotide comprises a plurality of guide polynucleotides and the Cas-alpha 10 polypeptide is a deactivated Cas-alpha 10 endonuclease complexed to a deaminase.

[0026] In some examples of the eighth aspect, the Cas-alpha 10 polypeptide is complexed to a heterologous protein domain through a linker, and wherein the heterologous protein domain has methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, reverse transcriptase activity, RNA aptamer binding activity, or nucleic acid binding activity. In further examples of this eighth aspect, the Cas-alpha 10 polypeptide has nickase activity. In yet further examples of this eighth aspect, the Cas-alpha 10 polypeptide is complexed to a reverse transcriptase or operably associated with a reverse transcriptase.

[0027] In a ninth aspect, the disclosure provides a method of editing a target polynucleotide in a cell, the method comprising: (a) providing to the cell a Cas-alpha 10 polypeptide comprising a PAM interacting (PI) domain, wherein the PI domain recognizes a PAM sequence on the target polynucleotide, and wherein the PAM sequence comprises 5'-YTC-3', 5'-TTTC-3', 5'-GTTC-3',

5'-GTYC-3', 5'-GTWC-3', 5'-DTHC-3', 5'-GTNC-3', 5'-NNNC-3', 5'-GNNC-3', 5'-HCTC-3', 5'-NTCC-3', 5'-TNC-3', 5'-GTTY-3', 5'-DTTY-3', 5'-GTTT-3', 5'-GTHC-3', 5'-NYTY-3', 5'-TTTY-3', 5'-TTYTY-3', 5'-TYTY-3', 5'-TTNN-3', 5'-NNCC-3', 5'-NCCC-3', 5'-NTNN-3', 5'-NCTY-3', 5'-NTHC-3', 5'-NNNN-3', or 5'-NNCN-3', wherein N = A, C, G, or T, Y = T or C, D = G, A, or T, W = A or T, and H = A, T, or C; (b) providing the cell with at least one guide polynucleotide comprising a region of complementarity to the target polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the guide polynucleotide and the complex binds the target polynucleotide; and (c) introducing at least one nucleotide modification in the target polynucleotide via the complex, wherein the target polynucleotide is heterologous to the Cas-alpha 10 polypeptide.

[0028] In some examples of the ninth aspect, the method further comprises providing the cell with a donor DNA molecular or a polynucleotide modification template.

[0029] In some examples of the ninth aspect, the cell is derived or obtained from an animal, a fungus, or a plant. In some examples, the plant is a dicot or a monocot. In some examples, the plant is maize, soybean, cotton, wheat, canola, oilseed rape, sorghum, rice, rye, barley, millet, oats, sugarcane, turfgrass, switchgrass, alfalfa, sunflower, tobacco, peanut, potato, Arabidopsis, safflower, or tomato.

[0030] In some examples of the ninth aspect, the Cas-alpha 10 polypeptide has endonuclease activity.

[0031] In some examples of the ninth aspect, the at least one guide polynucleotide comprises a plurality of guide polynucleotides and the Cas-alpha 10 polypeptide is a deactivated Cas-alpha 10 endonuclease complexed to a deaminase.

[0032] In some examples of the ninth aspect, the Cas-alpha 10 polypeptide is complexed to a heterologous protein domain through a linker, and wherein the heterologous protein domain has methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, reverse transcriptase activity, RNA aptamer binding activity, or nucleic acid binding activity. In further examples of this ninth aspect, the Cas-alpha 10 polypeptide has nickase activity. In yet further examples of this ninth aspect, the Cas-alpha 10 polypeptide is complexed to a reverse transcriptase or operably associated with a reverse transcriptase.

[0033] In a tenth aspect, the disclosure provides a method of editing a target polynucleotide in a cell, the method comprising: (a) providing to the cell a Cas-alpha 10 polypeptide comprising an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 2, wherein the amino acid sequence comprises a mutation relative to the amino acid position of SEQ ID NO: 2, wherein the mutation comprises a K85S mutation; a K85A mutation; a N92R mutation; a N92K mutation; a Q125K mutation; a N88H mutation; a N88K mutation; a N88Q mutation; a Q125F mutation; a Y72V mutation; a Y72E mutation; a Y72Q mutation; a Y72T mutation; a Y72C mutation; a Y72A mutation; a Y72S mutation; a Y72P mutation; a Y72G mutation; a Y72D mutation; a Y72L mutation; a K85G mutation; a K85D mutation; a K85N mutation; a N88D mutation; a Q89D mutation; a N92D mutation; a N92C mutation; a N92A mutation; a N92G mutation; a N92F mutation; a N92E mutation; a N92H mutation; a N92I mutation; a N92L mutation; a N92V mutation; a N92W mutation; a N92Q mutation; a N92S mutation; a N92P mutation; a N92Y mutation; a N92T mutation; a N92M mutation; a Q125R mutation; or a Q125P mutation; or one or more of the following combinations of mutations: a combination of a K85Q mutation and a N92L mutation; a combination of a N88H mutation and a Q89G mutation; a combination of a N88K mutation and a Q89G mutation; a combination of a N88Q mutation and a Q89G mutation; a combination of a K85S mutation and a N92L; a combination of a K85S mutation and a N92Q mutation; a combination of a K85S mutation and a N92C mutation; a combination of a K85S mutation and a N92H mutation; a combination of a K85S mutation and a N92A mutation; a combination of a K85S mutation and a N92M mutation; a combination of a K85N mutation and a N92L mutation; a combination of a K85N mutation and a N92H; a combination of a K85N mutation and a N92A mutation; a combination of a K85N mutation and a N92C mutation; a combination of a K85N mutation and a N92M mutation; a combination of a K85N mutation and a N92Q mutation; a combination of a K85N mutation and a N92I mutation; a combination of a K85S mutation, a N88D mutation, and a Q89G mutation; a combination of a K85S mutation, a N88H mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89D mutation, and a Q125R mutation; a combination of a K85S mutation, a N88D mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89G mutation, and a Q125R mutation; a combination of a K85S mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72S mutation, a K85D mutation, a Q125R mutation, and a N127R mutation; a combination of a Y72A mutation, a K85S mutation, a Q89D

mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88H mutation, a Q89G mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88D mutation, a Q89D mutation, a N92W mutation, and a Q125R mutation; or a combination of a K85Q and a N92W mutation, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide; (b) providing the cell with at least one guide polynucleotide comprising a region of complementarity to the target polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the guide polynucleotide and the complex binds the target polynucleotide; and (c) introducing at least one nucleotide modification in the target polynucleotide via the complex, wherein the target polynucleotide is heterologous to the Cas-alpha 10 polypeptide.

[0034] In some examples of this tenth aspect, the amino acid sequence of the Cas-alpha 10 polypeptide comprises the K85S mutation. In other examples of this tenth aspect, the amino acid sequence of the Cas-alpha 10 polypeptide comprises the combination of the K85N mutation and the N92L mutation.

[0035] In some examples of the tenth aspect, the PAM sequence recognized by the Cas-alpha 10 polypeptide comprises 5'-YTC-3', 5'-TTTC-3', 5'-GTTC-3', 5'-GTYC-3', 5'-GTWC-3', 5'-DTHC-3', 5'-GTNC-3', 5'-NNNC-3', 5'-GNNC-3', 5'-HCTC-3', 5'-NTCC-3', 5'-TNC-3', 5'-GTTY-3', 5'-DTTY-3', 5'-GTTT-3', 5'-GTHC-3', 5'-NYTY-3', 5'-TTTY-3', 5'-TTYTY-3', 5'-TYTY-3', 5'-TTNN-3', 5'-NNCC-3', 5'-NCCC-3', 5'-NTNN-3', 5'-NCTY-3', 5'-NTHC-3', 5'-NNNN-3', or 5'-NNCN-3', wherein N = A, C, G, or T, Y = T or C, D = G, A, or T, W = A or T, and H = A, T, or C.

[0036] In some examples of the tenth aspect, the method further comprises providing the cell with a donor DNA molecular or a polynucleotide modification template.

[0037] In some examples of the tenth aspect, the cell is derived or obtained from an animal, a fungus, or a plant. In some aspects, the plant is a dicot or a monocot. In some examples, the plant is maize, soybean, cotton, wheat, canola, oilseed rape, sorghum, rice, rye, barley, millet, oats, sugarcane, turfgrass, switchgrass, alfalfa, sunflower, tobacco, peanut, potato, Arabidopsis, safflower, or tomato.

[0038] In some examples of the tenth aspect, the Cas-alpha 10 polypeptide has endonuclease activity.

[0039] In some examples of the tenth aspect, the at least one guide polynucleotide comprises a plurality of guide polynucleotides and the Cas-alpha 10 polypeptide is a deactivated Cas-alpha 10 endonuclease complexed to a deaminase.

[0040] In some examples of the tenth aspect, the Cas-alpha 10 polypeptide is complexed to a heterologous protein domain through a linker, and wherein the heterologous protein domain has methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, reverse transcriptase activity, RNA aptamer binding activity, or nucleic acid binding activity. In further examples of this tenth aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide has nickase activity. In yet further examples of this tenth aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide is complexed to a reverse transcriptase or operably associated with a reverse transcriptase.

[0041] In an eleventh aspect, the disclosure provides an animal cell or a fungal cell comprising any of the synthetic or non-naturally occurring Cas-alpha 10 polypeptides as described herein.

[0042] In a twelfth aspect, the disclosure provides a plant cell, a plant part, a plantlet, or a plant comprising any of the synthetic or non-naturally occurring Cas-alpha 10 polypeptides as described herein.

[0043] In a thirteenth aspect, the disclosure provides methods for altering protospacer adjacent motif (PAM) specificity of a target Cas-alpha 10 polypeptide, the method comprising (a) comparing the PAM interacting (PI) domain of an orthologous Cas-alpha polypeptide with the PI domain of the target Cas-alpha 10 polypeptide; (b) selecting one or more amino acids and/or one or more polypeptide chains from the PI domain of the orthologous Cas-alpha polypeptide; (c) incorporating the one or more amino acids and/or the one or more polypeptide chains selected from the PI domain of the orthologous Cas-alpha polypeptide into one or more structurally similar positions of the target Cas-alpha 10 polypeptide resulting in a modified target Cas-alpha 10 polypeptide; and (d) determining PAM recognition of the modified target Cas-alpha 10 polypeptide. In some examples of this thirteenth aspect, the orthologous Cas-alpha polypeptide has different PAM recognition than the target Cas-alpha 10 polypeptide. In some examples of this thirteenth aspect, the Cas-alpha 10 polypeptide comprises an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 2. In further examples of this thirteenth example, the

orthologous Cas-alpha polypeptide is Cas-alpha 1, Cas-alpha 2, Cas-alpha 3, Cas-alpha 4, Cas-alpha 5, Cas-alpha 6, Cas-alpha 7, Cas-alpha 8, or Cas-alpha 11.

BRIEF DESCRIPTION OF THE DRAWINGS AND THE SEQUENCE LISTING

[0044] The disclosure can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing, which form a part of this application.

[0045] **FIG. 1** illustrates the phylogenetic relationships among some of the Cas-alpha orthologs. Three supergroups were identified (I, II, and III). Group I comprised Clade 1 (Candidate Archaea and Aureabacteria (Cas1, Cas2, Cas4 typically encoded in the locus)). Group II comprised Clade 2 (Aquificae (Sulfurihydrogenibium and Hydrogenivirga genera) and Deltaproteobacteria (Desulfovibrio genus)), Clade 3 (Candidate Archaea (Cas1, Cas2, and Cas4 typically encoded in the locus)), Clade 4 (Bacteroidetes (Prevotella and Bacteroides genera)), Clade 5 (Candidate Levybacterium), and Clade 6 (Clostridia (Dorea, Ruminococcus, Clostridium, Clostridioides, Peptocolstridium, Cellulosilyticym, Eubacterium, Syntrophomonas genera)). Group III comprised Clade 7 (Bacilli (Bacillus, Acidibacillus, Aneurinibacillus, Brevibacillus, Parageobacillus, Alicyclobacillus genera)), Clade 8 (Negativicutes (Phascolarctobacterium genus)), and Clade 9 (Flavobacteriia (Flavobacterium genus)). A diamond symbol represents Cas-alpha endonucleases 1-11 whose PAM recognition was described in US10934536.

[0046] **FIG. 2** illustrates an expression cassette expressing a Cas-alpha 10 or a synthetic or non-naturally occurring Cas-alpha 10 polypeptide as described herein.

[0047] **FIG. 3** illustrates a method for altering PAM specificity in a Cas-alpha polypeptide.

[0048] **FIG. 4** illustrates another method for altering PAM specificity in a Cas-alpha polypeptide.

[0049] **FIG. 5** illustrates yet another method for altering PAM specificity in a Cas-alpha polypeptide.

[0050] **FIG. 6** is a graph illustrating the number of target sites in the maize and human genome targeted by several PAM variants of Example 2.

[0051] **SEQ ID NO: 1** is the PRT sequence of a Cas-alpha 10 polypeptide from *Syntrophomonas palmitatica*.

[0052] **SEQ ID NO: 2** is a PRT sequence of a first exemplary synthetic or non-naturally occurring Cas-alpha 10 polypeptide.

[0053] **SEQ ID NO: 3** is a PRT sequence of a Cas-alpha 4 polypeptide from uncultured archaeon.

[0054] **SEQ ID NO: 4** is a PRT sequence of a Cas-alpha 8 polypeptide from *Acidibacillus sulfuroxidans*.

[0055] **SEQ ID NO: 5** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85A mutation relative to SEQ ID NO: 2.

[0056] **SEQ ID NO: 6** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85S mutation relative to SEQ ID NO: 2.

[0057] **SEQ ID NO: 7** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92R mutation relative to SEQ ID NO: 2.

[0058] **SEQ ID NO: 8** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92K mutation relative to SEQ ID NO: 2.

[0059] **SEQ ID NO: 9** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Q125K mutation relative to SEQ ID NO: 2.

[0060] **SEQ ID NO: 10** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N88H mutation relative to SEQ ID NO: 2.

[0061] **SEQ ID NO: 11** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N88H and a Q89G mutation relative to SEQ ID NO: 2.

[0062] **SEQ ID NO: 12** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N88K mutation relative to SEQ ID NO: 2.

[0063] **SEQ ID NO: 13** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N88K and a Q89G mutation relative to SEQ ID NO: 2.

[0064] **SEQ ID NO: 14** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N88Q mutation relative to SEQ ID NO: 2.

[0065] **SEQ ID NO: 15** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N88Q and a Q89G mutation relative to SEQ ID NO: 2.

[0066] **SEQ ID NO: 16** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Q125F mutation relative to SEQ ID NO: 2.

[0067] **SEQ ID NO: 17** is a PRT sequence Cas-alpha 1 polypeptide from Candidatus Micrarchaeota archaeon.

[0068] **SEQ ID NO: 18** is a PRT sequence Cas-alpha 2 polypeptide from *Candidatus Micrarchaeota* archaeon.

[0069] **SEQ ID NO: 19** is a PRT sequence Cas-alpha 3 polypeptide from *Candidatus Aureabacteria* bacterium.

[0070] **SEQ ID NO: 20** is a PRT sequence Cas-alpha 5 polypeptide from *Candidatus Micrarchaeota* archaeon.

[0071] **SEQ ID NO: 21** is a PRT sequence Cas-alpha 6 polypeptide from uncultured archaeon.

[0072] **SEQ ID NO: 22** is a PRT sequence Cas-alpha 7 polypeptide from *Parageobacillus thermoglucosidasius*.

[0073] **SEQ ID NO: 23** is a PRT sequence Cas-alpha 9 polypeptide from *Ruminococcus sp.*

[0074] **SEQ ID NO: 24** is a PRT sequence Cas-alpha 11 polypeptide from *Clostridium novyi*.

[0075] **SEQ ID NO: 25** is a PRT sequence Cas-alpha 13 polypeptide from *Clostridium paraputrificum*.

[0076] **SEQ ID NO: 26** is a PRT sequence Cas-alpha 24 polypeptide from *Bacillus toyonensis*.

[0077] **SEQ ID NO: 27** is a PRT sequence Cas-alpha 29 polypeptide from *Peptoclostridium sp.*

[0078] **SEQ ID NO: 28** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72V mutation relative to SEQ ID NO: 2.

[0079] **SEQ ID NO: 29** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72E mutation relative to SEQ ID NO: 2.

[0080] **SEQ ID NO: 30** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72Q mutation relative to SEQ ID NO: 2.

[0081] **SEQ ID NO: 31** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72T mutation relative to SEQ ID NO: 2.

[0082] **SEQ ID NO: 32** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72C mutation relative to SEQ ID NO: 2.

[0083] **SEQ ID NO: 33** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72A mutation relative to SEQ ID NO: 2.

[0084] **SEQ ID NO: 34** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72S mutation relative to SEQ ID NO: 2.

[0085] **SEQ ID NO: 35** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72P mutation relative to SEQ ID NO: 2.

[0086] **SEQ ID NO: 36** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72G mutation relative to SEQ ID NO: 2.

[0087] **SEQ ID NO: 37** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72D mutation relative to SEQ ID NO: 2.

[0088] **SEQ ID NO: 38** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72L mutation relative to SEQ ID NO: 2.

[0089] **SEQ ID NO: 39** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85G mutation relative to SEQ ID NO: 2.

[0090] **SEQ ID NO: 40** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85D mutation relative to SEQ ID NO: 2.

[0091] **SEQ ID NO: 41** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85N mutation relative to SEQ ID NO: 2.

[0092] **SEQ ID NO: 42** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N88D mutation relative to SEQ ID NO: 2.

[0093] **SEQ ID NO: 43** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Q89D mutation relative to SEQ ID NO: 2.

[0094] **SEQ ID NO: 44** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92D mutation relative to SEQ ID NO: 2.

[0095] **SEQ ID NO: 45** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92C mutation relative to SEQ ID NO: 2.

[0096] **SEQ ID NO: 46** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92A mutation relative to SEQ ID NO: 2.

[0097] **SEQ ID NO: 47** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92G mutation relative to SEQ ID NO: 2.

[0098] **SEQ ID NO: 48** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92F mutation relative to SEQ ID NO: 2.

[0099] **SEQ ID NO: 49** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92E mutation relative to SEQ ID NO: 2.

[0100] **SEQ ID NO: 50** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92H mutation relative to SEQ ID NO: 2.

[0101] **SEQ ID NO: 51** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92I mutation relative to SEQ ID NO: 2.

[0102] **SEQ ID NO: 52** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92L mutation relative to SEQ ID NO: 2.

[0103] **SEQ ID NO: 53** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92V mutation relative to SEQ ID NO: 2.

[0104] **SEQ ID NO: 54** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92W mutation relative to SEQ ID NO: 2.

[0105] **SEQ ID NO: 55** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92Q mutation relative to SEQ ID NO: 2.

[0106] **SEQ ID NO: 56** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92S mutation relative to SEQ ID NO: 2.

[0107] **SEQ ID NO: 57** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92P mutation relative to SEQ ID NO: 2.

[0108] **SEQ ID NO: 58** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92Y mutation relative to SEQ ID NO: 2.

[0109] **SEQ ID NO: 59** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92T mutation relative to SEQ ID NO: 2.

[0110] **SEQ ID NO: 60** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a N92M mutation relative to SEQ ID NO: 2.

[0111] **SEQ ID NO: 61** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Q125R mutation relative to SEQ ID NO: 2.

[0112] **SEQ ID NO: 62** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Q125P mutation relative to SEQ ID NO: 2.

[0113] **SEQ ID NO: 63** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85S and a N92L mutation relative to SEQ ID NO: 2.

[0114] **SEQ ID NO: 64** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85S and a N92Q mutation relative to SEQ ID NO: 2.

[0115] **SEQ ID NO: 65** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85S and a N92C mutation relative to SEQ ID NO: 2.

[0116] **SEQ ID NO: 66** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85S and a N92H mutation relative to SEQ ID NO: 2.

[0117] **SEQ ID NO: 67** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85S and a N92A mutation relative to SEQ ID NO: 2.

[0118] **SEQ ID NO: 68** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85S and a N92M mutation relative to SEQ ID NO: 2.

[0119] **SEQ ID NO: 69** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85N and a N92L mutation relative to SEQ ID NO: 2.

[0120] **SEQ ID NO: 70** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85N and a N92H mutation relative to SEQ ID NO: 2.

[0121] **SEQ ID NO: 71** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85N and a N92A mutation relative to SEQ ID NO: 2.

[0122] **SEQ ID NO: 72** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85N and a N92C mutation relative to SEQ ID NO: 2.

[0123] **SEQ ID NO: 73** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85N and a N92M mutation relative to SEQ ID NO: 2.

[0124] **SEQ ID NO: 74** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85N and a N92Q mutation relative to SEQ ID NO: 2.

[0125] **SEQ ID NO: 75** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85N and a N92I mutation relative to SEQ ID NO: 2.

[0126] **SEQ ID NO: 76** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85S, a N88D, and a Q89G mutation relative to SEQ ID NO: 2.

[0127] **SEQ ID NO: 77** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85S, a N88H, a Q89G, and a N92L mutation relative to SEQ ID NO: 2.

[0128] **SEQ ID NO: 78** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72A, a N88D, Q89D, and a Q125R mutation relative to SEQ ID NO: 2.

[0129] **SEQ ID NO: 79** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85S, a N88D, a Q89G, and a N92L mutation relative to SEQ ID NO: 2.

[0130] **SEQ ID NO: 80** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72A, a N88D, a Q89G, and a Q125R mutation relative to SEQ ID NO: 2.

[0131] **SEQ ID NO: 81** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85S, a N92L, and a Q125R mutation relative to SEQ ID NO: 2.

[0132] **SEQ ID NO: 82** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72S, a K85D, a Q125R, and a N127R mutation relative to SEQ ID NO: 2.

[0133] **SEQ ID NO: 83** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72A, a K85S, a Q89D, a N92L, and a Q125R mutation relative to SEQ ID NO: 2.

[0134] **SEQ ID NO: 84** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72C, a N88H, a Q89G, and a Q125R mutation relative to SEQ ID NO: 2.

[0135] **SEQ ID NO: 85** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a Y72C, a N88D, a Q89D, a N92W, and a Q125R mutation relative to SEQ ID NO: 2.

[0136] **SEQ ID NO: 86** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85Q and a N92L mutation relative to SEQ ID NO: 2.

[0137] **SEQ ID NO: 87** is a PRT sequence of a synthetic or non-naturally occurring Cas-alpha 10 polypeptide with a K85Q and a N92W mutation relative to SEQ ID NO: 2.

[0138] **SEQ ID NO: 88** is a PRT sequence of a second exemplary synthetic or non-naturally occurring Cas-alpha 10 polypeptide.

[0139] **SEQ ID NO: 89** is a PRT sequence of a third exemplary synthetic or non-naturally occurring Cas-alpha 10 polypeptide.

DETAILED DESCRIPTION

[0140] Compositions and methods are provided for novel CRISPR effector systems and elements comprising such systems, including, but not limiting to, novel guide polynucleotide/ endonuclease complexes, guide polynucleotides, guide RNA elements, Cas polypeptides, and endonucleases, as well as proteins comprising an endonuclease functionality (domain). Compositions and methods are also provided for direct delivery of endonucleases, cleavage ready complexes, guide RNAs, and guide RNA/Cas polypeptide complexes. The present disclosure further includes compositions and methods for genome modification of a target sequence in the genome of a cell, for gene editing, and for inserting a polynucleotide of interest into the genome of a cell.

[0141] Terms used in the claims and specification are defined as set forth below unless otherwise specified. It must be noted that, as used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise.

Definitions

[0142] As used herein, “nucleic acid” means a polynucleotide and includes a single or a double-stranded polymer of deoxyribonucleotide or ribonucleotide bases. Nucleic acids may also include fragments and modified nucleotides. Thus, the terms “polynucleotide”, “nucleic acid sequence”, “nucleotide sequence” and “nucleic acid fragment” are used interchangeably to denote a polymer of RNA and/or DNA and/or RNA-DNA that is single- or double-stranded, optionally comprising synthetic, non-natural, or altered nucleotide bases. Nucleotides (usually found in their 5’-monophosphate form) are referred to by their single letter designation as follows: “A” for adenosine or deoxyadenosine (for RNA or DNA, respectively), “C” for cytosine or deoxycytosine, “G” for guanosine or deoxyguanosine, “U” for uridine, “T” for deoxythymidine, “R” for purines (A or G), “Y” for pyrimidines (C or T), “K” for G or T, “H” for A or C or T, “I” for inosine, and “N” for any nucleotide.

[0143] The term “genome” as it applies to a prokaryotic and eukaryotic cell or organism cells encompasses not only chromosomal DNA found within the nucleus, but organelle DNA found within subcellular components (e.g., mitochondria, or plastid) of the cell.

[0144] “Open reading frame” is abbreviated ORF.

[0145] The term "selectively hybridizes" includes reference to hybridization, under stringent hybridization conditions, of a nucleic acid sequence to a specified nucleic acid target sequence to a detectably greater degree (e.g., at least 2-fold over background) than its hybridization to non-target nucleic acid sequences and to the substantial exclusion of non-target nucleic acids. Selectively hybridizing sequences typically have about at least 80% sequence identity, or 90% sequence identity, up to and including 100% sequence identity (i.e., fully complementary) with each other.

[0146] The term "stringent conditions" or “stringent hybridization conditions” includes reference to conditions under which a probe will selectively hybridize to its target sequence in an *in vitro* hybridization assay. Stringent conditions are sequence-dependent and will be different in different circumstances. By controlling the stringency of the hybridization and/or washing conditions, target sequences can be identified which are 100% complementary to the probe (homologous probing). Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). Generally, a probe is less than about 1000 nucleotides in length, optionally less than 500 nucleotides in length. Typically,

stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salt(s)) at pH 7.0 to 8.3, and at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C.

[0147] By "homology" is meant DNA sequences that are similar. For example, a "region of homology to a genomic region" that is found on the donor DNA is a region of DNA that has a similar sequence to a given "genomic region" in the cell or organism genome. A region of homology can be of any length that is sufficient to promote homologous recombination at the cleaved target site. For example, the region of homology can comprise at least 5-10, 5-15, 5-20, 5-25, 5-30, 5-35, 5-40, 5-45, 5- 50, 5-55, 5-60, 5-65, 5- 70, 5-75, 5-80, 5-85, 5-90, 5-95, 5-100, 5-200, 5-300, 5-400, 5-500, 5-600, 5-700, 5-800, 5-900, 5-1000, 5-1100, 5-1200, 5-1300, 5-1400, 5-1500, 5-1600, 5-1700, 5-1800, 5-1900, 5-2000, 5-2100, 5-2200, 5-2300, 5-2400, 5-2500, 5-2600, 5-2700, 5-2800, 5-2900, 5-3000, 5-3100 or more bases in length such that the region of homology has sufficient homology to undergo homologous recombination with the corresponding genomic region. "Sufficient homology" indicates that two polynucleotide sequences have sufficient structural similarity to act as substrates for a homologous recombination reaction. The structural similarity includes overall length of each polynucleotide fragment, as well as the sequence similarity of the polynucleotides. Sequence similarity can be described by the percent sequence identity over the whole length of the sequences, and/or by conserved regions comprising localized similarities such as contiguous nucleotides having 100% sequence identity, and percent sequence identity over a portion of the length of the sequences.

[0148] As used herein, a "genomic region" is a segment of a chromosome in the genome of a cell that is present on either side of the target site or, alternatively, also comprises a portion of the target site. The genomic region can comprise at least 5-10, 5-15, 5-20, 5-25, 5-30, 5-35, 5-40, 5-45, 5-

50, 5-55, 5-60, 5-65, 5- 70, 5-75, 5-80, 5-85, 5-90, 5-95, 5-100, 5-200, 5-300, 5-400, 5-500, 5-600, 5-700, 5-800, 5-900, 5-1000, 5-1100, 5-1200, 5-1300, 5-1400, 5-1500, 5-1600, 5-1700, 5-1800, 5-1900, 5-2000, 5-2100, 5-2200, 5-2300, 5-2400, 5-2500, 5-2600, 5-2700, 5-2800. 5-2900, 5-3000, 5-3100 or more bases such that the genomic region has sufficient homology to undergo homologous recombination with the corresponding region of homology.

[0149] As used herein, "homologous recombination" (HR) includes the exchange of DNA fragments between two DNA molecules at the sites of homology. The frequency of homologous recombination is influenced by a number of factors. Different organisms vary with respect to the amount of homologous recombination and the relative proportion of homologous to non-homologous recombination. Generally, the length of the region of homology affects the frequency of homologous recombination events: the longer the region of homology, the greater the frequency. The length of the homology region needed to observe homologous recombination is also species-variable. In many cases, at least 5 kb of homology has been utilized, but homologous recombination has been observed with as little as 25-50 bp of homology. See, for example, Singer *et al.*, (1982) Cell 31:25-33; Shen and Huang, (1986) Genetics 112:441-57; Watt *et al.*, (1985) *Proc. Natl. Acad. Sci. USA* 82:4768-72, Sugawara and Haber, (1992) *Mol Cell Biol* 12:563-75, Rubnitz and Subramani, (1984) *Mol Cell Biol* 4:2253-8; Ayares *et al.*, (1986) *Proc. Natl. Acad. Sci. USA* 83:5199-203; Liskay *et al.*, (1987) *Genetics* 115:161-7.

[0150] "Sequence identity" or "identity" in the context of nucleic acid or polypeptide sequences refers to the nucleic acid bases or amino acid residues in two sequences that are the same when aligned for maximum correspondence over a specified comparison window.

[0151] The term "percentage of sequence identity" refers to the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the results by 100 to yield the percentage of sequence identity. Useful examples of percent sequence identities include, but are not limited to, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%, or any

percentage from 50% to 100%. These identities can be determined using any of the programs described herein.

[0152] Sequence alignments and percent identity or similarity calculations may be determined using a variety of comparison methods designed to detect homologous sequences including, but not limited to, the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" will mean any set of values or parameters that originally load with the software when first initialized.

[0153] The "Clustal V method of alignment" corresponds to the alignment method labeled Clustal V (described by Higgins and Sharp, (1989) *CABIOS* 5:151-153; Higgins *et al.*, (1992) *Comput Appl Biosci* 8:189-191) and found in the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). For multiple alignments, the default values correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10. Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal method are KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids these parameters are KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. After alignment of the sequences using the Clustal V program, it is possible to obtain a "percent identity" by viewing the "sequence distances" Table in the same program. The "Clustal W method of alignment" corresponds to the alignment method labeled Clustal W (described by Higgins and Sharp, (1989) *CABIOS* 5:151-153; Higgins *et al.*, (1992) *Comput Appl Biosci* 8:189-191) and found in the MegAlign™ v6.1 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Default parameters for multiple alignment (GAP PENALTY=10, GAP LENGTH PENALTY=0.2, Delay Divergen Seqs (%)=30, DNA Transition Weight=0.5, Protein Weight Matrix=Gonnet Series, DNA Weight Matrix=IUB). After alignment of the sequences using the Clustal W program, it is possible to obtain a "percent identity" by viewing the "sequence distances" Table in the same program. Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using GAP Version 10 (GCG, Accelrys, San Diego, CA) using the following parameters:% identity and % similarity for a nucleotide sequence using a gap creation penalty weight of 50 and a gap length extension

penalty weight of 3, and the nwsgapdna.cmp scoring matrix; % identity and % similarity for an amino acid sequence using a GAP creation penalty weight of 8 and a gap length extension penalty of 2, and the BLOSUM62 scoring matrix (Henikoff and Henikoff, (1989) *Proc. Natl. Acad. Sci. USA* 89:10915). GAP uses the algorithm of Needleman and Wunsch, (1970) *J Mol Biol* 48:443-53, to find an alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP considers all possible alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps, using a gap creation penalty and a gap extension penalty in units of matched bases. "BLAST" is a searching algorithm provided by the National Center for Biotechnology Information (NCBI) used to find regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches to identify sequences having sufficient similarity to a query sequence such that the similarity would not be predicted to have occurred randomly. BLAST reports the identified sequences and their local alignment to the query sequence. It is well understood by one skilled in the art that many levels of sequence identity are useful in identifying polypeptides from other species or modified naturally or synthetically wherein such polypeptides have the same or similar function or activity. Useful examples of percent identities include, but are not limited to, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% or 95%, or any percentage from 50% to 100%. Indeed, any amino acid identity from 50% to 100% may be useful in describing the present disclosure, such as 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%.

[0154] Polynucleotide and polypeptide sequences, variants thereof, and the structural relationships of these sequences can be described by the terms "homology", "homologous", "substantially identical", "substantially similar" and "corresponding substantially" which are used interchangeably herein. These refer to polypeptide or nucleic acid sequences wherein changes in one or more amino acids or nucleotide bases do not affect the function of the molecule, such as the ability to mediate gene expression or to produce a certain phenotype. These terms also refer to modification(s) of nucleic acid sequences that do not substantially alter the functional properties of the resulting nucleic acid relative to the initial, unmodified nucleic acid. These modifications include deletion, substitution, and/or insertion of one or more nucleotides in the nucleic acid

fragment. Substantially similar nucleic acid sequences encompassed may be defined by their ability to hybridize (under moderately stringent conditions, e.g., 0.5X SSC, 0.1% SDS, 60°C) with the sequences exemplified herein, or to any portion of the nucleotide sequences disclosed herein and which are functionally equivalent to any of the nucleic acid sequences disclosed herein. Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions.

[0155] A "centimorgan" (cM) or "map unit" is the distance between two polynucleotide sequences, linked genes, markers, target sites, loci, or any pair thereof, wherein 1% of the products of meiosis are recombinant. Thus, a centimorgan is equivalent to a distance equal to a 1% average recombination frequency between the two linked genes, markers, target sites, loci, or any pair thereof.

[0156] An "isolated" or "purified" nucleic acid molecule, polynucleotide, polypeptide, or protein, or biologically active portion thereof, is substantially or essentially free from components that normally accompany or interact with the polynucleotide or protein as found in its naturally occurring environment. Thus, an isolated or purified polynucleotide or polypeptide or protein is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. Optimally, an "isolated" polynucleotide is free of sequences (optimally protein encoding sequences) that naturally flank the polynucleotide (i.e., sequences located at the 5' and 3' ends of the polynucleotide) in the genomic DNA of the organism from which the polynucleotide is derived. For example, in various aspects, the isolated polynucleotide can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotide sequence that naturally flank the polynucleotide in genomic DNA of the cell from which the polynucleotide is derived. Isolated polynucleotides may be purified from a cell in which they naturally occur. Conventional nucleic acid purification methods known to skilled artisans may be used to obtain isolated polynucleotides. The term also embraces recombinant polynucleotides and chemically synthesized polynucleotides.

[0157] The term “fragment” refers to a contiguous set of nucleotides or amino acids. In one aspect, a fragment is 2, 3, 4, 5, 6, 7 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or greater than 20 contiguous nucleotides. In one aspect, a fragment is 2, 3, 4, 5, 6, 7 8, 9, 10, 11, 12, 13, 14, 15, 16,

17, 18, 19, 20, or greater than 20 contiguous amino acids. A fragment may or may not exhibit the function of a sequence sharing some percent identity over the length of said fragment.

[0158] The terms "fragment that is functionally equivalent" and "functionally equivalent fragment" are used interchangeably herein. These terms refer to a portion or subsequence of an isolated nucleic acid fragment or polypeptide that displays the same activity or function as the longer sequence from which it derives. In one example, the fragment retains the ability to alter gene expression or produce a certain phenotype whether or not the fragment encodes an active protein. For example, the fragment can be used in the design of genes to produce the desired phenotype in a modified plant. Genes can be designed for use in suppression by linking a nucleic acid fragment, whether or not it encodes an active enzyme, in the sense or antisense orientation relative to a plant promoter sequence.

[0159] "Gene" includes a nucleic acid fragment that expresses a functional molecule such as, but not limited to, a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in its natural endogenous location with its own regulatory sequences.

[0160] By the term "endogenous" it is meant a sequence or other molecule that naturally occurs in a cell or organism. In some aspects,, an endogenous polynucleotide is normally found in the genome of a cell; that is, not heterologous.

[0161] An "allele" is one of several alternative forms of a gene occupying a given locus on a chromosome. When all the alleles present at a given locus on a chromosome are the same, that plant is homozygous at that locus. If the alleles present at a given locus on a chromosome differ, that plant is heterozygous at that locus.

[0162] "Coding sequence" refers to a polynucleotide sequence which codes for a specific amino acid sequence. "Regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences include, but are not limited to, promoters, translation leader sequences, 5' untranslated sequences, 3' untranslated sequences, introns, polyadenylation target sequences, RNA processing sites, effector binding sites, and stem-loop structures.

[0163] A "mutated gene" is a gene that has been altered through human intervention. Such a "mutated gene" has a sequence that differs from the sequence of the corresponding non-mutated

gene by at least one nucleotide addition, deletion, or substitution. In certain aspects of the disclosure, the mutated gene comprises an alteration that results from a guide polynucleotide/Cas polypeptide system as disclosed herein. A mutated plant is a plant comprising a mutated gene.

[0164] As used herein, a "targeted mutation" is a mutation in a gene (referred to as the target gene), including a native gene, that was made by altering a target sequence within the target gene using any method known to one skilled in the art, including a method involving a guided Cas polypeptide system as disclosed herein.

[0165] The terms "knock-out", "gene knock-out" and "genetic knock-out" are used interchangeably herein. A knock-out represents a DNA sequence of a cell that has been rendered partially or completely inoperative by targeting with a Cas polypeptide; for example, a DNA sequence prior to knock-out could have encoded an amino acid sequence, or could have had a regulatory function (e.g., promoter).

[0166] The terms "knock-in", "gene knock-in, "gene insertion" and "genetic knock-in" are used interchangeably herein. A knock-in represents the replacement or insertion of a DNA sequence at a specific DNA sequence in cell by targeting with a Cas polypeptide (for example by homologous recombination (HR), wherein a suitable donor DNA polynucleotide is also used). examples of knock-ins are a specific insertion of a heterologous amino acid coding sequence in a coding region of a gene, or a specific insertion of a transcriptional regulatory element in a genetic locus.

[0167] By "domain" it is meant a contiguous stretch of nucleotides (that can be RNA, DNA, and/or RNA-DNA-combination sequence) or amino acids.

[0168] The term "conserved domain" or "motif" means a set of polynucleotides or amino acids conserved at specific positions along an aligned sequence of evolutionarily related proteins. While amino acids at other positions can vary between homologous proteins, amino acids that are highly conserved at specific positions indicate amino acids that are essential to the structure, the stability, or the activity of a protein. Because they are identified by their high degree of conservation in aligned sequences of a family of protein homologues, they can be used as identifiers, or "signatures", to determine if a protein with a newly determined sequence belongs to a previously identified protein family.

[0169] A "codon-modified gene" or "codon-preferred gene" or "codon-optimized gene" is a gene having its frequency of codon usage designed to mimic the frequency of preferred codon usage of the host cell.

[0170] An “optimized” polynucleotide is a sequence that has been optimized for improved expression in a particular heterologous host cell.

[0171] A “plant-optimized nucleotide sequence" is a nucleotide sequence that has been optimized for expression in plants, particularly for increased expression in plants. A plant-optimized nucleotide sequence includes a codon-optimized gene. A plant-optimized nucleotide sequence can be synthesized by modifying a nucleotide sequence encoding a protein such as, for example, a Cas polypeptide as disclosed herein, using one or more plant-preferred codons for improved expression. *See*, for example, Campbell and Gowri (1990) *Plant Physiol.* 92:1-11 for a discussion of host-preferred codon usage.

[0172] A “promoter” is a region of DNA involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. An “enhancer” is a DNA sequence that can stimulate promoter activity, and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, and/or comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of some variation may have identical promoter activity.

[0173] Promoters that cause a gene to be expressed in most cell types at most times are commonly referred to as “constitutive promoters”. The term “inducible promoter” refers to a promoter that selectively express a coding sequence or functional RNA in response to the presence of an endogenous or exogenous stimulus, for example by chemical compounds (chemical inducers) or in response to environmental, hormonal, chemical, and/or developmental signals. Inducible or regulated promoters include, for example, promoters induced or regulated by light, heat, stress, flooding or drought, salt stress, osmotic stress, phytohormones, wounding, or chemicals such as ethanol, abscisic acid (ABA), jasmonate, salicylic acid, or safeners.

[0174] “Translation leader sequence” refers to a polynucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present

in the mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (e.g., Turner and Foster, (1995) *Mol Biotechnol* 3:225-236).

[0175] "3' non-coding sequences", "transcription terminator" or "termination sequences" refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht *et al.*, (1989) *Plant Cell* 1:671-680.

[0176] "RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complimentary copy of the DNA sequence, it is referred to as the primary transcript or pre-mRNA. A RNA transcript is referred to as the mature RNA or mRNA when it is a RNA sequence derived from post-transcriptional processing of the primary transcript pre-mRNA. "Messenger RNA" or "mRNA" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a DNA that is complementary to, and synthesized from, an mRNA template using the enzyme reverse transcriptase. The cDNA can be single-stranded or converted into double-stranded form using the Klenow fragment of DNA polymerase I. "Sense" RNA refers to RNA transcript that includes the mRNA and can be translated into protein within a cell or *in vitro*. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA, and that blocks the expression of a target gene (see, e.g., U.S. Patent No. 5,107,065). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes. The terms "complement" and "reverse complement" are used interchangeably herein with respect to mRNA transcripts, and are meant to define the antisense RNA of the message.

[0177] The term "genome" refers to the entire complement of genetic material (genes and non-coding sequences) that is present in each cell of an organism, or virus or organelle; and/or a complete set of chromosomes inherited as a (haploid) unit from one parent.

[0178] The term “operably linked” refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is regulated by the other. For example, a promoter is operably linked with a coding sequence when it is capable of regulating the expression of that coding sequence (i.e., the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in a sense or antisense orientation. In another example, the complementary RNA regions can be operably linked, either directly or indirectly, 5’ to the target mRNA, or 3’ to the target mRNA, or within the target mRNA, or a first complementary region is 5’ and its complement is 3’ to the target mRNA.

[0179] Generally, “host” refers to an organism or cell into which a heterologous component (polynucleotide, polypeptide, other molecule, cell) has been introduced. As used herein, a "host cell" refers to an *in vivo* or *in vitro* eukaryotic cell, prokaryotic cell (e.g., bacterial or archaeal cell), or cell from a multicellular organism (e.g., a cell line) cultured as a unicellular entity, into which a heterologous polynucleotide or polypeptide has been introduced. In some aspects, the cell is selected from the group consisting of: an archaeal cell, a bacterial cell, a eukaryotic cell, a eukaryotic single-cell organism, a somatic cell, a germ cell, a stem cell, a plant cell, an algal cell, an animal cell, in invertebrate cell, a vertebrate cell, a fish cell, a frog cell, a bird cell, an insect cell, a mammalian cell, a pig cell, a cow cell, a goat cell, a sheep cell, a rodent cell, a rat cell, a mouse cell, a non-human primate cell, and a human cell. In some cases, the cell is *in vitro*. In some cases, the cell is *in vivo*.

[0180] The term “recombinant” refers to an artificial combination of two otherwise separated segments of sequence, e.g., by chemical synthesis, or manipulation of isolated segments of nucleic acids by genetic engineering techniques.

[0181] The terms “plasmid”, “vector” and “cassette” refer to a linear or circular extra chromosomal element often carrying genes that are not part of the central metabolism of the cell, and usually in the form of double-stranded DNA. Such elements may be autonomously replicating sequences, genome integrating sequences, phage, or nucleotide sequences, in linear or circular form, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is

capable of introducing a polynucleotide of interest into a cell. "Transformation cassette" refers to a specific vector comprising a gene and having elements in addition to the gene that facilitates transformation of a particular host cell. "Expression cassette" refers to a specific vector comprising a gene and having elements in addition to the gene that allow for expression of that gene in a host.

[0182] The terms "recombinant DNA molecule", "recombinant DNA construct", "expression construct", "construct", and "recombinant construct" are used interchangeably herein. A recombinant DNA construct comprises an artificial combination of nucleic acid sequences, e.g., regulatory and coding sequences that are not all found together in nature. For example, a recombinant DNA construct may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. Such a construct may be used by itself or may be used in conjunction with a vector. If a vector is used, then the choice of vector is dependent upon the method that will be used to introduce the vector into the host cells as is well known to those skilled in the art. For example, a plasmid vector can be used. The skilled artisan is well aware of the genetic elements that must be present on the vector in order to successfully transform, select and propagate host cells. The skilled artisan will also recognize that different independent transformation events may result in different levels and patterns of expression (Jones *et al.*, (1985) *EMBO J* 4:2411-2418; De Almeida *et al.*, (1989) *Mol Gen Genetics* 218:78-86), and thus that multiple events are typically screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished standard molecular biological, biochemical, and other assays including Southern analysis of DNA, Northern analysis of mRNA expression, PCR, real time quantitative PCR (qPCR), reverse transcription PCR (RT-PCR), immunoblotting analysis of protein expression, enzyme or activity assays, and/or phenotypic analysis.

[0183] The term "heterologous" refers to the difference between the original environment, location, or composition of a particular polynucleotide or polypeptide sequence and its current environment, location, or composition. Non-limiting examples include differences in taxonomic derivation (*e.g.*, a polynucleotide sequence obtained from *Zea mays* would be heterologous if inserted into the genome of an *Oryza sativa* plant, or of a different variety or cultivar of *Zea mays*; or a polynucleotide obtained from a bacterium was introduced into a cell of a plant), or sequence (*e.g.*, a polynucleotide sequence obtained from *Zea mays*, isolated, modified, and re-introduced

into a maize plant). As used herein, "heterologous" in reference to a sequence can refer to a sequence that originates from a different species, variety, foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, a promoter operably linked to a heterologous polynucleotide is from a species different from the species from which the polynucleotide was derived, or, if from the same/analogous species, one or both are substantially modified from their original form and/or genomic locus, or the promoter is not the native promoter for the operably linked polynucleotide. Alternatively, one or more regulatory region(s) and/or a polynucleotide provided herein may be entirely synthetic. In another example, a target polynucleotide for cleavage by a Cas polypeptide may be of a different organism than that of the Cas polypeptide. In another example, a Cas polypeptide and guide RNA may be introduced to a target polynucleotide with an additional polynucleotide that acts as a template or donor for insertion into the target polynucleotide, wherein the additional polynucleotide is heterologous to the target polynucleotide and/or the Cas polypeptide.

[0184] The term "expression", as used herein, refers to the production of a functional end-product (e.g., an mRNA, guide RNA, or a protein) in either precursor or mature form.

[0185] A "mature" protein refers to a post-translationally processed polypeptide (i.e., one from which any pre- or pro-peptides present in the primary translation product have been removed).

[0186] "Precursor" protein refers to the primary product of translation of mRNA (i.e., with pre- and pro-peptides still present). Pre- and pro-peptides may be but are not limited to intracellular localization signals.

[0187] "CRISPR" (Clustered Regularly Interspaced Short Palindromic Repeats) loci refers to certain genetic loci encoding components of DNA cleavage systems, for example, used by bacterial and archaeal cells to destroy foreign DNA (Horvath and Barrangou, 2010, *Science* 327:167-170; WO2007025097, published 01 March 2007). A CRISPR locus can consist of a CRISPR array, comprising short direct repeats (CRISPR repeats) separated by short variable DNA sequences (called spacers), which can be flanked by diverse Cas (CRISPR-associated) genes.

[0188] As used herein, an "effector" or "effector protein" is a protein that encompasses an activity including recognizing, binding to, and/or cleaving or nicking a polynucleotide target. An effector, or effector protein, may also be an endonuclease. The "effector complex" of a CRISPR system includes Cas polypeptides involved in crRNA and target recognition and binding. Some of the

component Cas polypeptides may additionally comprise domains involved in target polynucleotide cleavage.

[0189] The term "Cas polypeptide" refers to a polypeptide encoded by a Cas (CRISPR-associated) gene. A Cas polypeptide includes proteins encoded by a gene in a cas locus, and include adaptation molecules as well as interference molecules. An interference molecule of a bacterial adaptive immunity complex includes endonucleases. A Cas endonuclease described herein comprises one or more nuclease domains. A Cas endonuclease includes but is not limited to: the novel Cas-alpha polypeptides disclosed herein, a Cas9 protein, a Cpf1 (Cas12) protein, a C2c1 protein, a C2c2 protein, a C2c3 protein, Cas3, Cas3-HD, Cas 5, Cas7, Cas8, Cas10, or combinations or complexes of these. A Cas polypeptide may be a "Cas endonuclease" or "Cas effector protein", that when in complex with a suitable polynucleotide component, is capable of recognizing, binding to, and optionally nicking or cleaving all or part of a specific polynucleotide target sequence. The Cas-alpha endonucleases of the disclosure include those having one or more RuvC nuclease domains. A Cas polypeptide is further defined as a functional fragment or functional variant of a native Cas polypeptide, or a protein that shares at least 50%, between 50% and 55%, at least 55%, between 55% and 60%, at least 60%, between 60% and 65%, at least 65%, between 65% and 70%, at least 70%, between 70% and 75%, at least 75%, between 75% and 80%, at least 80%, between 80% and 85%, at least 85%, between 85% and 90%, at least 90%, between 90% and 95%, at least 95%, between 95% and 96%, at least 96%, between 96% and 97%, at least 97%, between 97% and 98%, at least 98%, between 98% and 99%, at least 99%, between 99% and 100%, or 100% sequence identity with at least 50, between 50 and 100, at least 100, between 100 and 150, at least 150, between 150 and 200, at least 200, between 200 and 250, at least 250, between 250 and 300, at least 300, between 300 and 350, at least 350, between 350 and 400, at least 400, between 400 and 450, at least 500, or greater than 500 contiguous amino acids of a native Cas polypeptide, and retains at least partial activity of the native sequence.

[0190] A "functional fragment" of a Cas polypeptide refers to a portion or subsequence of the Cas polypeptide of the present disclosure in which the ability to recognize, bind to, and optionally unwind, nick or cleave (introduce a single or double-strand break in) the target site is retained. The portion or subsequence of the Cas polypeptide can comprise a complete or partial (functional) peptide of any one of its domains such as for example, but not limiting to a complete of functional

part of a Cas3 HD domain, a complete of functional part of a Cas3 Helicase domain, complete of functional part of a protein (such as but not limiting to a Cas5, Cas5d, Cas7 and Cas8b1).

[0191] The term "functional variant" of a Cas polypeptide or Cas effector protein refers to a variant of the Cas effector protein disclosed herein in which the ability to recognize, bind to, and optionally unwind, nick or cleave all or part of a target sequence is retained.

[0192] A Cas endonuclease may also include a multifunctional Cas endonuclease. The term "multifunctional Cas endonuclease" and "multifunctional Cas endonuclease polypeptide" are used interchangeably herein and includes reference to a single polypeptide that has Cas endonuclease functionality (comprising at least one protein domain that can act as a Cas endonuclease) and at least one other functionality, such as but not limited to, the functionality to form a complex (comprises at least a second protein domain that can form a complex with other proteins). In some aspects,, the multifunctional Cas endonuclease comprises at least one additional protein domain relative (either internally, upstream (5'), downstream (3'), or both internally 5' and 3', or any combination thereof) to those domains typical of a Cas endonuclease.

[0193] The terms "cascade" and "cascade complex" are used interchangeably herein and include reference to a multi-subunit protein complex that can assemble with a polynucleotide forming a polynucleotide-protein complex (PNP). Cascade is a PNP that relies on the polynucleotide for complex assembly and stability, and for the identification of target nucleic acid sequences. Cascade functions as a surveillance complex that finds and optionally binds target nucleic acids that are complementary to a variable targeting domain of the guide polynucleotide.

[0194]

[0195] The terms "5'-cap" and "7-methylguanylate (m7G) cap" are used interchangeably herein. A 7-methylguanylate residue is located on the 5′ terminus of messenger RNA (mRNA) in eukaryotes. RNA polymerase II (Pol II) transcribes mRNA in eukaryotes. Messenger RNA capping occurs generally as follows: the most terminal 5' phosphate group of the mRNA transcript is removed by RNA terminal phosphatase, leaving two terminal phosphates. A guanosine monophosphate (GMP) is added to the terminal phosphate of the transcript by a guanylyl transferase, leaving a 5′-5′ triphosphate-linked guanine at the transcript terminus. Finally, the 7-nitrogen of this terminal guanine is methylated by a methyl transferase.

[0196] The terminology "not having a 5'-cap" herein is used to refer to RNA having, for example, a 5'-hydroxyl group instead of a 5'-cap. Such RNA can be referred to as "uncapped RNA", for

example. Uncapped RNA can better accumulate in the nucleus following transcription, since 5'-capped RNA is subject to nuclear export. One or more RNA components herein are uncapped.

[0197] As used herein, the term "guide polynucleotide", relates to a polynucleotide sequence that can form a complex with a Cas polypeptide, including the Cas polypeptides described herein, and enables the Cas polypeptide to recognize, optionally bind to, and optionally cleave a DNA target site. The guide polynucleotide sequence can be a RNA sequence, a DNA sequence, or a combination thereof (a RNA-DNA combination sequence).

[0198] The term "functional fragment" of a guide RNA, crRNA or tracrRNA are used interchangeably herein, and refer to a portion or subsequence of the guide RNA, crRNA or tracrRNA, respectively, of the present disclosure in which the ability to function as a guide RNA, crRNA or tracrRNA, respectively, is retained.

[0199] The terms "functional variant" of a guide RNA, crRNA or tracrRNA (respectively) are used interchangeably herein, and refer to a variant of the guide RNA, crRNA or tracrRNA, respectively, of the present disclosure in which the ability to function as a guide RNA, crRNA or tracrRNA, respectively, is retained.

[0200] The terms "single guide RNA" and "sgRNA" are used interchangeably herein and relate to a synthetic fusion of two RNA molecules, a crRNA (CRISPR RNA) comprising a variable targeting domain (linked to a tracr mate sequence that hybridizes to a tracrRNA), fused to a tracrRNA (trans-activating CRISPR RNA). The single guide RNA can comprise a crRNA or crRNA fragment and a tracrRNA or tracrRNA fragment of the type II CRISPR/Cas system that can form a complex with a type II Cas polypeptide, wherein said guide RNA/Cas polypeptide complex can direct the Cas polypeptide to a DNA target site, enabling the Cas polypeptide to recognize, optionally bind to, and optionally nick or cleave (introduce a single or double-strand break) the DNA target site.

[0201] The term "variable targeting domain" or "VT domain" is used interchangeably herein and includes a nucleotide sequence that can hybridize (is complementary) to one strand (nucleotide sequence) of a double strand DNA target site. The percent complementation between the first nucleotide sequence domain (VT domain) and the target sequence can be at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 63%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100%. The variable

targeting domain can be at least 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30 nucleotides in length. In some aspects, the variable targeting domain comprises a contiguous stretch of 12 to 30 nucleotides. The variable targeting domain can be composed of a DNA sequence, a RNA sequence, a modified DNA sequence, a modified RNA sequence, or any combination thereof.

[0202] The term "Cas endonuclease recognition domain" or "CER domain" (of a guide polynucleotide) is used interchangeably herein and includes a nucleotide sequence that interacts with a Cas polypeptide. A CER domain comprises a (trans-acting) tracrNucleotide mate sequence followed by a tracrNucleotide sequence. The CER domain can be composed of a DNA sequence, a RNA sequence, a modified DNA sequence, a modified RNA sequence (see for example US20150059010A1, published 26 February 2015), or any combination thereof.

[0203] As used herein, the terms "guide polynucleotide/Cas polypeptide complex", "guide polynucleotide/Cas polypeptide system", "guide polynucleotide/Cas complex", "guide polynucleotide/Cas system", "guided Cas system", "Polynucleotide-guided endonuclease", "PGEN" are used interchangeably herein and refer to at least one guide polynucleotide and at least one Cas polypeptide, that are capable of forming a complex, wherein said guide polynucleotide/Cas polypeptide complex can direct the Cas polypeptide to a DNA target site, enabling the Cas polypeptide to recognize, bind to, and optionally nick or cleave (introduce a single or double-strand break) the DNA target site. A guide polynucleotide/Cas polypeptide complex herein can comprise Cas polypeptide(s) and suitable polynucleotide component(s) of any of the known CRISPR systems (Horvath and Barrangou, 2010, *Science* 327:167-170; Makarova *et al.* 2015, *Nature Reviews Microbiology* Vol. 13:1-15; Zetsche *et al.*, 2015, Cell 163, 1-13; Shmakov *et al.*, 2015, *Molecular Cell* 60, 1-13).

[0204] The terms "guide RNA/Cas polypeptide complex", "guide RNA/Cas polypeptide system", "guide RNA/Cas complex", "guide RNA/Cas system", "gRNA/Cas complex", "gRNA/Cas system", "RNA-guided endonuclease" , and "RGEN" are used interchangeably herein and refer to at least one RNA component and at least one Cas polypeptide that are capable of forming a complex, wherein said guide RNA/Cas polypeptide complex can direct the Cas polypeptide to a DNA target site, enabling the Cas polypeptide to recognize, bind to, and optionally nick or cleave (introduce a single or double-strand break) the DNA target site.

[0205] The terms "target site", "target sequence", "target site sequence, "target DNA", "target locus", "genomic target site", "genomic target sequence", "genomic target locus" and "protospacer", are used interchangeably herein and refer to a polynucleotide sequence such as, but not limited to, a nucleotide sequence on a chromosome, episome, a locus, or any other DNA molecule in the genome (including chromosomal, chloroplastic, mitochondrial DNA, plasmid DNA) of a cell, at which a guide polynucleotide/Cas polypeptide complex can recognize, bind to, and optionally nick or cleave . The target site can be an endogenous site in the genome of a cell, or alternatively, the target site can be heterologous to the cell and thereby not be naturally occurring in the genome of the cell, or the target site can be found in a heterologous genomic location compared to where it occurs in nature. As used herein, terms "endogenous target sequence" and "native target sequence" are used interchangeable herein to refer to a target sequence that is endogenous or native to the genome of a cell and is at the endogenous or native position of that target sequence in the genome of the cell. An "artificial target site" or "artificial target sequence" are used interchangeably herein and refer to a target sequence that has been introduced into the genome of a cell. Such an artificial target sequence can be identical in sequence to an endogenous or native target sequence in the genome of a cell but be located in a different position (*i.e.*, a non-endogenous or non-native position) in the genome of a cell.

[0206] A "protospacer adjacent motif" (PAM) herein refers to a short nucleotide sequence adjacent to a target sequence (protospacer) that is recognized (targeted) by a guide polynucleotide/Cas polypeptide system described herein. The Cas polypeptide may not successfully recognize a target DNA sequence if the target DNA sequence is not followed by a PAM sequence. The sequence and length of a PAM herein can differ depending on the Cas polypeptide or Cas polypeptide complex used. The PAM sequence can be of any length but is typically 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 nucleotides long.

[0207] An "altered target site", "altered target sequence", "modified target site", "modified target sequence" are used interchangeably herein and refer to a target sequence as disclosed herein that comprises at least one alteration when compared to non-altered target sequence. Such "alterations" include, for example: (i) replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, (iv) a chemical alteration of at least one nucleotide, or (v) any combination of (i) – (iv).

[0208] A “modified nucleotide” or “edited nucleotide” refers to a nucleotide sequence of interest that comprises at least one alteration when compared to its non-modified nucleotide sequence. Such “alterations” include, for example: (i) replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, (iv) a chemical alteration of at least one nucleotide, or (v) any combination of (i) – (iv).

[0209] Methods for “modifying a target site” and “altering a target site” are used interchangeably herein and refer to methods for producing an altered target site.

[0210] As used herein, “donor DNA” is a DNA construct that comprises a polynucleotide of interest to be inserted into a genomic target site by homology-directed repair.

[0211] The term “polynucleotide modification template” includes a polynucleotide that comprises at least one nucleotide modification when compared to the nucleotide sequence to be edited. A nucleotide modification can be at least one nucleotide substitution, addition or deletion. Optionally, the polynucleotide modification template can further comprise homologous nucleotide sequences flanking the at least one nucleotide modification, wherein the flanking homologous nucleotide sequences provide sufficient homology to the desired nucleotide sequence to be edited.

[0212] The term “plant-optimized Cas polypeptide” herein refers to a Cas polypeptide, including a multifunctional Cas polypeptide, encoded by a nucleotide sequence that has been optimized for expression in a plant cell or plant.

[0213] A “plant-optimized nucleotide sequence encoding a Cas polypeptide”, “plant-optimized construct encoding a Cas polypeptide” and a “plant-optimized polynucleotide encoding a Cas polypeptide” are used interchangeably herein and refer to a nucleotide sequence encoding a Cas polypeptide, or a variant or functional fragment thereof, that has been optimized for expression in a plant cell or plant. A plant comprising a plant-optimized Cas polypeptide includes a plant comprising the nucleotide sequence encoding for the Cas sequence and/or a plant comprising the Cas polypeptide. In some aspects,, the plant-optimized Cas polypeptide nucleotide sequence is a maize-optimized, rice-optimized, wheat-optimized, soybean-optimized, cotton-optimized, or canola-optimized Cas polypeptide.

[0214] The term “plant” generically includes whole plants, plant organs, plant tissues, seeds, plant cells, seeds and progeny of the same. The plant is a monocot or dicot. Plant cells include, without limitation, cells from seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen and microspores. A "plant element" is

intended to reference either a whole plant or a plant component, which may comprise differentiated and/or undifferentiated tissues, for example but not limited to plant tissues, parts, and cell types. In one aspect, a plant element is one of the following: whole plant, seedling, meristematic tissue, ground tissue, vascular tissue, dermal tissue, seed, leaf, root, shoot, stem, flower, fruit, stolon, bulb, tuber, corm, keiki, shoot, bud, tumor tissue, and various forms of cells and culture (e.g., single cells, protoplasts, embryos, callus tissue). It should be noted that a protoplast is not technically an "intact" plant cell (as naturally found with all components), as protoplasts lack a cell wall. The term "plant organ" refers to plant tissue or a group of tissues that constitute a morphologically and functionally distinct part of a plant. As used herein, a "plant element" is synonymous to a "portion" of a plant, and refers to any part of the plant, and can include distinct tissues and/or organs, and may be used interchangeably with the term "tissue" throughout. Similarly, a "plant reproductive element" is intended to generically reference any part of a plant that is able to initiate other plants via either sexual or asexual reproduction of that plant, for example but not limited to: seed, seedling, root, shoot, cutting, scion, graft, stolon, bulb, tuber, corm, keiki, or bud. The plant element may be in plant or in a plant organ, tissue culture, or cell culture.

[0215] "Progeny" comprises any subsequent generation of a plant.

[0216] As used herein, the term "plant part" refers to plant cells, plant protoplasts, plant cell tissue cultures from which plants can be regenerated, plant calli, plant clumps, and plant cells that are intact in plants or parts of plants such as embryos, pollen, ovules, seeds, leaves, flowers, branches, fruit, kernels, ears, cobs, husks, stalks, roots, root tips, anthers, and the like, as well as the parts themselves. Grain is intended to mean the mature seed produced by commercial growers for purposes other than growing or reproducing the species. Progeny, variants, and mutants of the regenerated plants are also included within the scope of the disclosure, provided that these parts comprise the introduced polynucleotides.

[0217] The term "monocotyledonous" or "monocot" refers to the subclass of angiosperm plants also known as "monocotyledoneae", whose seeds typically comprise only one embryonic leaf, or cotyledon. The term includes references to whole plants, plant elements, plant organs (e.g., leaves, stems, roots, etc.), seeds, plant cells, and progeny of the same.

[0218] The term "dicotyledonous" or "dicot" refers to the subclass of angiosperm plants also knows as "dicotyledoneae", whose seeds typically comprise two embryonic leaves, or cotyledons.

The term includes references to whole plants, plant elements, plant organs (e.g., leaves, stems, roots, etc.), seeds, plant cells, and progeny of the same.

[0219] The term "non-conventional yeast" herein refers to any yeast that is not a *Saccharomyces* (e.g., *S. cerevisiae*) or *Schizosaccharomyces* yeast species. (see "Non-Conventional Yeasts in Genetics, Biochemistry and Biotechnology:Practical Protocols", K. Wolf, K.D. Breunig, G. Barth, Eds., Springer-Verlag, Berlin, Germany, 2003).

[0220] The term "crossed" or "cross" or "crossing" in the context of this disclosure means the fusion of gametes via pollination to produce progeny (i.e., cells, seeds, or plants). The term encompasses both sexual crosses (the pollination of one plant by another) and selfing (self-pollination, i.e., when the pollen and ovule (or microspores and megaspores) are from the same plant or genetically identical plants).

[0221] The term "introgression" refers to the transmission of a desired allele of a genetic locus from one genetic background to another. For example, introgression of a desired allele at a specified locus can be transmitted to at least one progeny plant via a sexual cross between two parent plants, where at least one of the parent plants has the desired allele within its genome. Alternatively, for example, transmission of an allele can occur by recombination between two donor genomes, e.g., in a fused protoplast, where at least one of the donor protoplasts has the desired allele in its genome. The desired allele can be, e.g., a transgene, a modified (mutated or edited) native allele, or a selected allele of a marker or QTL.

[0222] "Introducing" is intended to mean presenting to a target, such as a cell or organism, a polynucleotide or polypeptide or polynucleotide-protein complex, in such a manner that the component(s) gains access to the interior of a cell of the organism or to the cell itself.

[0223] A "polynucleotide of interest" includes any nucleotide sequence encoding a protein or polypeptide that improves desirability of crops, i.e. a trait of agronomic interest. Polynucleotides of interest include, but are not limited to: polynucleotides encoding important traits for agronomics, herbicide-resistance, insecticidal resistance, disease resistance, nematode resistance, herbicide resistance, microbial resistance, fungal resistance, viral resistance, fertility or sterility, grain characteristics, commercial products, phenotypic marker, or any other trait of agronomic or commercial importance. A polynucleotide of interest may additionally be utilized in either the sense or anti-sense orientation. Further, more than one polynucleotide of interest may be utilized together, or "stacked", to provide additional benefit.

[0224] A “complex trait locus” includes a genomic locus that has multiple transgenes genetically linked to each other.

[0225] The terms "decreased," "fewer," "slower" and "increased" "faster" "enhanced" "greater" as used herein refers to a decrease or increase in a characteristic of the modified plant element or resulting plant compared to an unmodified plant element or resulting plant. For example, a decrease in a characteristic may be at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, between 5% and 10%, at least 10%, between 10% and 20%, at least 15%, at least 20%, between 20% and 30%, at least 25%, at least 30%, between 30% and 40%, at least 35%, at least 40%, between 40% and 50%, at least 45%, at least 50%, between 50% and 60%, at least about 60%, between 60% and 70%, between 70% and 80%, at least 75%, at least about 80%, between 80% and 90%, at least about 90%, between 90% and 100%, at least 100%, between 100% and 200%, at least 200%, at least about 300%, at least about 400%) or more lower than the untreated control and an increase may be at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, between 5% and 10%, at least 10%, between 10% and 20%, at least 15%, at least 20%, between 20% and 30%, at least 25%, at least 30%, between 30% and 40%, at least 35%, at least 40%, between 40% and 50%, at least 45%, at least 50%, between 50% and 60%, at least about 60%, between 60% and 70%, between 70% and 80%, at least 75%, at least about 80%, between 80% and 90%, at least about 90%, between 90% and 100%, at least 100%, between 100% and 200%, at least 200%, at least about 300%, at least about 400% or more higher than the untreated control.

[0226] As used herein, the term “before”, in reference to a sequence position, refers to an occurrence of one sequence upstream, or 5’, to another sequence.

[0227] The meaning of abbreviations is as follows: “sec” means second(s), “min” means minute(s), “h” means hour(s), “d” means day(s), “µL” means microliter(s), “mL” means milliliter(s), “L” means liter(s), “µM” means micromolar, “mM” means millimolar, “M” means molar, “mmol” means millimole(s), “µmole” or “umole” mean micromole(s), “g” means gram(s), “µg” or “ug” means microgram(s), “ng” means nanogram(s), “U” means unit(s), “bp” means base pair(s) and “kb” means kilobase(s).

Classification of CRISPR-Cas Systems

[0228] CRISPR-Cas systems have been classified according to sequence and structural analysis of components. Multiple CRISPR/Cas systems have been described including Class 1 systems, with multisubunit effector complexes (comprising type I, type III, and type IV), and Class 2

systems, with single protein effectors (comprising type II, type V, and type VI) (Makarova *et al.* 2015, *Nature Reviews Microbiology* Vol. 13:1-15; Zetsche *et al.*, 2015, *Cell* 163, 1-13; Shmakov *et al.*, 2015, *Molecular Cell* 60, 1-13; Haft *et al.*, 2005, *Computational Biology, PLoS Comput Biol* 1(6):e60; and Koonin *et al.* 2017, *Curr Opinion Microbiology* 37:67-78).

[0229] A CRISPR-Cas system comprises, at a minimum, a CRISPR RNA (crRNA) molecule and at least one CRISPR-associated (Cas) protein to form crRNA ribonucleoprotein (crRNP) effector complexes. CRISPR-Cas loci comprise an array of identical repeats interspersed with DNA-targeting spacers that encode the crRNA components and an operon-like unit of *cas* genes encoding the Cas polypeptide components. The resulting ribonucleoprotein complex recognizes a polynucleotide in a sequence-specific manner (Jore *et al.*, *Nature Structural & Molecular Biology* 18, 529–536 (2011)). The crRNA serves as a guide RNA for sequence specific binding of the effector (protein or complex) to double strand DNA sequences, by forming base pairs with the complementary DNA strand while displacing the noncomplementary strand to form a so called R-loop. (Jore *et al.*, 2011. *Nature Structural & Molecular Biology* 18, 529–536).

[0230] RNA transcripts of CRISPR loci (pre-crRNA) are cleaved specifically in the repeat sequences by CRISPR associated (Cas) endoribonucleases in type I and type III systems or by RNase III in type II systems. The number of CRISPR-associated genes at a given CRISPR locus can vary between species.

[0231] Different *cas* genes that encode proteins with different domains are present in different CRISPR systems. The *cas* operon comprises genes that encode for one or more effector endonucleases, as well as other Cas polypeptides. Protein subunits include those described in Makarova *et al.* 2011, *Nat Rev Microbiol.* 2011 9(6):467–477; Makarova *et al.* 2015, *Nature Reviews Microbiology* Vol. 13:1-15; and Koonin *et al.* 2017, *Current Opinion Microbiology* 37:67-78). The types of domains include those involved in Expression (pre-crRNA processing, for example Cas 6 or RNaseIII), Interference (including an effector module for crRNA and target binding, as well as domain(s) for target cleavage), Adaptation (spacer insertion, for example Cas1 or Cas2), and Ancillary (regulation or helper or unknown function). Some domains may serve more than one purpose, for example Cas9 comprises domains for endonuclease functionality as well as for target cleavage, among others.

[0232] The Cas polypeptide is guided by a single CRISPR RNA (crRNA) through direct RNA-DNA base-pairing to recognize a DNA target site that is in close vicinity to a protospacer adjacent

motif (PAM) (Jore, M.M. *et al.*, 2011, *Nat. Struct. Mol. Biol.* 18:529-536, Westra, E.R. *et al.*, 2012, *Molecular Cell* 46:595-605, and Sinkunas, T. *et al.*, 2013, *EMBO J.* 32:385-394).

Class I CRISPR-Cas Systems

[0233] Class I CRISPR-Cas systems comprise Types I, III, and IV. A characteristic feature of Class I systems is the presence of an effector endonuclease complex instead of a single protein. A Cascade complex comprises a RNA recognition motif (RRM) and a nucleic acid-binding domain that is the core fold of the diverse RAMP (Repeat-Associated Mysterious Proteins) protein superfamily (Makarova *et al.* 2013, *Biochem Soc Trans* 41, 1392-1400; Makarova *et al.* 2015, *Nature Reviews Microbiology* Vol. 13:1-15). RAMP protein subunits include Cas5 and Cas7 (which comprise the skeleton of the crRNA–effector complex), wherein the Cas5 subunit binds the 5' handle of the crRNA and interacts with the large subunit, and often includes Cas6 which is loosely associated with the effector complex and typically functions as the repeat-specific RNase in the pre-crRNA processing (Charpentier *et al.*, *FEMS Microbiol Rev* 2015, 39:428-441; Niewoehner *et al.*, *RNA* 2016, 22:318-329).

[0234] Type I CRISPR-Cas systems comprise a complex of effector proteins, termed Cascade (CRISPR-associated complex for antiviral defense) comprising at a minimum Cas5 and Cas7. The effector complex functions together with a single CRISPR RNA (crRNA) and Cas3 to defend against invading viral DNA (Brouns, S.J.J. *et al. Science* 321:960-964; Makarova *et al.* 2015, *Nature Reviews Microbiology* Vol. 13:1-15). Type I CRISPR-Cas loci comprise the signature gene *cas3* (or a variant cas3' or *cas3''*), which encodes a metal-dependent nuclease that possesses a single-stranded DNA (ssDNA)-stimulated superfamily 2 helicase with a demonstrated capacity to unwind double stranded DNA (dsDNA) and RNA–DNA duplexes (Makarova *et al.* 2015, Nature Reviews; Microbiology Vol. 13:1-15). Following target recognition, the Cas3 endonuclease is recruited to the Cascade-crRNA-target DNA complex to cleave and degrade the DNA target (Westra, E.R. *et al.* (2012) *Molecular Cell* 46:595-605, Sinkunas, T. *et al.* (2011) *EMBO J.* 30:1335-1342, and Sinkunas, T. *et al.* (2013) *EMBO J.* 32:385-394). In some type I systems, Cas6 can be the active endonuclease that is responsible for crRNA processing, and Cas5 and Cas7 function as non-catalytic RNA-binding proteins; although in type I-C systems, crRNA processing can be catalyzed by Cas5 (Makarova *et al.* 2015, *Nature Reviews Microbiology* Vol. 13:1-15). Type I systems are divided into seven subtypes (Makarova *et al.* 2011, *Nat Rev Microbiol.* 2011 9(6):467–477; Koonin *et al.* 2017, *Curr Opinion Microbiology* 37:67-78). A modified type I

CRISPR-associated complex for adaptive antiviral defense (Cascade) comprising at least the protein subunits Cas7, Cas5 and Cas6, wherein one of these subunits is synthetically fused to a Cas3 endonuclease or a modified restriction endonuclease, FokI, have been described (WO2013098244 published 04 July 4 2013).

[0235] Type III CRISPR-Cas systems, comprising a plurality of *cas7* genes, target either ssRNA or ssDNA, and function as either an RNase as well as a target RNA-activated DNA nuclease (Tamulaitis *et al.*, *Trends in Microbiology* 25(10)49-61, 2017). Csm (Type III-A) and Cmr (Type III-B) complexes function as RNA-activated single-stranded (ss) DNases that couple the target RNA binding/cleavage with ssDNA degradation. Upon foreign DNA infection, the CRISPR RNA (crRNA)-guided binding of the Csm or Cmr complex to the emerging transcript recruits Cas10 DNase to the actively transcribed phage DNA, resulting in degradation of both the transcript and phage DNA, but not the host DNA. The Cas10 HD-domain is responsible for the ssDNase activity, and Csm3/Cmr4 subunits are responsible for the endoribonuclease activity of the Csm/Cmr complex. The 3′-flanking sequence of the target RNA is critical for the ssDNase activity of Csm/Cmr: the basepairing with the 5′-handle of crRNA protects host DNA from degradation.

[0236] Type IV systems, although comprising typical type I *cas5* and *cas7* domains in addition to a *cas8*-like domain, may lack the CRISPR array that is characteristic of most other CRISPR-Cas systems.

Class II CRISPR-Cas Systems

[0237] Class II CRISPR-Cas systems comprise Types II, V, and VI. A characteristic feature of Class II systems is the presence of a single Cas effector protein instead of an effector complex. Types II and V Cas polypeptides comprise an RuvC endonuclease domain that adopts the RNase H fold.

[0238] Type II CRISPR/Cas systems employ a crRNA and tracrRNA (trans-activating CRISPR RNA) to guide the Cas polypeptide to its DNA target. The crRNA comprises a spacer region complementary to one strand of the double strand DNA target and a region that base pairs with the tracrRNA (trans-activating CRISPR RNA) forming a RNA duplex that directs the Cas polypeptide to cleave the DNA target, leaving a blunt end. Spacers are acquired through a not fully understood process involving Cas1 and Cas2 proteins. Type II CRISPR/Cas loci typically comprise *cas1* and *cas2* genes in addition to the *cas9* gene (Chylinski *et al.*, 2013, *RNA Biology* 10:726-737; Makarova *et al.* 2015, *Nature Reviews Microbiology* Vol. 13:1-15). Type II CRISR-Cas loci can

encode a tracrRNA, which is partially complementary to the repeats within the respective CRISPR array, and can comprise other proteins such as Csn1 and Csn2. The presence of *cas9* in the vicinity of *cas1* and *cas2* genes is the hallmark of type II loci (Makarova *et al.* 2015, *Nature Reviews Microbiology* Vol. 13:1-15).

[0239] Type V CRISPR/Cas systems comprise a single Cas polypeptide, including Cpf1 (Cas12) (Koonin *et al.*, *Curr Opinion Microbiology* 37:67-78, 2017), that is an active RNA-guided endonuclease that does not necessarily require the additional trans-activating CRISPR (tracr) RNA for target cleavage, unlike Cas9.

[0240] Type VI CRISPR-Cas systems comprise a *cas13* gene that encodes a nuclease with two HEPN (Higher Eukaryotes and Prokaryotes Nucleotide-binding) domains but no HNH or RuvC domains, and are not dependent upon tracrRNA activity. The majority of HEPN domains comprise conserved motifs that constitute a metal-independent endoRNase active site (Anantharam *et al.*, *Biol Direct* 8:15, 2013). Because of this feature, it is thought that type VI systems act on RNA targets instead of the DNA targets that are common to other CRISPR-Cas systems.

[0241] In a first aspect, the disclosure provides methods for altering protospacer adjacent motif (PAM) specificity of a target Cas-alpha polypeptide. As used herein, a "protospacer adjacent motif" (PAM) refers to a short nucleotide sequence adjacent to a target sequence (protospacer) that can be recognized (targeted) by a guide polynucleotide/Cas polypeptide system. The Cas polypeptide may not successfully recognize a target DNA sequence if the target DNA sequence is not followed by a PAM sequence. The sequence and length of a PAM herein can differ depending on the Cas polypeptide or Cas polypeptide complex used. The PAM sequence can be of any length but is typically 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 nucleotides long.

[0242] In some aspects, a method for altering PAM specificity of a target Cas-alpha polypeptide comprises: (a) comparing the PAM interacting (PI) domain of a heterologous, orthologous Cas-alpha polypeptide with the PI domain of the target Cas-alpha polypeptide, wherein the orthologous Cas-alpha polypeptide has different PAM recognition than the target Cas-alpha polypeptide; (b) selecting one or more amino acids and/or one or more polypeptide chains from the PI domain of the orthologous Cas-alpha polypeptide; (c) incorporating the one or more amino acids and/or the one or more polypeptide chains selected from the PI domain of the orthologous Cas-alpha polypeptide into one or more structurally similar positions of the target Cas-alpha polypeptide

resulting in a modified target Cas-alpha polypeptide; and (d) determining PAM recognition in the modified target Cas-alpha polypeptide.

[0243] As used herein, an "orthologous Cas-alpha polypeptide" or "Cas-alpha ortholog" refers to a Cas-alpha polypeptide comprising a CRISPR-Cas polypeptide comprising at least one zinc-finger-like domain, at least one bridge-helix-like domain, a tri-split RuvC domain (comprising non-contiguous RuvC-I domain, RuvC-II domain, and RuvC-III domain) that ranges in size from about 327 to 777 amino acids, and contains a GxxxG, ExL, and Cx_nC or $Cx_n(C,H)$ motif (where x represents any amino acid and n = one or more amino acids).

[0244] Cas-alpha orthologs that can be used in the methods disclosed herein include Cas-alpha 1, Cas-alpha 2, Cas-alpha 3, Cas-alpha 4, Cas-alpha 5, Cas-alpha 6, Cas-alpha 7, Cas-alpha 8, Cas-alpha 10, Cas-alpha 11, Cas-alpha 13, Cas-alpha 24, and Cas-alpha 29.

[0245] **FIG. 1** illustrates the phylogenetic relationship among some of the Cas-alpha orthologs divided into three supergroups (I, II, and III). Group I comprises Clade 1 (Candidate Archaea and Aureabacteria (Cas1, Cas2, Cas4 typically encoded in the locus)). Group II comprises Clade 2 (Aquificae (Sulfurihydrogenibium and Hydrogenivirga genera) and Deltaproteobacteria (Desulfovibrio genus)), Clade 3 (Candidate Archaea (Cas1, Cas2, and Cas4 typically encoded in the locus)), Clade 4 (Bacteroidetes (Prevotella and Bacteroides genera)), Clade 5 (Candidate Levybacterium), and Clade 6 (Clostridia (Dorea, Ruminococcus, Clostridium, Clostridioides, Peptocolstridium, Cellulosilyticym, Eubacterium, Syntrophomonas genera)). Group III comprises Clade 7 (Bacilli (Bacillus, Acidibacillus, Aneurinibacillus, Brevibacillus, Parageobacillus, Alicyclobacillus genera)), Clade 8 (Negativicutes (Phascolarctobacterium genus)), and Clade 9 (Flavobacteriia (Flavobacterium genus)). A diamond symbol represents orthologous Cas-alphas 1-11 endonucleases.

[0246] As used herein, a "structurally similar position" refers to a coordinate in a polypeptide that occupies a similar three-dimensional location when its structure, either predicted (e.g. neural network-based informatic models such as AlphaFold) or determined (e.g. using cryogenic electron microscopy (Cryo-EM), X-ray crystallography, and NMR spectroscopy), is aligned or superimposed with an orthologous structure either predicted or determined.

[0247] Methods for comparing the PI domain of a target Cas-alpha polypeptide and an orthologous Cas-alpha polypeptide include, but are not limited to, multiple sequence comparison by log-expectation (MUSCLE), multiple sequence comparison using Clustal Omega, root mean square

distance (RMSD), distance matrix alignment (DALI), structural homology by environment-based alignment (SHEBA), combinatorial extension (CE), homologous structure alignment database (HOMSTRAD), structural classification of proteins (SCOP), FatCat, and PhyreStorm.

[0248] In some examples of the disclosed methods for altering PAM specificity of a target Cas-alpha polypeptide, the step of selecting one or more amino acids from the PI domain of an orthologous Cas-alpha polypeptide comprises comparing the PI domain sequence and/or structure of the orthologous Cas-alpha polypeptide with that from the target Cas-alpha polypeptide whose PAM specificity is to be altered and substituting one or more amino acids from the orthologous Cas-alpha polypeptide into the Cas-alpha polypeptide target.

[0249] In some examples of the disclosed methods for altering PAM specificity of a target Cas-alpha polypeptide, the step of selecting one or more polypeptide chains from the PI domain of an orthologous Cas-alpha polypeptide comprises comparing the PI domain sequence and/or structure of the orthologous Cas-alpha polypeptide with that from the target Cas-alpha polypeptide whose PAM specificity is to be altered and substituting one or more polypeptide chains from the orthologous Cas-alpha polypeptide into the Cas-alpha polypeptide target.

[0250] Methods for determining the PAM recognition of a modified target Cas-alpha polypeptide include, but are not limited to, transcription and translation of the modified Cas-alpha polypeptide in a cell or in a cell-free mixture, complexing the modified Cas-alpha polypeptide with a guide RNA to form a ribonucleoprotein (RNP), incubating the RNP with DNA species that contain a fixed guide RNA target and a collection of different PAM sequences, capturing DNA molecules that support target cleavage, sequencing the PAM region from DNA species that support target cleavage, and calculating a consensus PAM using a position frequency matrix to summarize alterations to PAM specificity.

[0251] In some examples of the disclosed methods for altering PAM specificity of a target Cas-alpha polypeptide, the target Cas-alpha polypeptide is a Cas-alpha 10 polypeptide. The target Cas-alpha polypeptide can be a Cas-alpha 10 polypeptide having an amino acid sequence with at least 50%, alternatively at least 55%, alternatively at least 60%, alternatively at least 65%, alternatively at least 70%, alternatively at least 75%, alternatively at least 80%, alternatively at least 85%, alternatively at least 90% alternatively at least 95%, alternatively at least 96%, alternatively at least 97%, alternatively at least 98%, alternatively at least 99%, or alternatively 100% sequence identity to SEQ ID NO: 2.

[0252] In another aspect, the disclosure provides synthetic or non-naturally occurring Cas-alpha 10 polypeptides having altered PAM specificity. More specifically, disclosed herein are Cas-alpha 10 polypeptides comprising modified PAM interacting (PI) domains such that the resulting Cas-alpha 10 polypeptides recognize a PAM sequence in a target polynucleotide other than 5'-TTC-3'.

[0253] In another aspect, the disclosure provides synthetic or non-naturally occurring Cas-alpha 10 polypeptides having altered PAM specificity that comprise an amino acid sequence having at least 50%, alternatively at least 55%, alternatively at least 60%, alternatively at least 65%, alternatively at least 70%, alternatively at least 75%, alternatively at least 80%, alternatively at least 85%, alternatively at least 90% alternatively at least 95%, alternatively at least 96%, alternatively at least 97%, alternatively at least 98%, alternatively at least 99%, or alternatively 100% sequence identity to any one of SEQ ID Nos: 5-16 or 28-87.

[0254] In an example of this aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptides having altered PAM specificity can comprise an amino acid sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID Nos: 5-16 or 28-87.

[0255] In another example of this aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptides having altered PAM specificity that can comprise an amino acid sequence of SEQ ID Nos: 5-16 or 28-87.

[0256] In another aspect, the disclosure provides synthetic or non-naturally occurring Cas-alpha 10 polypeptides having altered PAM specificity, wherein the Cas-alpha 10 polypeptide comprises a PI domain that recognizes a PAM sequence comprising 5'-YTC-3', 5'-TTTC-3', 5'-GTTC-3', 5'-GTYC-3', 5'-GTWC-3', 5'-DTHC-3', 5'-GTNC-3', 5'-NNNC-3', 5'-GNNC-3', 5'-HCTC-3', 5'-NTCC-3', 5'-TNC-3', 5'-GTTY-3', 5'-DTTY-3', 5'-GTTT-3', 5'-GTHC-3', 5'-NYTY-3', 5'-TTTY-3', 5'-TTYTY-3', 5'-TYTY-3', 5'-TTNN-3', 5'-NNCC-3', 5'-NCCC-3', 5'-NTNN-3', 5'-NCTY-3', 5'-NTHC-3', 5'-NNNN-3', or 5'-NNCN-3', wherein N = A, C, G, or T, Y = T or C, D = G, A, or T, W = A or T, and H = A, T, or C.

[0257] In another aspect, the disclosure provides synthetic or non-naturally occurring Cas-alpha 10 polypeptides having altered PAM specificity, wherein the Cas-alpha 10 polypeptide comprises an amino acid sequence having at least 50%, alternatively at least 55%, alternatively at least 60%, alternatively at least 65%, alternatively at least 70%, alternatively at least 75%, alternatively at

least 80%, alternatively at least 85%, alternatively at least 90% alternatively at least 95%, alternatively at least 96%, alternatively at least 97%, alternatively at least 98%, alternatively at least 99%, or alternatively 100% sequence identity to SEQ ID NO: 2 and wherein, relative to the amino acid position of SEQ ID NO: 2, the synthetic or non-naturally occurring Cas-alpha 10 polypeptide sequence comprises a K85S mutation; a K85A mutation; a N92R mutation; a N92K mutation; a Q125K mutation; a N88H mutation; a N88K mutation; a N88Q mutation; a Q125F mutation; a Y72V mutation; a Y72E mutation; a Y72Q mutation; a Y72T mutation; a Y72C mutation; a Y72A mutation; a Y72S mutation; a Y72P mutation; a Y72G mutation; a Y72D mutation; a Y72L mutation; a K85G mutation; a K85D mutation; a K85N mutation; a N88D mutation; a Q89D mutation; a N92D mutation; a N92C mutation; a N92A mutation; a N92G mutation; a N92F mutation; a N92E mutation; a N92H mutation; a N92I mutation; a N92L mutation; a N92V mutation; a N92W mutation; a N92Q mutation; a N92S mutation; a N92P mutation; a N92Y mutation; a N92T mutation; a N92M mutation; a Q125R mutation; or a Q125P mutation.

[0258] In an example of this aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptides having altered PAM specificity can comprise an amino acid sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 2, wherein the Cas-alpha 10 polypeptide sequence comprises, relative to the amino acid position of SEQ ID NO: 2, a K85S mutation; a K85A mutation; a N92R mutation; a N92K mutation; a Q125K mutation; a N88H mutation; a N88K mutation; a N88Q mutation; a Q125F mutation; a Y72V mutation; a Y72E mutation; a Y72Q mutation; a Y72T mutation; a Y72C mutation; a Y72A mutation; a Y72S mutation; a Y72P mutation; a Y72G mutation; a Y72D mutation; a Y72L mutation; a K85G mutation; a K85D mutation; a K85N mutation; a N88D mutation; a Q89D mutation; a N92D mutation; a N92C mutation; a N92A mutation; a N92G mutation; a N92F mutation; a N92E mutation; a N92H mutation; a N92I mutation; a N92L mutation; a N92V mutation; a N92W mutation; a N92Q mutation; a N92S mutation; a N92P mutation; a N92Y mutation; a N92T mutation; a N92M mutation; a Q125R mutation; or a Q125P mutation.

[0259] In another example of this aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptides having altered PAM specificity can comprise an amino acid sequence of SEQ ID NO: 2, wherein the Cas-alpha 10 polypeptide sequence comprises, relative to the amino acid

position of SEQ ID NO: 2, a K85S mutation; a K85A mutation; a N92R mutation; a N92K mutation; a Q125K mutation; a N88H mutation; a N88K mutation; a N88Q mutation; a Q125F mutation; a Y72V mutation; a Y72E mutation; a Y72Q mutation; a Y72T mutation; a Y72C mutation; a Y72A mutation; a Y72S mutation; a Y72P mutation; a Y72G mutation; a Y72D mutation; a Y72L mutation; a K85G mutation; a K85D mutation; a K85N mutation; a N88D mutation; a Q89D mutation; a N92D mutation; a N92C mutation; a N92A mutation; a N92G mutation; a N92F mutation; a N92E mutation; a N92H mutation; a N92I mutation; a N92L mutation; a N92V mutation; a N92W mutation; a N92Q mutation; a N92S mutation; a N92P mutation; a N92Y mutation; a N92T mutation; a N92M mutation; a Q125R mutation; or a Q125P mutation.

[0260] In yet another aspect, the disclosure provides synthetic or non-naturally occurring Cas-alpha 10 polypeptides having altered PAM specificity, wherein the Cas-alpha 10 polypeptide can comprise an amino acid sequence having at least 50%, alternatively at least 55%, alternatively at least 60%, alternatively at least 65%, alternatively at least 70%, alternatively at least 75%, alternatively at least 80%, alternatively at least 85%, alternatively at least 90% alternatively at least 95%, alternatively at least 96%, alternatively at least 97%, alternatively at least 98%, alternatively at least 99%, or alternatively 100% sequence identity to SEQ ID NO: 2 and wherein, relative to the amino acid position of SEQ ID NO: 2, the Cas-alpha 10 polypeptide sequence comprises one or more of the following combinations: a combination of a K85Q mutation and a N92L mutation; a combination of a N88H mutation and a Q89G mutation; a combination of a N88K mutation and a Q89G mutation; a combination of a N88Q mutation and a Q89G mutation; a combination of a K85S mutation and a N92L; a combination of a K85S mutation and a N92Q mutation; a combination of a K85S mutation and a N92C mutation; a combination of a K85S mutation and a N92H mutation; a combination of a K85S mutation and a N92A mutation; a combination of a K85S mutation and a N92M mutation; a combination of a K85N mutation and a N92L mutation; a combination of a K85N mutation and a N92H; a combination of a K85N mutation and a N92A mutation; a combination of a K85N mutation and a N92C mutation; a combination of a K85N mutation and a N92M mutation; a combination of a K85N mutation and a N92Q mutation; a combination of a K85N mutation and a N92I mutation; a combination of a K85S mutation, a N88D mutation, and a Q89G mutation; a combination of a K85S mutation, a N88H mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89D

mutation, and a Q125R mutation; a combination of a K85S mutation, a N88D mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89G mutation, and a Q125R mutation; a combination of a K85S mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72S mutation, a K85D mutation, a Q125R mutation, and a N127R mutation; a combination of a Y72A mutation, a K85S mutation, a Q89D mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88H mutation, a Q89G mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88D mutation, a Q89D mutation, a N92W mutation, and a Q125R mutation; or a combination of a K85Q and a N92W mutation.

[0261] In an example of this aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptides having altered PAM specificity can comprise an amino acid sequence having at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 2, and wherein the Cas-alpha 10 polypeptide sequence comprises, relative to the amino acid position of SEQ ID NO: 2, one or more of the following combinations: a combination of a K85Q mutation and a N92L mutation; a combination of a N88H mutation and a Q89G mutation; a combination of a N88K mutation and a Q89G mutation; a combination of a N88Q mutation and a Q89G mutation; a combination of a K85S mutation and a N92L; a combination of a K85S mutation and a N92Q mutation; a combination of a K85S mutation and a N92C mutation; a combination of a K85S mutation and a N92H mutation; a combination of a K85S mutation and a N92A mutation; a combination of a K85S mutation and a N92M mutation; a combination of a K85N mutation and a N92L mutation; a combination of a K85N mutation and a N92H; a combination of a K85N mutation and a N92A mutation; a combination of a K85N mutation and a N92C mutation; a combination of a K85N mutation and a N92M mutation; a combination of a K85N mutation and a N92Q mutation; a combination of a K85N mutation and a N92I mutation; a combination of a K85S mutation, a N88D mutation, and a Q89G mutation; a combination of a K85S mutation, a N88H mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89D mutation, and a Q125R mutation; a combination of a K85S mutation, a N88D mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89G mutation, and a Q125R mutation; a combination of a K85S mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72S mutation, a K85D mutation, a Q125R

mutation, and a N127R mutation; a combination of a Y72A mutation, a K85S mutation, a Q89D mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88H mutation, a Q89G mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88D mutation, a Q89D mutation, a N92W mutation, and a Q125R mutation; or a combination of a K85Q and a N92W mutation.

[0262] In another example of this aspect, the synthetic or non-naturally occurring Cas-alpha 10 polypeptides having altered PAM specificity can comprise an amino acid sequence of SEQ ID NO: 2, wherein the Cas-alpha 10 polypeptide sequence comprises, relative to the amino acid position of SEQ ID NO: 2, one or more of the following combinations: a combination of a K85Q mutation and a N92L mutation; a combination of a N88H mutation and a Q89G mutation; a combination of a N88K mutation and a Q89G mutation; a combination of a N88Q mutation and a Q89G mutation; a combination of a K85S mutation and a N92L; a combination of a K85S mutation and a N92Q mutation; a combination of a K85S mutation and a N92C mutation; a combination of a K85S mutation and a N92H mutation; a combination of a K85S mutation and a N92A mutation; a combination of a K85S mutation and a N92M mutation; a combination of a K85N mutation and a N92L mutation; a combination of a K85N mutation and a N92H; a combination of a K85N mutation and a N92A mutation; a combination of a K85N mutation and a N92C mutation; a combination of a K85N mutation and a N92M mutation; a combination of a K85N mutation and a N92Q mutation; a combination of a K85N mutation and a N92I mutation; a combination of a K85S mutation, a N88D mutation, and a Q89G mutation; a combination of a K85S mutation, a N88H mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89D mutation, and a Q125R mutation; a combination of a K85S mutation, a N88D mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89G mutation, and a Q125R mutation; a combination of a K85S mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72S mutation, a K85D mutation, a Q125R mutation, and a N127R mutation; a combination of a Y72A mutation, a K85S mutation, a Q89D mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88H mutation, a Q89G mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88D mutation, a Q89D mutation, a N92W mutation, and a Q125R mutation; or a combination of a K85Q and a N92W mutation.

[0263] In yet a further aspect, the disclosure provides synthetic compositions comprising: (a) a Cas-alpha 10 polypeptide having DNA binding activity, the Cas-alpha 10 polypeptide comprising an amino acid amino acid sequence having at least 50%, alternatively at least 55%, alternatively at least 60%, alternatively at least 65%, alternatively at least 70%, alternatively at least 75%, alternatively at least 80%, alternatively at least 85%, alternatively at least 90% alternatively at least 95%, alternatively at least 96%, alternatively at least 97%, alternatively at least 98%, alternatively at least 99%, or alternatively 100% sequence identity to any one of SEQ ID Nos. 5-16 or 28-87; and (b) at least one guide polynucleotide comprising a region of complementarity to a target polynucleotide, wherein the target polynucleotide is heterologous to the source of the Cas-alpha 10 polypeptide, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide and forms a complex with the at least one guide polynucleotide, and wherein the complex binds to the target polynucleotide.

[0264] In an example of this aspect, a synthetic composition can comprise: (a) a Cas-alpha 10 polypeptide having DNA binding activity, the Cas-alpha 10 polypeptide comprising an amino acid amino acid sequence having 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 any one of SEQ ID Nos. 5-16 or 28-87; and (b) at least one guide polynucleotide comprising a region of complementarity to a target polynucleotide, wherein the target polynucleotide is heterologous to the source of the Cas-alpha 10 polypeptide, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide and forms a complex with the at least one guide polynucleotide, and wherein the complex binds to the target polynucleotide.

[0265] In another example of this aspect, a synthetic composition can comprise: (a) a Cas-alpha 10 polypeptide having DNA binding activity, the Cas-alpha 10 polypeptide comprising an amino acid amino acid sequence of any one of SEQ ID Nos. 5-16 or 28-87; and (b) at least one guide polynucleotide comprising a region of complementarity to a target polynucleotide, wherein the target polynucleotide is heterologous to the source of the Cas-alpha 10 polypeptide, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide and forms a complex with the at least one guide polynucleotide, and wherein the complex binds to the target polynucleotide.

[0266] In another aspect, the disclosure provides synthetic compositions comprising: (a) a Cas-alpha 10 polypeptide having DNA binding activity, the Cas-alpha 10 polypeptide comprising a

PAM interacting (PI) domain, wherein the PI domain recognizes a PAM sequence on a target polynucleotide, wherein the PAM sequence comprises 5'-YTC-3', 5'-TTTC-3', 5'-GTTC-3', 5'-GTYC-3', 5'-GTWC-3', 5'-DTHC-3', 5'-GTNC-3', 5'-NNNC-3', 5'-GNNC-3', 5'-HCTC-3', 5'-NTCC-3', 5'-TNC-3', 5'-GTTY-3', 5'-DTTY-3', 5'-GTTT-3', 5'-GTHC-3', 5'-NYTY-3', 5'-TTTY-3', 5'-TTYTY-3', 5'-TYTY-3', 5'-TTNN-3', 5'-NNCC-3', 5'-NCCC-3', 5'-NTNN-3', 5'-NCTY-3', 5'-NTHC-3', 5'-NNNN-3', or 5'-NNCN-3', wherein N = A, C, G, or T, Y = T or C, D = G, A, or T, W = A or T, and H = A, T, or C; and (b) at least one guide polynucleotide comprising a region of complementarity to a target polynucleotide, wherein the target polynucleotide is heterologous to the source of the Cas-alpha 10 polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the at least one guide polynucleotide, and wherein the complex binds to the target polynucleotide.

[0267] In yet another aspect, the disclosure provides synthetic compositions comprising: (a) a Cas-alpha 10 polypeptide having DNA binding activity, the Cas-alpha 10 polypeptide comprising an amino acid sequence having at least 50%, alternatively at least 55%, alternatively at least 60%, alternatively at least 65%, alternatively at least 70%, alternatively at least 75%, alternatively at least 80%, alternatively at least 85%, alternatively at least 90% alternatively at least 95%, alternatively at least 96%, alternatively at least 97%, alternatively at least 98%, alternatively at least 99%, or alternatively 100% sequence identity to SEQ ID NO: 2, and wherein the Cas-alpha 10 polypeptide comprises, relative to the amino acid position of SEQ ID NO: 2, a K85S mutation; a K85A mutation; a N92R mutation; a N92K mutation; a Q125K mutation; a N88H mutation; a N88K mutation; a N88Q mutation; a Q125F mutation; a Y72V mutation; a Y72E mutation; a Y72Q mutation; a Y72T mutation; a Y72C mutation; a Y72A mutation; a Y72S mutation; a Y72P mutation; a Y72G mutation; a Y72D mutation; a Y72L mutation; a K85G mutation; a K85D mutation; a K85N mutation; a N88D mutation; a Q89D mutation; a N92D mutation; a N92C mutation; a N92A mutation; a N92G mutation; a N92F mutation; a N92E mutation; a N92H mutation; a N92I mutation; a N92L mutation; a N92V mutation; a N92W mutation; a N92Q mutation; a N92S mutation; a N92P mutation; a N92Y mutation; a N92T mutation; a N92M mutation; a Q125R mutation; or a Q125P mutation; and (b) at least one guide polynucleotide comprising a region of complementarity to a target polynucleotide, wherein the target polynucleotide is heterologous to the source of the Cas-alpha 10 polypeptide, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide and forms a

complex with the at least one guide polynucleotide, and wherein the complex binds to the target polynucleotide.

[0268] In an example of this aspect, a synthetic composition can comprise: (a) a Cas-alpha 10 polypeptide having DNA binding activity, the Cas-alpha 10 polypeptide comprising an amino acid sequence having 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 SEQ ID NO: 2, and wherein the Cas-alpha 10 polypeptide comprises, relative to the amino acid position of SEQ ID NO: 2, a K85S mutation; a K85A mutation; a N92R mutation; a N92K mutation; a Q125K mutation; a N88H mutation; a N88K mutation; a N88Q mutation; a Q125F mutation; a Y72V mutation; a Y72E mutation; a Y72Q mutation; a Y72T mutation; a Y72C mutation; a Y72A mutation; a Y72S mutation; a Y72P mutation; a Y72G mutation; a Y72D mutation; a Y72L mutation; a K85G mutation; a K85D mutation; a K85N mutation; a N88D mutation; a Q89D mutation; a N92D mutation; a N92C mutation; a N92A mutation; a N92G mutation; a N92F mutation; a N92E mutation; a N92H mutation; a N92I mutation; a N92L mutation; a N92V mutation; a N92W mutation; a N92Q mutation; a N92S mutation; a N92P mutation; a N92Y mutation; a N92T mutation; a N92M mutation; a Q125R mutation; or a Q125P mutation; and (b) at least one guide polynucleotide comprising a region of complementarity to a target polynucleotide, wherein the target polynucleotide is heterologous to the source of the Cas-alpha 10 polypeptide, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide and forms a complex with the at least one guide polynucleotide, and wherein the complex binds to the target polynucleotide.

[0269] In another example of this aspect, a synthetic composition can comprise: (a) a Cas-alpha 10 polypeptide having DNA binding activity, the Cas-alpha 10 polypeptide comprising the amino acid sequence of SEQ ID NO: 2 and a K85S mutation; a K85A mutation; a N92R mutation; a N92K mutation; a Q125K mutation; a N88H mutation; a N88K mutation; a N88Q mutation; a Q125F mutation; a Y72V mutation; a Y72E mutation; a Y72Q mutation; a Y72T mutation; a Y72C mutation; a Y72A mutation; a Y72S mutation; a Y72P mutation; a Y72G mutation; a Y72D mutation; a Y72L mutation; a K85G mutation; a K85D mutation; a K85N mutation; a N88D mutation; a Q89D mutation; a N92D mutation; a N92C mutation; a N92A mutation; a N92G mutation; a N92F mutation; a N92E mutation; a N92H mutation; a N92I mutation; a N92L mutation; a N92V mutation; a N92W mutation; a N92Q mutation; a N92S mutation; a N92P

mutation; a N92Y mutation; a N92T mutation; a N92M mutation; a Q125R mutation; or a Q125P mutation; and (b) at least one guide polynucleotide comprising a region of complementarity to a target polynucleotide, wherein the target polynucleotide is heterologous to the source of the Cas-alpha 10 polypeptide, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide and forms a complex with the at least one guide polynucleotide, and wherein the complex binds to the target polynucleotide.

[0270] In yet another aspect, the disclosure provides synthetic compositions comprising: (a) a Cas-alpha 10 polypeptide having DNA binding activity, the Cas-alpha 10 polypeptide comprising an amino acid sequence having at least 50%, alternatively at least 55%, alternatively at least 60%, alternatively at least 65%, alternatively at least 70%, alternatively at least 75%, alternatively at least 80%, alternatively at least 85%, alternatively at least 90% alternatively at least 95%, alternatively at least 96%, alternatively at least 97%, alternatively at least 98%, alternatively at least 99%, or alternatively 100% sequence identity to SEQ ID NO: 2 and wherein the Cas-alpha 10 polypeptide comprises, relative to the amino acid position of SEQ ID NO: 2, one or more of the following combinations: a combination of a K85Q mutation and a N92L mutation; a combination of a N88H mutation and a Q89G mutation; a combination of a N88K mutation and a Q89G mutation; a combination of a N88Q mutation and a Q89G mutation; a combination of a K85S mutation and a N92L; a combination of a K85S mutation and a N92Q mutation; a combination of a K85S mutation and a N92C mutation; a combination of a K85S mutation and a N92H mutation; a combination of a K85S mutation and a N92A mutation; a combination of a K85S mutation and a N92M mutation; a combination of a K85N mutation and a N92L mutation; a combination of a K85N mutation and a N92H; a combination of a K85N mutation and a N92A mutation; a combination of a K85N mutation and a N92C mutation; a combination of a K85N mutation and a N92M mutation; a combination of a K85N mutation and a N92Q mutation; a combination of a K85N mutation and a N92I mutation; a combination of a K85S mutation, a N88D mutation, and a Q89G mutation; a combination of a K85S mutation, a N88H mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89D mutation, and a Q125R mutation; a combination of a K85S mutation, a N88D mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89G mutation, and a Q125R mutation; a combination of a K85S mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72S mutation, a K85D mutation, a Q125R mutation, and

a N127R mutation; a combination of a Y72A mutation, a K85S mutation, a Q89D mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88H mutation, a Q89G mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88D mutation, a Q89D mutation, a N92W mutation, and a Q125R mutation; or a combination of a K85Q and a N92W mutation; and (b) at least one guide polynucleotide comprising a region of complementarity to a target polynucleotide, wherein the target polynucleotide is heterologous to the source of the Cas-alpha 10 polypeptide, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide and forms a complex with the at least one guide polynucleotide, and wherein the complex binds to the target polynucleotide.

[0271] In an example of this aspect, a synthetic composition can comprise: (a) a Cas-alpha 10 polypeptide having DNA binding activity, the Cas-alpha 10 polypeptide comprising an amino acid sequence having 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 SEQ ID NO: 2 and wherein the Cas-alpha 10 polypeptide comprises, relative to the amino acid position of SEQ ID NO: 2, one or more of the following combinations: a combination of a K85Q mutation and a N92L mutation; a combination of a N88H mutation and a Q89G mutation; a combination of a N88K mutation and a Q89G mutation; a combination of a N88Q mutation and a Q89G mutation; a combination of a K85S mutation and a N92L; a combination of a K85S mutation and a N92Q mutation; a combination of a K85S mutation and a N92C mutation; a combination of a K85S mutation and a N92H mutation; a combination of a K85S mutation and a N92A mutation; a combination of a K85S mutation and a N92M mutation; a combination of a K85N mutation and a N92L mutation; a combination of a K85N mutation and a N92H; a combination of a K85N mutation and a N92A mutation; a combination of a K85N mutation and a N92C mutation; a combination of a K85N mutation and a N92M mutation; a combination of a K85N mutation and a N92Q mutation; a combination of a K85N mutation and a N92I mutation; a combination of a K85S mutation, a N88D mutation, and a Q89G mutation; a combination of a K85S mutation, a N88H mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89D mutation, and a Q125R mutation; a combination of a K85S mutation, a N88D mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89G mutation, and a Q125R mutation; a combination of a K85S mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72S mutation, a K85D mutation, a Q125R

mutation, and a N127R mutation; a combination of a Y72A mutation, a K85S mutation, a Q89D mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88H mutation, a Q89G mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88D mutation, a Q89D mutation, a N92W mutation, and a Q125R mutation; or a combination of a K85Q and a N92W mutation; and (b) at least one guide polynucleotide comprising a region of complementarity to a target polynucleotide, wherein the target polynucleotide is heterologous to the source of the Cas-alpha 10 polypeptide, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide and forms a complex with the at least one guide polynucleotide, and wherein the complex binds to the target polynucleotide.

[0272] In another example of this aspect, a synthetic or non-naturally occurring composition can comprise: (a) a Cas-alpha 10 polypeptide having DNA binding activity, the Cas-alpha 10 polypeptide comprising the amino acid sequence of SEQ ID NO: 2 and one or more of the following combinations: a combination of a K85Q mutation and a N92L mutation; a combination of a N88H mutation and a Q89G mutation; a combination of a N88K mutation and a Q89G mutation; a combination of a N88Q mutation and a Q89G mutation; a combination of a K85S mutation and a N92L; a combination of a K85S mutation and a N92Q mutation; a combination of a K85S mutation and a N92C mutation; a combination of a K85S mutation and a N92H mutation; a combination of a K85S mutation and a N92A mutation; a combination of a K85S mutation and a N92M mutation; a combination of a K85N mutation and a N92L mutation; a combination of a K85N mutation and a N92H; a combination of a K85N mutation and a N92A mutation; a combination of a K85N mutation and a N92C mutation; a combination of a K85N mutation and a N92M mutation; a combination of a K85N mutation and a N92Q mutation; a combination of a K85N mutation and a N92I mutation; a combination of a K85S mutation, a N88D mutation, and a Q89G mutation; a combination of a K85S mutation, a N88H mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89D mutation, and a Q125R mutation; a combination of a K85S mutation, a N88D mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89G mutation, and a Q125R mutation; a combination of a K85S mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72S mutation, a K85D mutation, a Q125R mutation, and a N127R mutation; a combination of a Y72A mutation, a K85S mutation, a Q89D mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88H mutation, a Q89G mutation, and a

Q125R mutation; a combination of a Y72C mutation, a N88D mutation, a Q89D mutation, a N92W mutation, and a Q125R mutation; or a combination of a K85Q and a N92W mutation; and (b) at least one guide polynucleotide comprising a region of complementarity to a target polynucleotide, wherein the target polynucleotide is heterologous to the source of the Cas-alpha 10 polypeptide, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide and forms a complex with the at least one guide polynucleotide, and wherein the complex binds to the target polynucleotide.

[0273] In a further aspect, the disclosure provides a method of editing a target polynucleotide in a cell, the method comprising: (a) providing to the cell a Cas-alpha 10 polypeptide having at least 50%, alternatively at least 55%, alternatively at least 60%, alternatively at least 65%, alternatively at least 70%, alternatively at least 75%, alternatively at least 80%, alternatively at least 85%, alternatively at least 90% alternatively at least 95%, alternatively at least 96%, alternatively at least 97%, alternatively at least 98%, alternatively at least 99%, or alternatively 100% sequence identity to any one of SEQ ID Nos. 5-16 or 28-87, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide; (b) providing the cell with at least one guide polynucleotide comprising a region of complementarity to the target polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the guide polynucleotide and the complex binds the target polynucleotide; and (c) introducing at least one nucleotide modification in the target polynucleotide via the complex, wherein the target polynucleotide is heterologous to the Cas-alpha 10 polypeptide.

[0274] In an example of this aspect, a method of editing a target polynucleotide in a cell can comprise: (a) providing to the cell a Cas-alpha 10 polypeptide having 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 identify to any one of SEQ ID Nos. 5-16 or 28-87, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide; (b) providing the cell with at least one guide polynucleotide comprising a region of complementarity to the target polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the guide polynucleotide and the complex binds the target polynucleotide; and (c) introducing at least one nucleotide modification in the target polynucleotide via the complex, wherein the target polynucleotide is heterologous to the Cas-alpha 10 polypeptide.

[0275] In another example of this aspect, a method of editing a target polynucleotide in a cell can comprise: (a) providing to the cell a Cas-alpha 10 polypeptide of any one of SEQ ID Nos. 5-16 or 28-87, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide; (b) providing the cell with at least one guide polynucleotide comprising a region of complementarity to the target polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the guide polynucleotide and the complex binds the target polynucleotide; and (c) introducing at least one nucleotide modification in the target polynucleotide via the complex, wherein the target polynucleotide is heterologous to the Cas-alpha 10 polypeptide.

[0276] In some examples of the foregoing methods of editing a target polynucleotide in a cell, the PAM sequence recognized by the Cas-alpha 10 polypeptide comprises 5'-YTC-3', 5'-TTTC-3', 5'-GTTC-3', 5'-GTYC-3', 5'-GTWC-3', 5'-DTHC-3', 5'-GTNC-3', 5'-NNNC-3', 5'-GNNC-3', 5'-HCTC-3', 5'-NTCC-3', 5'-TNC-3', 5'-GTTY-3', 5'-DTTY-3', 5'-GTTT-3', 5'-GTHC-3', 5'-NYTY-3', 5'-TTTY-3', 5'-TTYTY-3', 5'-TYTY-3', 5'-TTNN-3', 5'-NNCC-3', 5'-NCCC-3', 5'-NTNN-3', 5'-NCTY-3', 5'-NTHC-3', 5'-NNNN-3', or 5'-NNCN-3', wherein N = A, C, G, or T, Y = T or C, D = G, A, or T, W = A or T, and H = A, T, or C.

[0277] In some examples, the disclosed methods of editing a target polynucleotide in a cell, the method further comprises providing the cell with a donor DNA molecule or a polynucleotide modification template.

[0278] In yet a further aspect, the disclosure provides methods of editing a target polynucleotide in a cell, the method comprising: (a) providing to the cell a Cas-alpha 10 polypeptide comprising a PAM interacting (PI) domain, wherein the PI domain recognizes a PAM sequence on the target polynucleotide, and wherein the PAM sequence comprises 5'-YTC-3', 5'-TTTC-3', 5'-GTTC-3', 5'-GTYC-3', 5'-GTWC-3', 5'-DTHC-3', 5'-GTNC-3', 5'-NNNC-3', 5'-GNNC-3', 5'-HCTC-3', 5'-NTCC-3', 5'-TNC-3', 5'-GTTY-3', 5'-DTTY-3', 5'-GTTT-3', 5'-GTHC-3', 5'-NYTY-3', 5'-TTTY-3', 5'-TTYTY-3', 5'-TYTY-3', 5'-TTNN-3', 5'-NNCC-3', 5'-NCCC-3', 5'-NTNN-3', 5'-NCTY-3', 5'-NTHC-3', 5'-NNNN-3', or 5'-NNCN-3', wherein N = A, C, G, or T, Y = T or C, D = G, A, or T, W = A or T, and H = A, T, or C; (b) providing the cell with at least one guide polynucleotide comprising a region of complementarity to the target polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the guide polynucleotide and the complex binds the target polynucleotide; and (c) introducing at least one nucleotide modification in the target

polynucleotide via the complex, wherein the target polynucleotide is heterologous to the Cas-alpha 10 polypeptide.

[0279] In some aspects of the method of editing a target polynucleotide in a cell, the method further comprises providing the cell with a donor DNA molecule or a polynucleotide modification template.

[0280] In another aspect, the disclosure provides a method of editing a target polynucleotide in a cell, the method comprising: (a) providing to the cell a Cas-alpha 10 polypeptide comprising an amino acid sequence having at least 50%, alternatively at least 55%, alternatively at least 60%, alternatively at least 65%, alternatively at least 70%, alternatively at least 75%, alternatively at least 80%, alternatively at least 85%, alternatively at least 90% alternatively at least 95%, alternatively at least 96%, alternatively at least 97%, alternatively at least 98%, alternatively at least 99%, or alternatively 100% sequence identity to SEQ ID NO: 2, wherein the Cas-alpha 10 polypeptide sequence comprises, relative to the amino acid position of SEQ ID NO: 2, a K85S mutation; a K85A mutation; a N92R mutation; a N92K mutation; a Q125K mutation; a N88H mutation; a N88K mutation; a N88Q mutation; a Q125F mutation; a Y72V mutation; a Y72E mutation; a Y72Q mutation; a Y72T mutation; a Y72C mutation; a Y72A mutation; a Y72S mutation; a Y72P mutation; a Y72G mutation; a Y72D mutation; a Y72L mutation; a K85G mutation; a K85D mutation; a K85N mutation; a N88D mutation; a Q89D mutation; a N92D mutation; a N92C mutation; a N92A mutation; a N92G mutation; a N92F mutation; a N92E mutation; a N92H mutation; a N92I mutation; a N92L mutation; a N92V mutation; a N92W mutation; a N92Q mutation; a N92S mutation; a N92P mutation; a N92Y mutation; a N92T mutation; a N92M mutation; a Q125R mutation; or a Q125P mutation, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide; (b) providing the cell with at least one guide polynucleotide comprising a region of complementarity to the target polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the guide polynucleotide and the complex binds the target polynucleotide; and (c) introducing at least one nucleotide modification in the target polynucleotide via the complex, wherein the target polynucleotide is heterologous to the Cas-alpha 10 polypeptide.

[0281] In an example of this aspect, a method of editing a target polynucleotide in a cell can comprise: (a) providing to the cell a Cas-alpha 10 polypeptide comprising an amino acid sequence having 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 SEQ ID NO: 2, wherein the Cas-alpha 10 polypeptide sequence comprises, relative to the amino acid position of SEQ ID NO: 2, a K85S mutation; a K85A mutation; a N92R mutation; a N92K mutation; a Q125K mutation; a N88H mutation; a N88K mutation; a N88Q mutation; a Q125F mutation; a Y72V mutation; a Y72E mutation; a Y72Q mutation; a Y72T mutation; a Y72C mutation; a Y72A mutation; a Y72S mutation; a Y72P mutation; a Y72G mutation; a Y72D mutation; a Y72L mutation; a K85G mutation; a K85D mutation; a K85N mutation; a N88D mutation; a Q89D mutation; a N92D mutation; a N92C mutation; a N92A mutation; a N92G mutation; a N92F mutation; a N92E mutation; a N92H mutation; a N92I mutation; a N92L mutation; a N92V mutation; a N92W mutation; a N92Q mutation; a N92S mutation; a N92P mutation; a N92Y mutation; a N92T mutation; a N92M mutation; a Q125R mutation; or a Q125P mutation, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide; (b) providing the cell with at least one guide polynucleotide comprising a region of complementarity to the target polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the guide polynucleotide and the complex binds the target polynucleotide; and (c) introducing at least one nucleotide modification in the target polynucleotide via the complex, wherein the target polynucleotide is heterologous to the Cas-alpha 10 polypeptide.

[0282] In another example of this aspect, a method of editing a target polynucleotide in a cell can comprise: (a) providing to the cell a Cas-alpha 10 polypeptide comprising an amino acid of SEQ ID NO: 2, wherein the Cas-alpha 10 polypeptide comprises, relative to the amino acid position of SEQ ID NO: 2, a K85S mutation; a K85A mutation; a N92R mutation; a N92K mutation; a Q125K mutation; a N88H mutation; a N88K mutation; a N88Q mutation; a Q125F mutation; a Y72V mutation; a Y72E mutation; a Y72Q mutation; a Y72T mutation; a Y72C mutation; a Y72A mutation; a Y72S mutation; a Y72P mutation; a Y72G mutation; a Y72D mutation; a Y72L mutation; a K85G mutation; a K85D mutation; a K85N mutation; a N88D mutation; a Q89D mutation; a N92D mutation; a N92C mutation; a N92A mutation; a N92G mutation; a N92F mutation; a N92E mutation; a N92H mutation; a N92I mutation; a N92L mutation; a N92V mutation; a N92W mutation; a N92Q mutation; a N92S mutation; a N92P mutation; a N92Y mutation; a N92T mutation; a N92M mutation; a Q125R mutation; or a Q125P mutation, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide; (b) providing the cell with at least one guide polynucleotide comprising a region of complementarity

to the target polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the guide polynucleotide and the complex binds the target polynucleotide; and (c) introducing at least one nucleotide modification in the target polynucleotide via the complex, wherein the target polynucleotide is heterologous to the Cas-alpha 10 polypeptide.

[0283] In another aspect, the disclosure provides a method of editing a target polynucleotide in a cell, the method comprising: (a) providing to the cell a Cas-alpha 10 polypeptide comprising an amino acid sequence having at least 50%, alternatively at least 55%, alternatively at least 60%, alternatively at least 65%, alternatively at least 70%, alternatively at least 75%, alternatively at least 80%, alternatively at least 85%, alternatively at least 90% alternatively at least 95%, alternatively at least 96%, alternatively at least 97%, alternatively at least 98%, alternatively at least 99%, or alternatively 100% sequence identity to SEQ ID NO: 2, wherein the Cas-alpha 10 polypeptide sequence comprises, relative to the amino acid position of SEQ ID NO: 2, one or more of the following combinations: a combination of a K85Q mutation and a N92L mutation; a combination of a N88H mutation and a Q89G mutation; a combination of a N88K mutation and a Q89G mutation; a combination of a N88Q mutation and a Q89G mutation; a combination of a K85S mutation and a N92L; a combination of a K85S mutation and a N92Q mutation; a combination of a K85S mutation and a N92C mutation; a combination of a K85S mutation and a N92H mutation; a combination of a K85S mutation and a N92A mutation; a combination of a K85S mutation and a N92M mutation; a combination of a K85N mutation and a N92L mutation; a combination of a K85N mutation and a N92H; a combination of a K85N mutation and a N92A mutation; a combination of a K85N mutation and a N92C mutation; a combination of a K85N mutation and a N92M mutation; a combination of a K85N mutation and a N92Q mutation; a combination of a K85N mutation and a N92I mutation; a combination of a K85S mutation, a N88D mutation, and a Q89G mutation; a combination of a K85S mutation, a N88H mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89D mutation, and a Q125R mutation; a combination of a K85S mutation, a N88D mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89G mutation, and a Q125R mutation; a combination of a K85S mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72S mutation, a K85D mutation, a Q125R mutation, and a N127R mutation; a combination of a Y72A mutation, a K85S mutation, a Q89D mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88H mutation, a

Q89G mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88D mutation, a Q89D mutation, a N92W mutation, and a Q125R mutation; or a combination of a K85Q and a N92W mutation, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide; (b) providing the cell with at least one guide polynucleotide comprising a region of complementarity to the target polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the guide polynucleotide and the complex binds the target polynucleotide; and (c) introducing at least one nucleotide modification in the target polynucleotide via the complex, wherein the target polynucleotide is heterologous to the Cas-alpha 10 polypeptide.

[0284] In an example of this aspect, a method of editing a target polynucleotide in a cell can comprise: (a) providing to the cell a Cas-alpha 10 polypeptide comprising an amino acid sequence having 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 SEQ ID NO: 2, wherein the Cas-alpha 10 polypeptide sequence comprises, relative to the amino acid position of SEQ ID NO: 2, one or more of the following combinations: a combination of a K85Q mutation and a N92L mutation; a combination of a N88H mutation and a Q89G mutation; a combination of a N88K mutation and a Q89G mutation; a combination of a N88Q mutation and a Q89G mutation; a combination of a K85S mutation and a N92L; a combination of a K85S mutation and a N92Q mutation; a combination of a K85S mutation and a N92C mutation; a combination of a K85S mutation and a N92H mutation; a combination of a K85S mutation and a N92A mutation; a combination of a K85S mutation and a N92M mutation; a combination of a K85N mutation and a N92L mutation; a combination of a K85N mutation and a N92H; a combination of a K85N mutation and a N92A mutation; a combination of a K85N mutation and a N92C mutation; a combination of a K85N mutation and a N92M mutation; a combination of a K85N mutation and a N92Q mutation; a combination of a K85N mutation and a N92I mutation; a combination of a K85S mutation, a N88D mutation, and a Q89G mutation; a combination of a K85S mutation, a N88H mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89D mutation, and a Q125R mutation; a combination of a K85S mutation, a N88D mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89G mutation, and a Q125R mutation; a combination of a K85S mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72S mutation, a K85D mutation, a Q125R mutation, and a N127R mutation; a combination of a Y72A mutation, a K85S mutation, a Q89D

mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88H mutation, a Q89G mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88D mutation, a Q89D mutation, a N92W mutation, and a Q125R mutation; or a combination of a K85Q and a N92W mutation, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide; (b) providing the cell with at least one guide polynucleotide comprising a region of complementarity to the target polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the guide polynucleotide and the complex binds the target polynucleotide; and (c) introducing at least one nucleotide modification in the target polynucleotide via the complex, wherein the target polynucleotide is heterologous to the Cas-alpha 10 polypeptide.

[0285] In another example of this aspect, a method of editing a target polynucleotide in a cell can comprise: (a) providing to the cell a Cas-alpha 10 polypeptide comprising an amino acid of SEQ ID NO: 2, wherein the Cas-alpha 10 polypeptide sequence comprises, relative to the amino acid position of SEQ ID NO: 2, one or more of the following combinations: a combination of a K85Q mutation and a N92L mutation; a combination of a N88H mutation and a Q89G mutation; a combination of a N88K mutation and a Q89G mutation; a combination of a N88Q mutation and a Q89G mutation; a combination of a K85S mutation and a N92L; a combination of a K85S mutation and a N92Q mutation; a combination of a K85S mutation and a N92C mutation; a combination of a K85S mutation and a N92H mutation; a combination of a K85S mutation and a N92A mutation; a combination of a K85S mutation and a N92M mutation; a combination of a K85N mutation and a N92L mutation; a combination of a K85N mutation and a N92H; a combination of a K85N mutation and a N92A mutation; a combination of a K85N mutation and a N92C mutation; a combination of a K85N mutation and a N92M mutation; a combination of a K85N mutation and a N92Q mutation; a combination of a K85N mutation and a N92I mutation; a combination of a K85S mutation, a N88D mutation, and a Q89G mutation; a combination of a K85S mutation, a N88H mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89D mutation, and a Q125R mutation; a combination of a K85S mutation, a N88D mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89G mutation, and a Q125R mutation; a combination of a K85S mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72S mutation, a K85D mutation, a Q125R mutation, and a N127R mutation; a combination of a Y72A mutation, a K85S mutation, a Q89D

mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88H mutation, a Q89G mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88D mutation, a Q89D mutation, a N92W mutation, and a Q125R mutation; or a combination of a K85Q and a N92W mutation, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide; (b) providing the cell with at least one guide polynucleotide comprising a region of complementarity to the target polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the guide polynucleotide and the complex binds the target polynucleotide; and (c) introducing at least one nucleotide modification in the target polynucleotide via the complex, wherein the target polynucleotide is heterologous to the Cas-alpha 10 polypeptide.

[0286] In some aspects of the method of editing a target polynucleotide in a cell, the PAM sequence recognized by the Cas-alpha 10 polypeptide comprises 5'-YTC-3', 5'-TTTC-3', 5'-GTTC-3', 5'-GTYC-3', 5'-GTWC-3', 5'-DTHC-3', 5'-GTNC-3', 5'-NNNC-3', 5'-GNNC-3', 5'-HCTC-3', 5'-NTCC-3', 5'-TNC-3', 5'-GTTY-3', 5'-DTTY-3', 5'-GTTT-3', 5'-GTHC-3', 5'-NYTY-3', 5'-TTTY-3', 5'-TTYTY-3', 5'-TYTY-3', 5'-TTNN-3', 5'-NNCC-3', 5'-NCCC-3', 5'-NTNN-3', 5'-NCTY-3', 5'-NTHC-3', 5'-NNNN-3', or 5'-NNCN-3', wherein N = A, C, G, or T, Y = T or C, D = G, A, or T, W = A or T, and H = A, T, or C.

[0287] In some aspects of the method of editing a target polynucleotide in a cell, the method further comprises providing the cell with a donor DNA molecule or a polynucleotide modification template.

CRISPR-Cas System Components

Cas polypeptides

[0288] A number of proteins may be encoded in the CRISPR *cas* operon, including those involved in adaptation (spacer insertion), interference (effector module target binding, target nicking or cleavage – *e.g.* endonuclease activity), expression (pre-crRNA processing), regulation, or other.

[0289] Two proteins, Cas1 and Cas2, are conserved among many CRISPR systems (for example, as described in Koonin *et al.*, *Curr Opinion Microbiology* 37:67-78, 2017). Cas1 is a metal-dependent DNA-specific endonuclease that produces double-stranded DNA fragments. In some systems Cas1 forms a stable complex with Cas2, which is essential to spacer acquisition and insertion for CRISPR systems (Nuñez *et al.*, *Nature Str Mol Biol* 21:528-534, 2014).

[0290] A number of other proteins have been identified across different systems, including Cas4 (which may have similarity to a RecB nuclease) and is thought to play a role in the capture of new viral DNA sequences for incorporation into the CRISPR array (Zhang et al., PLOS One 7(10):e47232, 2012).

[0291] Some proteins may encompass a plurality of functions. For example, Cas9, the signature protein of Class 2 type II systems, has been demonstrated to be involved in pre-crRNA processing, target binding, as well as target cleavage.

Cas Endonucleases and Effectors

[0292] Endonucleases are enzymes that cleave the phosphodiester bond within a polynucleotide chain, and include restriction endonucleases that cleave DNA at specific sites without damaging the bases. Examples of endonucleases include restriction endonucleases, meganucleases, TAL effector nucleases (TALENs), zinc finger nucleases, and Cas (CRISPR-associated) effector endonucleases.

[0293] Cas endonucleases, either as single effector proteins or in an effector complex with other components, unwind the DNA duplex at the target sequence and optionally cleave at least one DNA strand, as mediated by recognition of the target sequence by a polynucleotide (such as, but not limited to, a crRNA or guide RNA) that is in complex with the Cas effector protein. Such recognition and cutting of a target sequence by a Cas endonuclease typically occurs if the correct protospacer-adjacent motif (PAM) is located at or adjacent to the 3' end of the DNA target sequence. Alternatively, a Cas endonuclease herein may lack DNA cleavage or nicking activity, but can still specifically bind to a DNA target sequence when complexed with a suitable RNA component. (See also U.S. Patent Application US20150082478 published 19 March 2015 and US20150059010 published 26 February 2015).

[0294] Cas endonucleases may occur as individual effectors (Class 2 CRISPR systems) or as part of larger effector complexes (Class I CRISPR systems).

[0295] Cas endonucleases that have been described include, but are not limited to, for example: Cas9, Cas12f (Cas-alpha, Cas14), Cas12l (Cas-beta), Cas12a (Cpf1), Cas12b (a C2c1 protein), Cas13 (a C2c2 protein), Cas12c (a C2c3 protein), Cas12d, Cas12e, Cas12g, Cas12h, Cas12i, Cas12j, Cas12k, Cas3, Cas3-HD, Cas 5, Cas6, Cas7, Cas8, Cas10, or combinations or complexes of these. Cas endonucleases and effector proteins can be used for targeted genome editing (via simplex and multiplex double-strand breaks and nicks) and targeted genome regulation (via

tethering of epigenetic effector domains to either the Cas polypeptide or sgRNA. A Cas endonuclease can also be engineered to function as an RNA-guided recombinase, and via RNA tethers could serve as a scaffold for the assembly of multiprotein and nucleic acid complexes (Mali *et al.*, 2013, *Nature Methods Vol.* 10:957-963).

Cas-alpha Endonucleases

[0296] A Cas-alpha endonuclease (e.g., Cas-alpha 10, also known as Cas12f) is defined as a functional RNA-guided, PAM-dependent dsDNA cleavage protein of fewer than 800 amino acids, comprising: a C-terminal RuvC catalytic domain split into three subdomains and further comprising bridge-helix and one or more Zinc finger motif(s); and an N-terminal Rec subunit with a helical bundle, WED wedge-like (or "Oligonucleotide Binding Domain", OBD) domain, and, optionally, a Zinc finger motif. Some exemplary Cas-alpha endonucleases are described, for example, in US10934536 and WO2022082179.

[0297] RuvC domains have been demonstrated in the literature to encompass endonuclease functionality. A Cas-alpha endonuclease may be isolated or identified from a locus that comprises a *cas-alpha* gene encoding an effector protein, and an array comprising a plurality repeats. In some aspects, a *cas-alpha* locus may further comprise a partial or whole *cas1* gene, a *cas2* gene, and/or a *cas4* gene.

[0298] Zinc finger motifs are domains that coordinate one or more zinc ions, usually through Cysteine and Histidine sidechains, to stabilize their fold. Zinc fingers are named for the pattern of Cysteine and Histidine residues that coordinate the zinc ion (*e.g.*, C4 means a zinc ion is coordinated by four Cysteine residues; C3H means a zinc ion is coordinated by three Cysteine residues and one Histidine residue).

[0299] Cas-alpha polypeptides comprise one or more Zinc Finger (ZFN) coordination motif(s) that may form a Zinc binding domain. Zinc Finger-like motifs can aid in target and non-target strand separation and loading of the guide RNA into the DNA target. Cas-alpha polypeptides comprising one or more Zinc Finger motifs may provide additional stability to the ribonucleoprotein complex on the target polynucleotide. Cas-alpha polypeptides comprise C4 or C3H zinc binding domains.

[0300] Cas-alpha endonucleases are RNA-guided endonucleases capable of binding to, and cleaving, a double-strand DNA target that comprises: (1) a sequence sharing homology with a

nucleotide sequence of the guide RNA, and (2) a PAM sequence. In some aspects, the PAM is T-rich. In some aspects, the PAM is C-rich.

[0301] A Cas-alpha endonuclease is functional as a double-strand-break-inducing agent, and may also be a nickase, or a single-strand-break inducing agent. In some aspects, a catalytically inactive Cas-alpha endonuclease may be used to target or recruit to a target DNA sequence but not induce cleavage. In some aspects, a catalytically inactive Cas-alpha polypeptide may be used with a functional endonuclease, to cleave a target sequence. In some aspects, a catalytically inactive Cas-alpha polypeptide may be combined with a base editing molecule, such as a deaminase. A "deaminase" is an enzyme that catalyzes a deamination reaction. For example, deamination of adenine with an adenine deaminase results in the formation of inosine. Inosine selectively base pairs with cytosine instead of thymine. This results in a post-replicative transition mutation, such that the original A – T base pair transforms into a G – C base pair. In another example, cytosine deamination results in the formation of uracil, which can be repaired by cellular repair mechanisms back to a C – T base pair or to a T – A, G – C, or A – T base pair. This heterogeneity in repair can be suppressed by the introduction of a uracil glycosylase inhibitor, such that DNA repair or replication transforms the original C – T base pair into a T – A base pair (Burnett *et al.* (2022) *Frontiers in Genome Editing*. 4, 923718). In the case of both adenine and cytosine deaminases, the introduction of a nick promotes the respective base pair change (Burnett *et al.*, 2022). In some aspects, a deaminase may be a cytidine deaminase. In some aspects, a deaminase may be an adenine deaminase. In some aspects, a deaminase may be ADAR-2.

[0302] A "functional fragment" of a Cas-alpha endonuclease retains the ability to recognize, or bind, or nick a single strand of a double-stranded polynucleotide, or cleave both strands of a double-stranded polynucleotide, or any combination of the preceding.

[0303] A Cas polypeptide, effector protein, or functional fragment thereof, for use in the disclosed methods, can be isolated from a native source, or from, a recombinant source where the genetically modified host cell is modified to express the nucleic acid sequence encoding the protein. Alternatively, the Cas polypeptide can be produced using cell free protein expression systems, or be synthetically produced. Effector Cas nucleases may be isolated and introduced into a heterologous cell, or may be modified from its native form to exhibit a different type or magnitude of activity than what it would exhibit in its native source. Such modifications include but are not limited to: fragments, variants, substitutions, deletions, and insertions.

[0304] Fragments and variants of Cas polypeptides and Cas effector proteins can be obtained via methods such as site-directed mutagenesis and synthetic construction. Methods for measuring endonuclease activity are well known in the art such as, but not limiting to, WO2013166113 published 07 November 2013, WO2016186953 published 24 November 2016, and WO2016186946 published 24 November 2016.

[0305] The Cas polypeptides disclosed herein can be modified. The modified form of the Cas polypeptide can include an amino acid change (e.g., deletion, insertion, or substitution) that reduces the naturally-occurring nuclease activity of the Cas polypeptide. For example, in some instances, the modified form of the Cas polypeptide has less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% of the nuclease activity of the corresponding wild-type Cas polypeptide (US20140068797 published 06 March 2014). In some cases, the modified form of the Cas polypeptide has no substantial nuclease activity and is referred to as catalytically "inactivated Cas" or "deactivated Cas (dCas)." An inactivated Cas/deactivated Cas includes a deactivated Cas endonuclease (dCas). A catalytically inactive Cas effector protein can be fused to a heterologous sequence to induce or modify activity.

[0306] Cas polypeptides disclosed herein can be part of a fusion protein comprising one or more heterologous protein domains (e.g., 1, 2, 3, or more domains in addition to the Cas polypeptide). Such a fusion protein may comprise any additional protein sequence, and optionally a linker sequence between any two domains, such as between Cas and a first heterologous domain. Examples of protein domains that may be fused to a Cas polypeptide herein include, without limitation, epitope tags (e.g., histidine [His], V5, FLAG, influenza hemagglutinin [HA], myc, VSV-G, thioredoxin [Trx]), reporters (e.g., glutathione-5-transferase [GST], horseradish peroxidase [HRP], chloramphenicol acetyltransferase [CAT], beta-galactosidase, beta-glucuronidase [GUS], luciferase, green fluorescent protein [GFP], HcRed, DsRed, cyan fluorescent protein [CFP], yellow fluorescent protein [YFP], blue fluorescent protein [BFP]), and domains having one or more of the following activities: methylase activity, demethylase activity, transcription activation activity (e.g., VP16 or VP64), transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity and nucleic acid binding activity. A disclosed Cas polypeptide can also be in fusion with a protein that binds DNA molecules or other molecules, such as maltose binding protein (MBP), S-tag, Lex A

DNA binding domain (DBD), GAL4A DNA binding domain, and herpes simplex virus (HSV) VP16.

[0307] A catalytically active and/or inactive Cas polypeptide can be fused to a heterologous sequence (US20140068797 published 06 March 2014). Suitable fusion partners include, but are not limited to, a polypeptide that provides an activity that indirectly increases transcription by acting directly on the target DNA or on a polypeptide (e.g., a histone or other DNA-binding protein) associated with the target DNA. Additional suitable fusion partners include, but are not limited to, a polypeptide that provides for methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, or demyristoylation activity. Further suitable fusion partners include, but are not limited to, a polypeptide that directly provides for increased transcription of the target nucleic acid (e.g., a transcription activator or a fragment thereof, a protein or fragment thereof that recruits a transcription activator, a small molecule/drug-responsive transcription regulator, etc.). A partially active or catalytically inactive Cas-alpha endonuclease can also be fused to another protein or domain, for example Clo51 or FokI nuclease, to generate double-strand breaks (Guilinger *et al. Nature Biotechnology*, volume 32, number 6, June 2014).

[0308] A catalytically active or inactive Cas polypeptide, such as the Cas-alpha polypeptides disclosed herein, can also be in fusion with a molecule that directs editing of single or multiple bases in a polynucleotide sequence, for example a site-specific deaminase that can change the identity of a nucleotide, for example from C•G to T•A or an A•T to G•C (Gaudelli et al., Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage." Nature (2017); Nishida et al. "Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems." Science 353 (6305) (2016); Komor et al. "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage." Nature 533 (7603) (2016):420-4. A base editing fusion protein may comprise, for example, an active (double strand break creating), partially active (nickase) or deactivated (catalytically inactive) Cas-alpha endonuclease and a deaminase (such as, but not limited to, a cytidine deaminase, an adenine deaminase, APOBEC1, APOBEC3A, BE2, BE3, BE4, ABEs, or the like). Base edit repair

inhibitors and glycosylase inhibitors (*e.g.*, uracil glycosylase inhibitor (to prevent uracil removal)) are contemplated as other components of a base editing system, in some aspects.

[0309] Any of the Cas polypeptides (e.g., catalytically inactive Cas polypeptide) disclosed herein can also be complexed with a base editing molecule via an RNA-aptamer system such any of those described in WO 2021/055459 published 25 March 2021. RNA aptamer systems can include for example (A) RNA motifs such as a (1) telomerase Ku binding motif, (2) telomerase Sm7 binding motif, (3) MS2 phage operator stem-loop, (4) PP7 phage operator stem-loop, (5) SfMu phage Com stem-loop, (6) chemically modified version of the such aptamers or (7) non-natural RNA aptamer and (B) the corresponding aptamer ligand or an RNA-binding section thereof. See also US Patent 11,479,793 and WO 2018/129129 published 12 July 2018.

[0310] The Cas polypeptides and endonucleases described herein can be expressed and purified by methods known in the art, for example as described in WO/2016/186953 published 24 November 2016.

[0311] Many Cas polypeptides and endonucleases have been described to date that can recognize specific PAM sequences (WO2016186953 published 24 November 2016, WO2016186946 published 24 November 2016, and Zetsche B *et al.* 2015. *Cell* 163, 1013) and cleave the target DNA at a specific position. It is understood that based on the methods and aspects described herein utilizing a novel guided Cas system one skilled in the art can now tailor these methods such that they can utilize any guided endonuclease system.

[0312] A Cas effector protein can comprise a heterologous nuclear localization sequence (NLS). A heterologous NLS amino acid sequence herein may be of sufficient strength to drive accumulation of a Cas polypeptide in a detectable amount in the nucleus of a yeast cell herein, for example. An NLS may comprise one (monopartite) or more (e.g., bipartite) short sequences (e.g., 2 to 20 residues) of basic, positively charged residues (e.g., lysine and/or arginine), and can be located anywhere in a Cas amino acid sequence but such that it is exposed on the protein surface. An NLS may be operably linked to the N-terminus or C-terminus of a Cas polypeptide herein, for example. Two or more NLS sequences can be linked to a Cas polypeptide, for example, such as on both the N- and C-termini of a Cas polypeptide. The Cas polypeptide gene can be operably linked to a SV40 nuclear targeting signal upstream of the Cas codon region and a bipartite VirD2 nuclear localization signal (Tinland *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:7442-6)

downstream of the Cas codon region. Non-limiting examples of suitable NLS sequences herein include those disclosed in U.S. Patent Nos. 6,660,830 and 7,309,576.

Guide Polynucleotides

[0313] The guide polynucleotide enables target recognition, binding, and optionally cleavage by the Cas polypeptides, and can be a single molecule or a double molecule. The guide polynucleotide sequence can be a RNA sequence, a DNA sequence, or a combination thereof (a RNA-DNA combination sequence). Optionally, the guide polynucleotide can comprise at least one nucleotide, phosphodiester bond or linkage modification such as, but not limited, to Locked Nucleic Acid (LNA), 5-methyl dC, 2,6-Diaminopurine, 2'-Fluoro A, 2'-Fluoro U, 2'-O-Methyl RNA, phosphorothioate bond, linkage to a cholesterol molecule, linkage to a polyethylene glycol molecule, linkage to a spacer 18 (hexaethylene glycol chain) molecule, or 5' to 3' covalent linkage resulting in circularization. A guide polynucleotide that solely comprises ribonucleic acids is also referred to as a "guide RNA" or "gRNA" (US20150082478 published 19 March 2015 and US20150059010 published 26 February 2015). A guide polynucleotide may be engineered or synthetic.

[0314] A guide polynucleotide includes a chimeric non-naturally occurring guide RNA comprising regions that are not found together in nature (i.e., they are heterologous with each other). For example, a chimeric non-naturally occurring guide RNA comprising a first nucleotide sequence domain (referred to as Variable Targeting domain or VT domain) that can hybridize to a nucleotide sequence in a target DNA, linked to a second nucleotide sequence that can recognize the Cas polypeptide, such that the first and second nucleotide sequence are not found linked together in nature.

[0315] A guide polynucleotide can be a double molecule (also referred to as duplex guide polynucleotide) comprising a crNucleotide sequence (such as a crRNA) and a tracrNucleotide (such as a tracrRNA) sequence. In some cases, there is a linker polynucleotide that connects the crRNA and tracrRNA to form a single guide, for example an sgRNA.

[0316] In some aspects, the crNucleotide includes a first nucleotide sequence domain (referred to as Variable Targeting domain or VT domain) that can hybridize to a nucleotide sequence in a target DNA and a second nucleotide sequence (also referred to as a tracr mate sequence) that is part of a Cas polypeptide recognition (CER) domain. The tracr mate sequence can hybridized to a tracrNucleotide along a region of complementarity and together form the Cas polypeptide

recognition domain or CER domain. The CER domain is capable of interacting with a Cas polypeptide. The crNucleotide and the tracrNucleotide of the duplex guide polynucleotide can be RNA, DNA, and/or RNA-DNA- combination sequences. In some aspects, the crNucleotide molecule of the duplex guide polynucleotide is referred to as "crDNA" (when composed of a contiguous stretch of DNA nucleotides) or "crRNA" (when composed of a contiguous stretch of RNA nucleotides), or "crDNA-RNA" (when composed of a combination of DNA and RNA nucleotides). The crNucleotide can comprise a fragment of the crRNA naturally occurring in Bacteria and Archaea. The size of the fragment of the crRNA naturally occurring in Bacteria and Archaea that can be present in a crNucleotide disclosed herein can range from, but is not limited to, 2, 3, 4, 5, 6, 7, 8, 9,10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides.

[0317] In some aspects, the tracrNucleotide is referred to as "tracrRNA" (when composed of a contiguous stretch of RNA nucleotides) or "tracrDNA" (when composed of a contiguous stretch of DNA nucleotides) or "tracrDNA-RNA" (when composed of a combination of DNA and RNA nucleotides. In one aspect, the RNA that guides the RNA/ Cas9 endonuclease complex is a duplexed RNA comprising a duplex crRNA-tracrRNA. The tracrRNA (trans-activating CRISPR RNA) comprises, in the 5'-to-3' direction, (i) a sequence that anneals with the repeat region of CRISPR type II crRNA and (ii) a stem loop-comprising portion (Deltcheva *et al.*, *Nature* 471:602-607). The duplex guide polynucleotide can form a complex with a Cas polypeptide, wherein said guide polynucleotide/Cas polypeptide complex (also referred to as a guide polynucleotide/Cas polypeptide system) can direct the Cas polypeptide to a genomic target site, enabling the Cas polypeptide to recognize, bind to, and optionally nick or cleave (introduce a single or double-strand break) into the target site. (US20150082478 published 19 March 2015 and US20150059010 published 26 February 2015).

[0318] In some aspects, the guide polynucleotide is a guide polynucleotide capable of forming a PGEN as described herein, wherein said guide polynucleotide comprises a first nucleotide sequence domain that is complementary to a nucleotide sequence in a target DNA, and a second nucleotide sequence domain that interacts with said Cas polypeptide.

[0319] In some aspects, the guide polynucleotide is a guide polynucleotide described herein, wherein each of the first nucleotide sequence and the second nucleotide sequence domain is selected from the group consisting of a DNA sequence, a RNA sequence, and a combination thereof.

[0320] In some aspects,, the guide polynucleotide is a guide polynucleotide described herein, further comprising a second nucleotide sequence domain is selected from the group consisting of RNA backbone modifications that enhance stability, DNA backbone modifications that enhance stability, and a combination thereof (see Kanasty *et al.*, 2013, Common RNA-backbone modifications, *Nature Materials* 12:976-977; US20150082478 published 19 March 2015 and US20150059010 published 26 February 2015)

[0321] The guide RNA can include a dual molecule comprising a chimeric non-naturally occurring crRNA linked to at least one tracrRNA. A chimeric non-naturally occurring crRNA includes a crRNA that comprises regions that are not found together in nature (i.e., they are heterologous with each other. For example, a crRNA comprising a first nucleotide sequence domain (referred to as Variable Targeting domain or VT domain) that can hybridize to a nucleotide sequence in a target DNA, linked to a second nucleotide sequence (also referred to as a tracr mate sequence) such that the first and second sequence are not found linked together in nature.

[0322] The guide polynucleotide can also be a single molecule (also referred to as single guide polynucleotide) comprising a crNucleotide sequence linked to a tracrNucleotide sequence. The single guide polynucleotide comprises a first nucleotide sequence domain (referred to as Variable Targeting domain or VT domain) that can hybridize to a nucleotide sequence in a target DNA and a Cas endonuclease recognition domain (CER domain), that interacts with a Cas polypeptide.

[0323] The VT domain and /or the CER domain of a single guide polynucleotide can comprise a RNA sequence, a DNA sequence, or a RNA-DNA-combination sequence. The single guide polynucleotide being comprised of sequences from the crNucleotide and the tracrNucleotide may be referred to as "single guide RNA" (when composed of a contiguous stretch of RNA nucleotides) or "single guide DNA" (when composed of a contiguous stretch of DNA nucleotides) or "single guide RNA-DNA" (when composed of a combination of RNA and DNA nucleotides). The single guide polynucleotide can form a complex with a Cas polypeptide, wherein said guide polynucleotide/Cas polypeptide complex (also referred to as a guide polynucleotide/Cas polypeptide system) can direct the Cas polypeptide to a genomic target site, enabling the Cas polypeptide to recognize, bind to, and optionally nick or cleave (introduce a single or double-strand break) the target site. (US20150082478 published 19 March 2015 and US20150059010 published 26 February 2015).

[0324] A chimeric non-naturally occurring single guide RNA (sgRNA) includes a sgRNA that comprises regions that are not found together in nature (i.e., they are heterologous with each other. For example, a sgRNA comprising a first nucleotide sequence domain (referred to as Variable Targeting domain or VT domain) that can hybridize to a nucleotide sequence in a target DNA linked to a second nucleotide sequence (also referred to as a tracr mate sequence) that are not found linked together in nature.

[0325] The nucleotide sequence linking the crNucleotide and the tracrNucleotide of a single guide polynucleotide can comprise a RNA sequence, a DNA sequence, or a RNA-DNA combination sequence. In one aspect, the nucleotide sequence linking the crNucleotide and the tracrNucleotide of a single guide polynucleotide (also referred to as "loop") can be at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100 nucleotides in length. In another aspect, the nucleotide sequence linking the crNucleotide and the tracrNucleotide of a single guide polynucleotide can comprise a tetraloop sequence, such as, but not limiting to a GAAA tetraloop sequence.

[0326] The guide polynucleotide can be produced by any method known in the art, including chemically synthesizing guide polynucleotides (such as but not limiting to Hendel *et al.* 2015, *Nature Biotechnology* 33, 985–989), *in vitro* generated guide polynucleotides, and/or self-splicing guide RNAs (such as but not limited to Xie *et al.* 2015, *PNAS* 112:3570-3575).

Guide Polynucleotide/Cas polypeptide Complexes

[0327] A guide polynucleotide/Cas polypeptide complex described herein is capable of recognizing, binding to, and optionally nicking, unwinding, or cleaving (e.g., a Cas endonuclease or a Cas polypeptide having nicking or cleaving activity or a Cas polypeptide having nuclease or endonuclease activity) all or part of a target sequence.

[0328] A guide polynucleotide/Cas polypeptide complex that can cleave both strands of a DNA target sequence typically comprises a Cas polypeptide that has all of its endonuclease domains in a functional state (e.g., wild type endonuclease domains or variants thereof retaining some or all activity in each endonuclease domain). Thus, a wild type Cas polypeptide (e.g., a Cas polypeptide disclosed herein), or a variant thereof retaining some or all activity in each endonuclease domain

of the Cas polypeptide, is a suitable example of a Cas endonuclease that can cleave both strands of a DNA target sequence.

[0329] A guide polynucleotide/Cas endonuclease complex that can cleave one strand of a DNA target sequence can be characterized herein as having nickase activity (e.g., partial cleaving capability). A Cas nickase typically comprises one functional endonuclease domain that allows the Cas to cleave only one strand (i.e., make a nick) of a DNA target sequence. For example, a Cas9 nickase may comprise (i) a mutant, dysfunctional RuvC domain and (ii) a functional HNH domain (e.g., wild type HNH domain). As another example, a Cas9 nickase may comprise (i) a functional RuvC domain (e.g., wild type RuvC domain) and (ii) a mutant, dysfunctional HNH domain. Non-limiting examples of Cas9 nickases suitable for use herein are disclosed in US20140189896 published on 03 July 2014. A pair of Cas nickases can be used to increase the specificity of DNA targeting. In general, this can be done by providing two Cas nickases that, by virtue of being associated with RNA components with different guide sequences, target and nick nearby DNA sequences on opposite strands in the region for desired targeting. Such nearby cleavage of each DNA strand creates a double-strand break (i.e., a DSB with single-stranded overhangs), which is then recognized as a substrate for non-homologous-end-joining, NHEJ (prone to imperfect repair leading to mutations) or homologous recombination, HR. Each nick in these aspects can be at least about 5, between 5 and 10, at least 10, between 10 and 15, at least15, between 15 and 20, at least 20, between 20 and 30, at least 30, between 30 and 40, at least 40, between 40 and 50, at least 50, between 50 and 60, at least 60, between 60 and 70, at least 70, between 70 and 80, at least 80, between 80 and 90, at least 90, between 90 and 100, or 100 or greater (or any integer between 5 and 100) bases apart from each other, for example. One or two Cas nickase proteins herein can be used in a Cas nickase pair. For example, a Cas9 nickase with a mutant RuvC domain, but functioning HNH domain (i.e., Cas9 HNH+/RuvC-), can be used (e.g., *Streptococcus pyogenes* Cas9 HNH+/RuvC-). Each Cas9 nickase (e.g., Cas9 HNH+/RuvC-) can be directed to specific DNA sites nearby each other (up to 100 base pairs apart) by using suitable RNA components herein with guide RNA sequences targeting each nickase to each specific DNA site.

[0330] A guide polynucleotide/Cas polypeptide complex in certain aspects can bind to a DNA target site sequence, but does not cleave any strand at the target site sequence. Such a complex may comprise a Cas polypeptide in which all of its nuclease domains are mutant, dysfunctional. For example, a Cas9 protein that can bind to a DNA target site sequence, but does not cleave any

strand at the target site sequence, may comprise both a mutant, dysfunctional RuvC domain and a mutant, dysfunctional HNH domain. A Cas polypeptide herein that binds, but does not cleave, a target DNA sequence can be used to modulate gene expression, for example, in which case the Cas polypeptide could be fused with a transcription factor (or portion thereof) (e.g., a repressor or activator, such as any of those disclosed herein).

[0331] In some aspects, the guide polynucleotide/Cas endonuclease complex (PGEN) described herein is a PGEN, wherein said Cas endonuclease is optionally covalently or non-covalently linked, or assembled to at least one protein subunit, or functional fragment thereof.

[0332] In some aspects, the guide polynucleotide/Cas endonuclease complex is a guide polynucleotide/Cas endonuclease complex (PGEN) comprising at least one guide polynucleotide and at least one Cas endonuclease polypeptide, wherein said Cas endonuclease polypeptide comprises at least one protein subunit, or a functional fragment thereof, wherein said guide polynucleotide is a chimeric non-naturally occurring guide polynucleotide, wherein said guide polynucleotide/Cas endonuclease complex is capable of recognizing, binding to, and optionally nicking, unwinding, or cleaving all or part of a target sequence.

[0333] The Cas effector protein can be a Cas-alpha effector protein as disclosed herein.

[0334] In some aspects, the guide polynucleotide/Cas effector complex is a guide polynucleotide/Cas effector protein complex (PGEN) comprising at least one guide polynucleotide and a Cas-alpha effector protein, wherein said guide polynucleotide/Cas effector protein complex is capable of recognizing, binding to, and optionally nicking, unwinding, or cleaving all or part of a target sequence.

[0335] The PGEN can be a guide polynucleotide/Cas effector protein complex, wherein said Cas effector protein further comprises one copy or multiple copies of at least one protein subunit, or a functional fragment thereof. In some aspects, said protein subunit is selected from the group consisting of a Cas1 protein subunit, a Cas2 protein subunit, a Cas4 protein subunit, and any combination thereof. The PGEN can be a guide polynucleotide/Cas effector protein complex, wherein said Cas effector protein further comprises at least two different protein subunits of selected from the group consisting of a Cas1, Cas2, and Cas4.

[0336] The PGEN can be a guide polynucleotide/Cas effector protein complex, wherein said Cas effector protein further comprises at least three different protein subunits, or functional fragments

thereof, selected from the group consisting of Cas1, Cas2, and one additional Cas polypeptide, optionally comprising Cas4.

[0337] In some aspects, the guide polynucleotide/Cas effector protein complex described herein is a PGEN, wherein said Cas effector protein is covalently or non-covalently linked to at least one protein subunit, or functional fragment thereof. The PGEN can be a guide polynucleotide/Cas effector protein complex, wherein said Cas effector protein polypeptide is covalently or non-covalently linked, or assembled to one copy or multiple copies of at least one protein subunit, or a functional fragment thereof, selected from the group consisting of a Cas1 protein subunit, a Cas2 protein subunit, a one additional Cas polypeptide optionally comprising Cas4 protein subunit, and any combination thereof. The PGEN can be a guide polynucleotide/Cas effector protein complex, wherein said Cas effector protein is covalently or non-covalently linked or assembled to at least two different protein subunits selected from the group consisting of a Cas1, a Cas2, and one additional Cas polypeptide, optionally comprising Cas4. The PGEN can be a guide polynucleotide/Cas effector protein complex, wherein said Cas effector protein is covalently or non-covalently linked to at least three different protein subunits, or functional fragments thereof, selected from the group consisting of a Cas1, a Cas2, and one additional Cas polypeptide, optionally comprising Cas4, and any combination thereof.

[0338] Any component of the guide polynucleotide/Cas effector protein complex, the guide polynucleotide/Cas effector protein complex itself, as well as the polynucleotide modification template(s) and/or donor DNA(s), can be introduced into a heterologous cell or organism by any method known in the art.

Recombinant Constructs for Transformation of Cells

[0339] The disclosed guide polynucleotides, Cas polypeptides, polynucleotide modification templates, donor DNAs, guide polynucleotide/Cas polypeptide systems disclosed herein, and any one combination thereof, optionally further comprising one or more polynucleotide(s) of interest, can be introduced into a cell. Cells include, but are not limited to, human, non-human, animal, bacterial, fungal, insect, yeast, non-conventional yeast, and plant cells as well as plants and seeds produced by the methods described herein.

[0340] Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook *et al.*, *Molecular Cloning:A Laboratory*

Manual; Cold Spring Harbor Laboratory:Cold Spring Harbor, NY (1989). Transformation methods are well known to those skilled in the art and are described *infra*.

[0341] Vectors and constructs include circular plasmids, and linear polynucleotides, comprising a polynucleotide of interest and optionally other components including linkers, adapters, regulatory or analysis. In some examples a recognition site and/or target site can be comprised within an intron, coding sequence, 5' UTRs, 3' UTRs, and/or regulatory regions.

[0342] ***NHEJ and HDR***

[0001] In some aspects, the Cas polypeptides described herein can be part of a genome editing system further comprising one or more guide polynucleotides and optionally donor DNA, and editing a target polynucleotide sequence comprises nonhomologous end-joining (NHEJ) or homologous recombination (HR) following a Cas polypeptide-mediated double-strand break. Once a double-strand break is induced in the DNA, the cell's DNA repair mechanism is activated to repair the break. The most common repair mechanism to bring the broken ends together is the nonhomologous end-joining pathway (Bleuyard et al., (2006) DNA Repair 5:1-12). The structural integrity of chromosomes is typically preserved by the repair, but deletions, insertions, or other rearrangements are possible (Siebert and Puchta, (2002) Plant Cell 14:1121-31; Pacher et al., (2007) Genetics 175:21-9). Alternatively, the double-strand break can be repaired by homologous recombination between homologous DNA sequences. Once the sequence around the double-strand break is altered, for example, by exonuclease activities involved in the maturation of double-strand breaks, gene conversion pathways can restore the original structure if a homologous sequence is available, such as a homologous chromosome in non-dividing somatic cells, or a sister chromatid after DNA replication (Molinier et al., (2004) Plant Cell 16:342-52). Ectopic and/or epigenic DNA sequences may also serve as a DNA repair template for homologous recombination (Puchta, (1999) Genetics 152:1173-81).

[0343] As used herein, "donor DNA" is a DNA construct that comprises a polynucleotide of interest to be inserted into a genomic target site, wherein insertion is mediated by a Cas polypeptide. Once a double-strand break is introduced in the target site by the endonuclease, the first and second regions of homology of the donor DNA can undergo homologous recombination with their corresponding genomic regions of homology resulting in exchange of DNA between the donor and the target genome. As such, the provided methods result in the integration of the

polynucleotide of interest of the donor DNA into the double-strand break in the target site in the plant genome, thereby altering the original target site and producing an altered genomic target site.

[0344] ***Base Editing***

[0345] In some aspects, the Cas polypeptides described herein can be part of a genome editing system further comprising a base editing agent and a plurality of guide polynucleotides and editing a target polynucleotide sequence comprises introducing a plurality of nucleobase edits in the target polynucleotide sequence resulting in a variant nucleotide sequence.

[0346] One or more nucleobases of a target polynucleotide can be chemically altered, in some cases to change the base from one type to another, for example from a Cytosine to a Thymine, or an Adenine to a Guanine. In some aspects, a plurality of bases, for example 2 or more, 5 or more, 10 or more, 20 or more, 30 or more, 40 or more, 50 or more, 60 or more, 70 or more, 80 or more 90 or more, 100 or more, or even greater than 100, 200 or more, up to thousands of bases may be modified or altered, to produce a plant with a plurality of modified bases.

[0347] Any base editing complex, such as a base editing agent associated with an RNA-guided protein, may be used to target and bind to a desired locus in the genome of an organism and chemically modify one or more components of a target polynucleotide.

[0348] Site-specific base conversions can be achieved to engineer one or more nucleotide changes to create one or more edits into the genome. These include for example, a site-specific base edit mediated by an C•G to T•A or an A•T to G•C base editing deaminase enzymes (Gaudelli et al., Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage." Nature (2017); Nishida et al. "Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems." Science 353 (6305) (2016); Komor et al. "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage." Nature 533 (7603) (2016):420-4. A catalytically "dead" or inactive Cas9 (dCas9), for example a catalytically inactive "dead" version of a Cas endonuclease disclosed herein, fused to a cytidine deaminase or an adenine deaminase protein becomes a specific base editor that can alter DNA bases without inducing a DNA break. Base editors convert C->T (or G->A on the opposite strand) or an adenine base editor that would convert adenine to inosine, resulting in an A->G change within an editing window specified by the gRNA. Any molecule that effects a change in a nucleobase is a "base editing agent".

[0349] For many traits of interest, the creation of single double-strand breaks and the subsequent repair via HDR or NHEJ is not ideal for quantitative traits. An observed phenotype includes both genotype effects and environmental effects. The genotype effects further comprise additive effects, dominance effects, and epistatic effects. The probability of no effect per any single edit can be greater than zero, and any single phenotypic effect can be small, depending on the method used and site selected. Double-stranded break repair can additionally be "noisy" and have low repeatability.

[0350] One approach to ameliorate the probability of no effect per edit or small phenotypic effect outcome is to multiplex genome modification, such that a plurality of target sites are modified. Methods to modify a genomic sequence that do not introduce double-strand breaks would allow for single base substitutions. Combining these approaches, multiplexed base editing is beneficial for creating large numbers of genotype edits that can produce observable phenotype modifications. In some cases, dozens or hundreds or thousands of sites can be edited within one or a few generations of an organism.

[0351] A multiplexed approach to base editing in an organism, has the potential to create a plurality of significant phenotypic variations in one or a few generations, with a positive directional bias to the effects. In some aspects, the organism is a plant. A plant or a population of plants with a plurality of edits can be cross-bred to produce progeny plants, some of which will comprise multiple pluralities of edits from the parental lines. In this way, accelerated breeding of desired traits can be accomplished in parallel in one or a few generations, replacing time-consuming traditional sequential crossing and breeding across multiple generations.

[0352] A base editing deaminase, such as a cytidine deaminase or an adenine deaminase, may be fused to an RNA-guided endonuclease that can be deactivated ("dCas", such as a deactivated Cas9) or partially active ("nCas", such as a Cas9 nickase) so that it does not cleave a target site to which it is guided. The dCas forms a functional complex with a guide polynucleotide that shares homology with a polynucleotide sequence at the target site, and is further complexed with the deaminase molecule. The guided Cas polypeptide recognizes and binds to a double-stranded target sequence, opening the double-strand to expose individual bases. In the case of a cytidine deaminase, the deaminase deaminates the cytosine base and creates a uracil. Uracil glycosylase inhibitor (UGI) is provided to prevent the conversion of U back to C. DNA replication or repair mechanisms then convert the Uracil to a thymine (U to T), and subsequent repair of the opposing

base (formerly G in the original G-C pair) to an Adenine, creating a T-A pair. For example, see Komor *et al.* Nature Volume 533, Pages 420-424, 19 May 2016. Base editing performance may be enhanced using RNA aptamer systems, e.g., those discussed in WO 2021/055459 published 25 March 2021; US Patent 11,479,793; and WO 2018/129129 published 12 July 2018.

[0353] ***Prime Editing***

[0354] In some aspects, the Cas polypeptides described herein can be part of a genome editing system further comprising a prime editing agent and a guide polynucleotide and editing a target nucleotide sequence comprises introducing one or more insertions, deletions, or nucleobase swaps in a target nucleotide sequence without generating a double-stranded DNA break. For example, see Anzalone, A.V., Randolph, P.B., Davis, J.R. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157 (2019).

[0355] In some aspects, the prime editing agent is a Cas polypeptide fused to a reverse transcriptase, wherein the Cas polypeptide is modified to nick DNA rather than generating double-strand break. This Cas-polypeptide-reverse transcriptase fusion can also be referred to as a "prime editor" or "PE". In some aspects, the guide polynucleotide comprises a prime editing guide polynucleotide (pegRNA), and is larger than standard sgRNAs commonly used for CRISPR gene editing (e.g., >100 nucleobases). The pegRNA comprises a primer binding sequence (PBS) and a template containing the desired or target RNA sequence at its 3' end.

[0356] During prime editing, the PE:pegRNA complex binds to a target DNA sequence and the modified Cas polypeptide nicks one target DNA strand resulting in a flap. The PBS on the pegRNA binds to the DNA flap and the target RNA sequence is reverse transcribed using the reverse transcriptase. The edited strand is incorporated into the target DNA at the end of the nicked flap, and the target DNA sequence is repaired with the new reverse transcribed DNA.

<u>Other Reverse Transcriptase- Based Genome Modification Systems</u>

[0357] Another option for genome modification through a CRISPR-Cas system that appears to rely on a reverse transcriptase-based method that reverse transcribes the desired genome edit into the complement of the PAM-containing strand target strand DNA (i.e., the target strand), compared to the non-target strand DNA as described in prime editing. An aspect of this RNA encoded DNA replacement of alleles with CRISPR (designated as REDRAW) is disclosed in, e.g., Kim *et al.*, *bioRxiv* 2022.12.13.520319 (2022) and US20210130835A1, the disclosures of which are

incorporated by reference herein to the extent necessary for use with the CRISPR-Cas polypeptides disclosed herein.

Components for Expression and Utilization of Novel CRISPR-Cas Systems in Prokaryotic and Eukaryotic cells

[0358] The disclosure further provides expression constructs for expressing in a prokaryotic or eukaryotic cell/organism a guide RNA/Cas system that is capable of recognizing, binding to, and optionally nicking, unwinding, or cleaving all or part of a target sequence.

[0359] In some aspects, the expression constructs of the disclosure comprise a promoter operably linked to a nucleotide sequence encoding a Cas gene (or plant optimized, including a Cas polypeptide gene described herein) and a promoter operably linked to a guide RNA of the present disclosure. The promoter is capable of driving expression of an operably linked nucleotide sequence in a prokaryotic or eukaryotic cell/organism.

[0360] Nucleotide sequence modification of the guide polynucleotide, VT domain and/or CER domain can be selected from, but not limited to , the group consisting of a 5' cap, a 3' polyadenylated tail, a riboswitch sequence, a stability control sequence, a sequence that forms a dsRNA duplex, a modification or sequence that targets the guide poly nucleotide to a subcellular location, a modification or sequence that provides for tracking , a modification or sequence that provides a binding site for proteins , a Locked Nucleic Acid (LNA), a 5-methyl dC nucleotide, a 2,6-Diaminopurine nucleotide, a 2'-Fluoro A nucleotide, a 2'-Fluoro U nucleotide; a 2'-O-Methyl RNA nucleotide, a phosphorothioate bond, linkage to a cholesterol molecule, linkage to a polyethylene glycol molecule, linkage to a spacer 18 molecule, a 5' to 3' covalent linkage, or any combination thereof. These modifications can result in at least one additional beneficial feature, wherein the additional beneficial feature is selected from the group of a modified or regulated stability, a subcellular targeting, tracking, a fluorescent label, a binding site for a protein or protein complex, modified binding affinity to complementary target sequence, modified resistance to cellular degradation, and increased cellular permeability.

[0361] A method of expressing RNA components such as gRNA in eukaryotic cells for performing Cas9-mediated DNA targeting has been to use RNA polymerase III (Pol III) promoters, which allow for transcription of RNA with precisely defined, unmodified, 5'- and 3'-ends (DiCarlo *et al.*, *Nucleic Acids Res.* 41:4336-4343; Ma *et al.*, *Mol. Ther. Nucleic Acids* 3:e161). This strategy has been successfully applied in cells of several different species including maize and soybean

(US20150082478 published 19 March 2015). Methods for expressing RNA components that do not have a 5' cap have been described (WO2016/025131 published 18 February 2016).

[0362] Various methods and compositions can be employed to obtain a cell or organism having a polynucleotide of interest inserted in a target site for a Cas polypeptide. Such methods can employ homologous recombination (HR) to provide integration of the polynucleotide of interest at the target site. In one method described herein, a polynucleotide of interest is introduced into the organism cell via a donor DNA construct.

[0363] The donor DNA construct further comprises a first and a second region of homology that flank the polynucleotide of interest. The first and second regions of homology of the donor DNA share homology to a first and a second genomic region, respectively, present in or flanking the target site of the cell or organism genome.

[0364] The donor DNA can be tethered to the guide polynucleotide. Tethered donor DNAs can allow for co-localizing target and donor DNA, useful in genome editing, gene insertion, and targeted genome regulation, and can also be useful in targeting post-mitotic cells where function of endogenous HR machinery is expected to be highly diminished (Mali *et al.*, 2013, *Nature Methods* Vol. 10:957-963).

[0365] The amount of homology or sequence identity shared by a target and a donor polynucleotide can vary and includes total lengths and/or regions having unit integral values in the ranges of about 1-20 bp, 20-50 bp, 50-100 bp, 75-150 bp, 100-250 bp, 150-300 bp, 200-400 bp, 250-500 bp, 300-600 bp, 350-750 bp, 400-800 bp, 450-900 bp, 500-1000 bp, 600-1250 bp, 700-1500 bp, 800-1750 bp, 900-2000 bp, 1-2.5 kb, 1.5–3 kb, 2-4 kb, 2.5-5 kb, 3-6 kb, 3.5-7 kb, 4-8 kb, 5-10 kb, or up to and including the total length of the target site. These ranges include every integer within the range, for example, the range of 1-20 bp includes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 bps. The amount of homology can also be described by percent sequence identity over the full aligned length of the two polynucleotides which includes percent sequence identity at least of about 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, between 98% and 99%, 99%, between 99% and 100%, or 100%. Sufficient homology includes any combination of polynucleotide length, global percent sequence identity, and optionally conserved regions of contiguous nucleotides or local percent sequence identity, for example sufficient homology can be described as a region of 75-150 bp having at least

80% sequence identity to a region of the target locus. Sufficient homology can also be described by the predicted ability of two polynucleotides to specifically hybridize under high stringency conditions, see, for example, Sambrook *et al.*, (1989) *Molecular Cloning:A Laboratory Manual*, (Cold Spring Harbor Laboratory Press, NY); *Current Protocols in Molecular Biology*, Ausubel *et al.*, Eds (1994) *Current Protocols*, (Greene Publishing Associates, Inc. and John Wiley & Sons, Inc.); and, Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, (Elsevier, New York).

[0366] The structural similarity between a given genomic region and the corresponding region of homology found on the donor DNA can be any degree of sequence identity that allows for homologous recombination to occur. For example, the amount of homology or sequence identity shared by the "region of homology" of the donor DNA and the "genomic region" of the organism genome can be at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, such that the sequences undergo homologous recombination

[0367] The region of homology on the donor DNA can have homology to any sequence flanking the target site. While in some instances the regions of homology share significant sequence homology to the genomic sequence immediately flanking the target site, it is recognized that the regions of homology can be designed to have sufficient homology to regions that may be further 5' or 3' to the target site. The regions of homology can also have homology with a fragment of the target site along with downstream genomic regions

[0368] In one aspect, the first region of homology further comprises a first fragment of the target site and the second region of homology comprises a second fragment of the target site, wherein the first and second fragments are dissimilar.

Polynucleotides of Interest

[0369] Polynucleotides of interest are further described herein and include polynucleotides reflective of the commercial markets and interests of those involved in the development of the crop. Crops and markets of interest change, and as developing nations open up world markets, new crops and technologies will emerge also. In addition, as our understanding of agronomic traits and characteristics such as yield and heterosis increase, the choice of genes for genetic engineering will change accordingly.

[0370] General categories of polynucleotides of interest include, for example, genes of interest involved in information, such as zinc fingers, those involved in communication, such as kinases, and those involved in housekeeping, such as heat shock proteins. More specific polynucleotides of interest include, but are not limited to, genes involved in traits of agronomic interest such as but not limited to:crop yield, grain quality, crop nutrient content, starch and carbohydrate quality and quantity as well as those affecting kernel size, sucrose loading, protein quality and quantity, nitrogen fixation and/or utilization, fatty acid and oil composition, genes encoding proteins conferring resistance to abiotic stress (such as drought, nitrogen, temperature, salinity, toxic metals or trace elements, or those conferring resistance to toxins such as pesticides and herbicides), genes encoding proteins conferring resistance to biotic stress (such as attacks by fungi, viruses, bacteria, insects, and nematodes, and development of diseases associated with these organisms).

[0371] Agronomically important traits such as oil, starch, and protein content can be genetically altered in addition to using traditional breeding methods. Modifications include increasing content of oleic acid, saturated and unsaturated oils, increasing levels of lysine and sulfur, providing essential amino acids, and also modification of starch. Hordothionin protein modifications are described in U.S. Patent Nos. 5,703,049, 5,885,801, 5,885,802, and 5,990,389.

[0372] Polynucleotide sequences of interest may encode proteins involved in providing disease or pest resistance. By "disease resistance" or "pest resistance" is intended that the plants avoid the harmful symptoms that are the outcome of the plant-pathogen interactions. Pest resistance genes may encode resistance to pests that have great yield drag such as rootworm, cutworm, European Corn Borer, and the like. Disease resistance and insect resistance genes such as lysozymes or cecropins for antibacterial protection, or proteins such as defensins, glucanases or chitinases for antifungal protection, or *Bacillus thuringiensis* endotoxins, protease inhibitors, collagenases, lectins, or glycosidases for controlling nematodes or insects are all examples of useful gene products. Genes encoding disease resistance traits include detoxification genes, such as against fumonisin (U.S. Patent No. 5,792,931); avirulence (avr) and disease resistance (R) genes (Jones *et al.* (1994) *Science* 266:789; Martin *et al.* (1993) *Science* 262:1432; and Mindrinos *et al.* (1994) *Cell* 78:1089); and the like. Insect resistance genes may encode resistance to pests that have great yield drag such as rootworm, cutworm, European Corn Borer, and the like. Such genes include, for example, *Bacillus thuringiensis* toxic protein genes (U.S. Patent Nos. 5,366,892; 5,747,450; 5,736,514; 5,723,756; 5,593,881; and Geiser *et al.* (1986) *Gene* 48:109); and the like.

[0373] An "herbicide resistance protein" or a protein resulting from expression of an "herbicide resistance-encoding nucleic acid molecule" includes proteins that confer upon a cell the ability to tolerate a higher concentration of an herbicide than cells that do not express the protein, or to tolerate a certain concentration of an herbicide for a longer period of time than cells that do not express the protein. Herbicide resistance traits may be introduced into plants by genes coding for resistance to herbicides that act to inhibit the action of acetolactate synthase (ALS, also referred to as acetohydroxyacid synthase, AHAS), in particular the sulfonylurea (UK:sulphonylurea) type herbicides, genes coding for resistance to herbicides that act to inhibit the action of glutamine synthase, such as phosphinothricin or basta (e.g., the bar gene), glyphosate (e.g., the EPSP synthase gene and the GAT gene), HPPD inhibitors (e.g., the HPPD gene) or other such genes known in the art. See, for example, US Patent Nos. 7,626,077, 5,310,667, 5,866,775, 6,225,114, 6,248,876, 7,169,970, 6,867,293, and 9,187,762. The *bar* gene encodes resistance to the herbicide basta, the *nptII* gene encodes resistance to the antibiotics kanamycin and geneticin, and the ALS-gene mutants encode resistance to the herbicide chlorsulfuron.

[0374] Furthermore, it is recognized that the polynucleotide of interest may also comprise antisense sequences complementary to at least a portion of the messenger RNA (mRNA) for a targeted gene sequence of interest. Antisense nucleotides are constructed to hybridize with the corresponding mRNA. Modifications of the antisense sequences may be made as long as the sequences hybridize to and interfere with expression of the corresponding mRNA. In this manner, antisense constructions having 70%, 80%, or 85% sequence identity to the corresponding antisense sequences may be used. Furthermore, portions of the antisense nucleotides may be used to disrupt the expression of the target gene. Generally, sequences of at least 50 nucleotides, 100 nucleotides, 200 nucleotides, or greater may be used.

[0375] In addition, the polynucleotide of interest may also be used in the sense orientation to suppress the expression of endogenous genes in plants. Methods for suppressing gene expression in plants using polynucleotides in the sense orientation are known in the art. The methods generally involve transforming plants with a DNA construct comprising a promoter that drives expression in a plant operably linked to at least a portion of a nucleotide sequence that corresponds to the transcript of the endogenous gene. Typically, such a nucleotide sequence has substantial sequence identity to the sequence of the transcript of the endogenous gene, generally greater than about 65%

sequence identity, about 85% sequence identity, or greater than about 95% sequence identity. See U.S. Patent Nos. 5,283,184 and 5,034,323.

[0376] The polynucleotide of interest can also be a phenotypic marker. A phenotypic marker is screenable or a selectable marker that includes visual markers and selectable markers whether it is a positive or negative selectable marker. Any phenotypic marker can be used. Specifically, a selectable or screenable marker comprises a DNA segment that allows one to identify, or select for or against a molecule or a cell that comprises it, often under particular conditions. These markers can encode an activity, such as, but not limited to, production of RNA, peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds or compositions and the like.

[0377] Examples of selectable markers include, but are not limited to, DNA segments that comprise restriction enzyme sites; DNA segments that encode products which provide resistance against otherwise toxic compounds including antibiotics, such as, spectinomycin, ampicillin, kanamycin, tetracycline, Basta, neomycin phosphotransferase II (NEO) and hygromycin phosphotransferase (HPT)); DNA segments that encode products which are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); DNA segments that encode products which can be readily identified (e.g., phenotypic markers such as β-galactosidase, GUS; fluorescent proteins such as green fluorescent protein (GFP), cyan (CFP), yellow (YFP), red (RFP), and cell surface proteins); the generation of new primer sites for PCR (e.g., the juxtaposition of two DNA sequence not previously juxtaposed), the inclusion of DNA sequences not acted upon or acted upon by a restriction endonuclease or other DNA modifying enzyme, chemical, etc.; and, the inclusion of a DNA sequences required for a specific modification (e.g., methylation) that allows its identification.

[0378] Additional selectable markers include genes that confer resistance to herbicidal compounds, such as sulphonylureas, glufosinate ammonium, bromoxynil, imidazolinones, and 2,4-dichlorophenoxyacetate (2,4-D). See for example, Acetolactase synthase (ALS) for resistance to sulfonylureas, imidazolinones, triazolopyrimidine sulfonamides, pyrimidinylsalicylates and sulphonylaminocarbonyl-triazolinones (Shaner and Singh, 1997, Herbicide Activity:*Toxicol Biochem Mol Biol* 69-110); glyphosate resistant 5-enolpyruvylshikimate-3-phosphate (EPSPS) (Saroha *et al.* 1998, *J. Plant Biochemistry & Biotechnology* Vol 7:65-72);

[0379] Polynucleotides of interest includes genes that can be stacked or used in combination with other traits, such as but not limited to herbicide resistance or any other trait described herein. Polynucleotides of interest and/or traits can be stacked together in a complex trait locus as described in US20130263324 published 03 Oct 2013 and in WO/2013/112686, published 01 August 2013.

[0380] A polypeptide of interest includes any protein or polypeptide that is encoded by a polynucleotide of interest described herein.

[0381] Further provided are methods for identifying at least one plant cell, comprising in its genome, a polynucleotide of interest integrated at the target site. A variety of methods are available for identifying those plant cells with insertion into the genome at or near to the target site. Such methods can be viewed as directly analyzing a target sequence to detect any change in the target sequence, including but not limited to PCR methods, sequencing methods, nuclease digestion, Southern blots, and any combination thereof. See, for example, US20090133152 published 21 May 2009. The method also comprises recovering a plant from the plant cell comprising a polynucleotide of interest integrated into its genome. The plant may be sterile or fertile. It is recognized that any polynucleotide of interest can be provided, integrated into the plant genome at the target site, and expressed in a plant.

Optimization of Sequences for Expression in Plants

[0382] Methods are available in the art for synthesizing plant-preferred genes. See, for example, U.S. Patent Nos. 5,380,831, and 5,436,391, and Murray *et al.* (1989) *Nucleic Acids Res.* 17:477-498. Additional sequence modifications are known to enhance gene expression in a plant host. These include, for example, elimination of: one or more sequences encoding spurious polyadenylation signals, one or more exon-intron splice site signals, one or more transposon-like repeats, and other such well-characterized sequences that may be deleterious to gene expression. The G-C content of the sequence may be adjusted to levels average for a given plant host, as calculated by reference to known genes expressed in the host plant cell. When possible, the sequence is modified to avoid one or more predicted hairpin secondary mRNA structures. Thus, "a plant-optimized nucleotide sequence" of the present disclosure comprises one or more of such sequence modifications.

Expression Elements

[0383] Any polynucleotide encoding a Cas polypeptide or other CRISPR system component disclosed herein may be functionally linked to a heterologous expression element, to facilitate transcription or regulation in a host cell. Such expression elements include but are not limited to: promoter, leader, intron, and terminator. Expression elements may be "minimal" – meaning a shorter sequence derived from a native source, that still functions as an expression regulator or modifier. Alternatively, an expression element may be "optimized" – meaning that its polynucleotide sequence has been altered from its native state in order to function with a more desirable characteristic in a particular host cell (for example, but not limited to, a bacterial promoter may be "maize-optimized" to improve its expression in corn plants). Alternatively, an expression element may be "synthetic" – meaning that it is designed in silico and synthesized for use in a host cell. Synthetic expression elements may be entirely synthetic, or partially synthetic (comprising a fragment of a naturally-occurring polynucleotide sequence).

[0384] It has been shown that certain promoters are able to direct RNA synthesis at a higher rate than others. These are called "strong promoters". Certain other promoters have been shown to direct RNA synthesis at higher levels only in particular types of cells or tissues and are often referred to as "tissue specific promoters", or "tissue-preferred promoters" if the promoters direct RNA synthesis preferably in certain tissues but also in other tissues at reduced levels.

[0385] A plant promoter includes a promoter capable of initiating transcription in a plant cell. For a review of plant promoters, see, Potenza *et al.*, 2004, *In vitro Cell Dev Biol* 40:1-22; Porto *et al.*, 2014, Molecular Biotechnology (2014), 56(1), 38-49.

[0386] Constitutive promoters include, for example, the core CaMV 35S promoter (Odell *et al.*, (1985) *Nature* 313:810-2); rice actin (McElroy *et al.*, (1990) *Plant Cell* 2:163-71); ubiquitin (Christensen *et al.*, (1989) *Plant Mol Biol* 12:619-32; ALS promoter (U.S. Patent No. 5,659,026) and the like.

[0387] Tissue-preferred promoters can be utilized to target enhanced expression within a particular plant tissue. Tissue-preferred promoters include, for example, WO2013103367 published 11 July 2013, Kawamata *et al.*, (1997) *Plant Cell Physiol* 38:792-803; Hansen *et al.*, (1997) *Mol Gen Genet* 254:337-43; Russell *et al.*, (1997) *Transgenic Res* 6:157-68; Rinehart *et al.*, (1996) *Plant Physiol* 112:1331-41; Van Camp *et al.*, (1996) *Plant Physiol* 112:525-35; Canevascini *et al.*, (1996) *Plant Physiol* 112:513-524; Lam, (1994) *Results Probl Cell Differ* 20:181-96; and Guevara-

Garcia *et al.*, (1993) *Plant J* 4:495-505. Leaf-preferred promoters include, for example, Yamamoto *et al.*, (1997) *Plant J* 12:255-65; Kwon *et al.*, (1994) *Plant Physiol* 105:357-67; Yamamoto *et al.*, (1994) *Plant Cell Physiol* 35:773-8; Gotor *et al.*, (1993) *Plant J* 3:509-18; Orozco *et al.*, (1993) *Plant Mol Biol* 23:1129-38; Matsuoka *et al.*, (1993) *Proc. Natl. Acad. Sci. USA* 90:9586-90; Simpson *et al.*, (1958) *EMBO J* 4:2723-9; Timko *et al.*, (1988) *Nature* 318:57-8. Root-preferred promoters include, for example, Hire *et al.*, (1992) *Plant Mol Biol* 20:207-18 (soybean root-specific glutamine synthase gene); Miao *et al.*, (1991) *Plant Cell* 3:11-22 (cytosolic glutamine synthase (GS)); Keller and Baumgartner, (1991) *Plant Cell* 3:1051-61 (root-specific control element in the GRP 1.8 gene of French bean); Sanger *et al.*, (1990) *Plant Mol Biol* 14:433-43 (root-specific promoter of *A. tumefaciens* mannopine synthase (MAS)); Bogusz *et al.*, (1990) *Plant Cell* 2:633-41 (root-specific promoters isolated from *Parasponia andersonii* and *Trema tomentosa*); Leach and Aoyagi, (1991) *Plant Sci* 79:69-76 (*A. rhizogenes* rolC and rolD root-inducing genes); Teeri *et al.*, (1989) *EMBO J* 8:343-50 (*Agrobacterium* wound-induced TR1' and TR2' genes); VfENOD-GRP3 gene promoter (Kuster *et al.*, (1995) *Plant Mol Biol* 29:759-72); and rolB promoter (Capana *et al.*, (1994) *Plant Mol Biol* 25:681-91; phaseolin gene (Murai *et al.*, (1983) *Science* 23:476-82; Sengopta-Gopalen *et al.*, (1988) *Proc. Natl. Acad. Sci. USA* 82:3320-4). See also, U.S. Patent Nos. 5,837,876; 5,750,386; 5,633,363; 5,459,252; 5,401,836; 5,110,732 and 5,023,179.

[0388] Seed-preferred promoters include both seed-specific promoters active during seed development, as well as seed-germinating promoters active during seed germination. See, Thompson *et al.*, (1989) *BioEssays* 10:108. Seed-preferred promoters include, but are not limited to, Cim1 (cytokinin-induced message); cZ19B1 (maize 19 kDa zein); and milps (myo-inositol-1-phosphate synthase); and for example those disclosed in WO2000011177 published 02 March 2000 and U.S. Patent 6,225,529. For dicots, seed-preferred promoters include, but are not limited to, bean β-phaseolin, napin, β-conglycinin, soybean lectin, cruciferin, and the like. For monocots, seed-preferred promoters include, but are not limited to, maize 15 kDa zein, 22 kDa zein, 27 kDa gamma zein, waxy, shrunken 1, shrunken 2, globulin 1, oleosin, and nuc1. See also, WO2000012733 published 09 March 2000, where seed-preferred promoters from *END1* and *END2* genes are disclosed.

[0389] Chemical inducible (regulated) promoters can be used to modulate the expression of a gene in a prokaryotic and eukaryotic cell or organism through the application of an exogenous chemical

regulator. The promoter may be a chemical-inducible promoter, where application of the chemical induces gene expression, or a chemical-repressible promoter, where application of the chemical represses gene expression. Chemical-inducible promoters include, but are not limited to, the maize In2-2 promoter, activated by benzene sulfonamide herbicide safeners (De Veylder *et al.*, (1997) *Plant Cell Physiol* 38:568-77), the maize GST promoter (GST-II-27, WO1993001294 published 21 January 1993), activated by hydrophobic electrophilic compounds used as pre-emergent herbicides, and the tobacco PR-1a promoter (Ono *et al.*, (2004) *Biosci Biotechnol Biochem* 68:803-7) activated by salicylic acid. Other chemical-regulated promoters include steroid-responsive promoters (see, for example, the glucocorticoid-inducible promoter (Schena *et al.*, (1991) *Proc. Natl. Acad. Sci. USA* 88:10421-5; McNellis *et al.*, (1998) *Plant J* 14:247-257); tetracycline-inducible and tetracycline-repressible promoters (Gatz *et al.*, (1991) *Mol Gen Genet* 227:229-37; U.S. Patent Nos. 5,814,618 and 5,789,156).

[0390] Pathogen inducible promoters induced following infection by a pathogen include, but are not limited to those regulating expression of PR proteins, SAR proteins, beta-1,3-glucanase, chitinase, *etc.*

[0391] A stress-inducible promoter includes the RD29A promoter (Kasuga *et al.* (1999) Nature Biotechnol. 17:287-91). One of ordinary skill in the art is familiar with protocols for simulating stress conditions such as drought, osmotic stress, salt stress and temperature stress and for evaluating stress tolerance of plants that have been subjected to simulated or naturally-occurring stress conditions.

[0392] Another example of an inducible promoter useful in plant cells, is the ZmCAS1 promoter, described in US20130312137 published 21 November 2013.

[0393] New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg, (1989) In *The Biochemistry of Plants*, Vol. 115, Stumpf and Conn, eds (New York, NY:Academic Press), pp. 1-82.

Gene Targeting

[0394] The guide polynucleotide/Cas systems described herein can be used for gene targeting.

[0395] In general, DNA targeting can be performed by cleaving one or both strands at a specific polynucleotide sequence in a cell with a Cas polypeptide associated with a suitable polynucleotide component. Once a single or double-strand break is induced in the DNA, the cell's DNA repair

mechanism is activated to repair the break via nonhomologous end-joining (NHEJ) or Homology-Directed Repair (HDR) processes which can lead to modifications at the target site.

[0396] The length of the DNA sequence at the target site can vary, and includes, for example, target sites that are at least 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more than 30 nucleotides in length. It is further possible that the target site can be palindromic, that is, the sequence on one strand reads the same in the opposite direction on the complementary strand. The nick/cleavage site can be within the target sequence or the nick/cleavage site could be outside of the target sequence. In another variation, the cleavage could occur at nucleotide positions immediately opposite each other to produce a blunt end cut or, in other cases, the incisions could be staggered to produce single-stranded overhangs, also called "sticky ends", which can be either 5' overhangs, or 3' overhangs. Active variants of genomic target sites can also be used. Such active variants can comprise at least 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the given target site, wherein the active variants retain biological activity and hence are capable of being recognized and cleaved by an Cas polypeptide.

[0397] Assays to measure the single or double-strand break of a target site by an endonuclease are known in the art and generally measure the overall activity and specificity of the agent on DNA substrates comprising recognition sites.

[0398] A targeting method herein can be performed in such a way that two or more DNA target sites are targeted in the method, for example. Such a method can optionally be characterized as a multiplex method. Two, three, four, five, six, seven, eight, nine, ten, or more target sites can be targeted at the same time in certain aspects. A multiplex method is typically performed by a targeting method herein in which multiple different RNA components are provided, each designed to guide a guide polynucleotide/Cas polypeptide complex to a unique DNA target site.

Gene Editing

[0399] The process for editing a genomic sequence combining DSB and modification templates generally comprises: introducing into a host cell a DSB-inducing agent, or a nucleic acid encoding a DSB-inducing agent, that recognizes a target sequence in the chromosomal sequence and is able to induce a DSB in the genomic sequence, and at least one polynucleotide modification template comprising at least one nucleotide alteration when compared to the nucleotide sequence to be edited. The polynucleotide modification template can further comprise nucleotide sequences

flanking the at least one nucleotide alteration, in which the flanking sequences are substantially homologous to the chromosomal region flanking the DSB. Genome editing using DSB-inducing agents, such as Cas-gRNA complexes, has been described, for example in US20150082478 published on 19 March 2015, WO2015026886 published on 26 February 2015, WO2016007347 published 14 January 2016, and WO/2016/025131 published on 18 February 2016.

[0400] Some uses for guide RNA/Cas polypeptide systems have been described (see for example:US20150082478 A1 published 19 March 2015, WO2015026886 published 26 February 2015, and US20150059010 published 26 February 2015) and include but are not limited to modifying or replacing nucleotide sequences of interest (such as a regulatory elements), insertion of polynucleotides of interest, gene knock-out, gene-knock in, modification of splicing sites and/or introducing alternate splicing sites, modifications of nucleotide sequences encoding a protein of interest, amino acid and/or protein fusions, and gene silencing by expressing an inverted repeat into a gene of interest.

[0401] Proteins may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known. For example, amino acid sequence variants of the protein(s) can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations include, for example, Kunkel, (1985) *Proc. Natl. Acad. Sci. USA* 82:488-92; Kunkel *et al.*, (1987) *Meth Enzymol* 154:367-82; U.S. Patent No. 4,873,192; Walker and Gaastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan Publishing Company, New York) and the references cited therein. Guidance regarding amino acid substitutions not likely to affect biological activity of the protein is found, for example, in the model of Dayhoff *et al.*, (1978) *Atlas of Protein Sequence and Structure* (Natl Biomed Res Found, Washington, D.C.). Conservative substitutions, such as exchanging one amino acid with another having similar properties, may be preferable. Conservative deletions, insertions, and amino acid substitutions are not expected to produce radical changes in the characteristics of the protein, and the effect of any substitution, deletion, insertion, or combination thereof can be evaluated by routine screening assays. Assays for double-strand-break-inducing activity are known and generally measure the overall activity and specificity of the agent on DNA substrates comprising target sites.

[0402] Described herein are methods for genome editing with a Cas polypeptide and complexes with a Cas polypeptide and a guide polynucleotide. Following characterization of the guide RNA

and PAM sequence, components of the endonuclease and associated CRISPR RNA (crRNA) may be utilized to modify chromosomal DNA in other organisms including plants. To facilitate optimal expression and nuclear localization (for eukaryotic cells), the genes comprising the complex may be optimized as described in WO2016186953 published 24 November 2016, and then delivered into cells as DNA expression cassettes by methods known in the art. The components necessary to comprise an active complex may also be delivered as RNA with or without modifications that protect the RNA from degradation or as mRNA capped or uncapped (Zhang, Y. *et al.*, 2016, *Nat. Commun.* 7:12617) or Cas polypeptide guide polynucleotide complexes (WO2017070032 published 27 April 2017), or any combination thereof. Additionally, a part or part(s) of the complex and crRNA may be expressed from a DNA construct while other components are delivered as RNA with or without modifications that protect the RNA from degradation or as mRNA capped or uncapped (Zhang et al. 2016 *Nat. Commun.* 7:12617) or Cas polypeptide guide polynucleotide complexes (WO2017070032 published 27 April 2017) or any combination thereof. To produce crRNAs *in-vivo*, tRNA derived elements may also be used to recruit endogenous RNAses to cleave crRNA transcripts into mature forms capable of guiding the complex to its DNA target site, as described, for example, in WO2017105991 published 22 June 2017. Nickase complexes may be utilized separately or concertedly to generate a single or multiple DNA nicks on one or both DNA strands. Furthermore, the cleavage activity of the Cas endonuclease may be deactivated by altering key catalytic residues in its cleavage domain (Sinkunas, T. *et al.*, 2013, *EMBO J.* 32:385-394) resulting in a RNA guided helicase that may be used to enhance homology directed repair, induce transcriptional activation, or remodel local DNA structures. Moreover, the activity of the Cas cleavage and helicase domains may both be knocked-out and used in combination with other DNA cutting, DNA nicking, DNA binding, transcriptional activation, transcriptional repression, DNA remodeling, DNA deamination, DNA unwinding, DNA recombination enhancing, DNA integration, DNA inversion, and DNA repair agents.

[0403] The transcriptional direction of the tracrRNA for the CRISPR-Cas system (if present) and other components of the CRISPR-Cas system (such as variable targeting domain, crRNA repeat, loop, anti-repeat) can be deduced as described in WO2016186946 published 24 November 2016, and WO2016186953 published 24 November 2016.

[0404] As described herein, once the appropriate guide RNA requirement is established, the PAM preferences for each new system disclosed herein may be examined. If the cleavage complex

results in degradation of the randomized PAM library, the complex can be converted into a nickase by disabling the ATPase dependent helicase activity either through mutagenesis of critical residues or by assembling the reaction in the absence of ATP as described previously (Sinkunas, T. *et al.*, 2013, *EMBO J.* 32:385-394). Two regions of PAM randomization separated by two protospacer targets may be utilized to generate a double-stranded DNA break which may be captured and sequenced to examine the PAM sequences that support cleavage by the respective complex.

[0405] In some aspects, a method for modifying a target site in the genome of a cell comprises introducing into a cell at least one PGEN described herein, and identifying at least one cell that has a modification at said target, wherein the modification at said target site is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, the chemical alteration of at least one nucleotide, and (v) any combination of (i) – (iv).

[0406] The nucleotide to be edited can be located within or outside a target site recognized and cleaved by a Cas endonuclease. In one aspect, the at least one nucleotide modification is not a modification at a target site recognized and cleaved by a Cas endonuclease. In another aspect, there are at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 900 or 1000 nucleotides between the at least one nucleotide to be edited and the genomic target site.

[0407] A knock-out may be produced by an indel (insertion or deletion of nucleotide bases in a target DNA sequence through NHEJ), or by specific removal of sequence that reduces or completely destroys the function of sequence at or near the targeting site.

[0408] A guide polynucleotide/Cas endonuclease induced targeted mutation can occur in a nucleotide sequence that is located within or outside a genomic target site that is recognized and cleaved by the Cas endonuclease.

[0409] In one aspect, the disclosure describes a method for modifying a target site in the genome of a cell, the method comprising introducing into a cell at least one PGEN described herein and at least one donor DNA, wherein said donor DNA comprises a polynucleotide of interest, and optionally, further comprising identifying at least one cell that said polynucleotide of interest integrated in or near said target site.

[0410] In some aspects, the methods disclosed herein may employ homologous recombination (HR) to provide integration of the polynucleotide of interest at the target site.

[0411] Various methods and compositions can be employed to produce a cell or organism having a polynucleotide of interest inserted in a target site via activity of a CRISPR-Cas system component described herein. In one method described herein, a polynucleotide of interest is introduced into the organism cell via a donor DNA construct. As used herein, "donor DNA" is a DNA construct that comprises a polynucleotide of interest to be inserted into a genomic target site of a Cas polypeptide. The donor DNA construct further comprises a first and a second region of homology that flank the polynucleotide of interest. The first and second regions of homology of the donor DNA share homology to a first and a second genomic region, respectively, present in or flanking the target site of the cell or organism genome.

[0412] The donor DNA can be tethered to the guide polynucleotide. Tethered donor DNAs can allow for co-localizing target and donor DNA, useful in genome editing, gene insertion, and targeted genome regulation, and can also be useful in targeting post-mitotic cells where function of endogenous HR machinery is expected to be highly diminished (Mali *et al.*, 2013, *Nature Methods* Vol. 10:957-963).

[0413] The amount of homology or sequence identity shared by a target and a donor polynucleotide can vary and includes total lengths and/or regions having unit integral values in the ranges of about 1-20 bp, 20-50 bp, 50-100 bp, 75-150 bp, 100-250 bp, 150-300 bp, 200-400 bp, 250-500 bp, 300-600 bp, 350-750 bp, 400-800 bp, 450-900 bp, 500-1000 bp, 600-1250 bp, 700-1500 bp, 800-1750 bp, 900-2000 bp, 1-2.5 kb, 1.5–3 kb, 2-4 kb, 2.5-5 kb, 3-6 kb, 3.5-7 kb, 4-8 kb, 5-10 kb, or up to and including the total length of the target site. These ranges include every integer within the range, for example, the range of 1-20 bp includes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 bps. The amount of homology can also be described by percent sequence identity over the full aligned length of the two polynucleotides which includes percent sequence identity of about at least 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100%. Sufficient homology includes any combination of polynucleotide length, global percent sequence identity, and optionally conserved regions of contiguous nucleotides or local percent sequence identity, for example sufficient homology can be described as a region of 75-150 bp having at least 80% sequence identity to a region of the target locus. Sufficient homology can also be described by the predicted ability of two polynucleotides to specifically hybridize under high stringency conditions, see, for example, Sambrook *et al.*,

(1989) *Molecular Cloning:A Laboratory Manual*, (Cold Spring Harbor Laboratory Press, NY); *Current Protocols in Molecular Biology*, Ausubel *et al.*, Eds (1994) Current Protocols, (Greene Publishing Associates, Inc. and John Wiley & Sons, Inc.); and, Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, (Elsevier, New York).

[0414] Episomal DNA molecules can also be ligated into the double-strand break, for example, integration of T-DNAs into chromosomal double-strand breaks (Chilton and Que, (2003) *Plant Physiol* 133:956-65; Salomon and Puchta, (1998) *EMBO J.* 17:6086-95). Once the sequence around the double-strand breaks is altered, for example, by exonuclease activities involved in the maturation of double-strand breaks, gene conversion pathways can restore the original structure if a homologous sequence is available, such as a homologous chromosome in non-dividing somatic cells, or a sister chromatid after DNA replication (Molinier *et al.*, (2004) *Plant Cell* 16:342-52). Ectopic and/or epigenic DNA sequences may also serve as a DNA repair template for homologous recombination (Puchta, (1999) *Genetics* 152:1173-81).

[0415] In one aspect, the disclosure comprises a method for editing a nucleotide sequence in the genome of a cell, the method comprising introducing into at least one PGEN described herein, and a polynucleotide modification template, wherein said polynucleotide modification template comprises at least one nucleotide modification of said nucleotide sequence, and optionally further comprising selecting at least one cell that comprises the edited nucleotide sequence.

[0416] The guide polynucleotide/Cas endonuclease system can be used in combination with at least one polynucleotide modification template to allow for editing (modification) of a genomic nucleotide sequence of interest. (See also US20150082478, published 19 March 2015 and WO2015026886 published 26 February 2015).

[0417] Polynucleotides of interest and/or traits can be stacked together in a complex trait locus as described in WO2012129373 published 27 September 2012, and in WO2013112686, published 01 August 2013. The guide polynucleotide/Cas9 endonuclease system described herein provides for an efficient system to generate double-strand breaks and allows for traits to be stacked in a complex trait locus.

[0418] A guide polynucleotide/Cas system as described herein, mediating gene targeting, can be used in methods for directing heterologous gene insertion and/or for producing complex trait loci comprising multiple heterologous genes in a fashion similar as disclosed in WO2012129373

published 27 September 2012, where instead of using a double-strand break inducing agent to introduce a gene of interest, a guide polynucleotide/Cas system as disclosed herein is used. By inserting independent transgenes within 0.1, 0.2, 0.3, 0.4, 0.5, 1.0, 2, or even 5 centimorgans (cM) from each other, the transgenes can be bred as a single genetic locus (see, for example, US20130263324 published 03 October 2013 or WO2012129373 published 14 March 2013). After selecting a plant comprising a transgene, plants comprising (at least) one transgenes can be crossed to form an F1 that comprises both transgenes. In progeny from these F1 (F2 or BC1) 1/500 progeny would have the two different transgenes recombined onto the same chromosome. The complex locus can then be bred as single genetic locus with both transgene traits. This process can be repeated to stack as many traits as desired.

[0419] Further uses for guide RNA/Cas polypeptide systems have been described (See for example:US20150082478 published 19 March 2015, WO2015026886 published 26 February 2015, US20150059010 published 26 February 2015, WO2016007347 published 14 January 2016, and PCT application WO2016025131 published 18 February 2016) and include but are not limited to modifying or replacing nucleotide sequences of interest (such as a regulatory elements), insertion of polynucleotides of interest, gene knock-out, gene-knock in, modification of splicing sites and/or introducing alternate splicing sites, modifications of nucleotide sequences encoding a protein of interest, amino acid and/or protein fusions, and gene silencing by expressing an inverted repeat into a gene of interest.

[0420] Resulting characteristics from the gene editing compositions and methods described herein may be evaluated. Chromosomal intervals that correlate with a phenotype or trait of interest can be identified. A variety of methods well known in the art are available for identifying chromosomal intervals. The boundaries of such chromosomal intervals are drawn to encompass markers that will be linked to the gene controlling the trait of interest. In other words, the chromosomal interval is drawn such that any marker that lies within that interval (including the terminal markers that define the boundaries of the interval) can be used as a marker for a particular trait. In one aspect, the chromosomal interval comprises at least one QTL, and furthermore, may indeed comprise more than one QTL. Close proximity of multiple QTLs in the same interval may obfuscate the correlation of a particular marker with a particular QTL, as one marker may demonstrate linkage to more than one QTL. Conversely, e.g., if two markers in close proximity show co-segregation with the desired phenotypic trait, it is sometimes unclear if each of those markers identifies the

same QTL or two different QTL. The term "quantitative trait locus" or "QTL" refers to a region of DNA that is associated with the differential expression of a quantitative phenotypic trait in at least one genetic background, e.g., in at least one breeding population. The region of the QTL encompasses or is closely linked to the gene or genes that affect the trait in question. An "allele of a QTL" can comprise multiple genes or other genetic factors within a contiguous genomic region or linkage group, such as a haplotype. An allele of a QTL can denote a haplotype within a specified window wherein said window is a contiguous genomic region that can be defined, and tracked, with a set of one or more polymorphic markers. A haplotype can be defined by the unique fingerprint of alleles at each marker within the specified window.

Introduction of CRISPR-Cas System Components into a Cell

[0421] The methods and compositions described herein do not depend on a particular method for introducing a sequence into an organism or cell, only that the polynucleotide or polypeptide gains access to the interior of at least one cell of the organism. Introducing includes reference to the incorporation of a nucleic acid into a eukaryotic or prokaryotic cell where the nucleic acid may be incorporated into the genome of the cell, and includes reference to the transient (direct) provision of a nucleic acid, protein or polynucleotide-protein complex (PGEN, RGEN) to the cell.

[0422] Methods for introducing polynucleotides or polypeptides or a polynucleotide-protein complex into cells or organisms are known in the art including, but not limited to, microinjection, electroporation, stable transformation methods, transient transformation methods, ballistic particle acceleration (particle bombardment), whiskers mediated transformation, *Agrobacterium*-mediated transformation, direct gene transfer, viral-mediated introduction, transfection, transduction, cell-penetrating peptides, mesoporous silica nanoparticle (MSN)-mediated direct protein delivery, topical applications, sexual crossing , sexual breeding, and any combination thereof.

[0423] For example, the guide polynucleotide (guide RNA, crNucleotide + tracrNucleotide, guide DNA and/or guide RNA-DNA molecule) can be introduced into a cell directly (transiently) as a single stranded or double stranded polynucleotide molecule. The guide RNA (or crRNA + tracrRNA) can also be introduced into a cell indirectly by introducing a recombinant DNA molecule comprising a heterologous nucleic acid fragment encoding the guide RNA (or crRNA + tracrRNA), operably linked to a specific promoter that is capable of transcribing the guide RNA (crRNA+tracrRNA molecules) in said cell. The specific promoter can be, but is not limited to, a RNA polymerase III promoter, which allow for transcription of RNA with precisely defined,

unmodified, 5'- and 3'-ends (Ma *et al.*, 2014, *Mol. Ther. Nucleic Acids* 3:e161; DiCarlo *et al.*, 2013, *Nucleic Acids Res.* 41:4336-4343; WO2015026887, published 26 February 2015). Any promoter capable of transcribing the guide RNA in a cell can be used and includes a heat shock /heat inducible promoter operably linked to a nucleotide sequence encoding the guide RNA.

[0424] Plant cells differ from animal cells (such as human cells), fungal cells (such as yeast cells) and protoplasts, including for example plant cells comprise a plant cell wall which may act as a barrier to the delivery of components.

[0425] Delivery of the Cas polypeptide, and/or the guide RNA, and/or a ribonucleoprotein complex, and/or a polynucleotide encoding any one or more of the preceding, into plant cells can be achieved through methods known in the art, for example but not limited to: *Rhizobiales*-mediated transformation (*e.g.*, *Agrobacterium*, *Ochrobactrum*), particle mediated delivery (particle bombardment), polyethylene glycol (PEG)-mediated transfection (for example to protoplasts), electroporation, cell-penetrating peptides, or mesoporous silica nanoparticle (MSN)-mediated direct protein delivery.

[0426] The Cas polypeptide, such as the Cas polypeptides described herein, can be introduced into a cell by directly introducing the Cas polypeptide itself (referred to as direct delivery of Cas polypeptide), the mRNA encoding the Cas polypeptide, and/ or the guide polynucleotide/Cas polypeptide complex itself, using any method known in the art. The Cas polypeptide can also be introduced into a cell indirectly by introducing a recombinant DNA molecule that encodes the Cas polypeptide. The endonuclease can be introduced into a cell transiently or can be incorporated into the genome of the host cell using any method known in the art. Uptake of the endonuclease and/or the guided polynucleotide into the cell can be facilitated with a Cell Penetrating Peptide (CPP) as described in WO2016073433 published 12 May 2016. Any promoter capable of expressing the Cas polypeptide in a cell can be used and includes a heat shock /heat inducible promoter operably linked to a nucleotide sequence encoding the Cas polypeptide.

[0427] Direct delivery of a polynucleotide modification template into plant cells can be achieved through particle mediated delivery, and any other direct method of delivery, such as but not limiting to, polyethylene glycol (PEG)-mediated transfection to protoplasts, whiskers mediated transformation, electroporation, particle bombardment, cell-penetrating peptides, or mesoporous silica nanoparticle (MSN)-mediated direct protein delivery can be successfully used for delivering a polynucleotide modification template in eukaryotic cells, such as plant cells.

[0428] The donor DNA can be introduced by any means known in the art. The donor DNA may be provided by any transformation method known in the art including, for example, *Agrobacterium*-mediated transformation or biolistic particle bombardment. The donor DNA may be present transiently in the cell or it could be introduced via a viral replicon. In the presence of the Cas endonuclease and the target site, the donor DNA is inserted into the transformed plant's genome.

[0429] Direct delivery of any one of the guided Cas system components can be accompanied by direct delivery (co-delivery) of other mRNAs that can promote the enrichment and/or visualization of cells receiving the guide polynucleotide/Cas polypeptide complex components. For example, direct co-delivery of the guide polynucleotide/Cas polypeptide components (and/or guide polynucleotide/Cas polypeptide complex itself) together with mRNA encoding phenotypic markers (such as but not limiting to transcriptional activators such as CRC (Bruce *et al.* 2000 *The Plant Cell* 12:65-79) can enable the selection and enrichment of cells without the use of an exogenous selectable marker by restoring function to a non-functional gene product as described in WO2017070032 published 27 April 2017.

[0430] Introducing a guide RNA/Cas polypeptide complex described herein, (representing the cleavage ready complex described herein) into a cell includes introducing the individual components of said complex either separately or combined into the cell, and either directly (direct delivery as RNA for the guide and protein for the Cas polypeptide and protein subunits, or functional fragments thereof) or via recombination constructs expressing the components (guide RNA, Cas polypeptide, protein subunits, or functional fragments thereof). Introducing a guide RNA/Cas endonuclease complex (RGEN) into a cell includes introducing the guide RNA/Cas polypeptide complex as a ribonucleotide-protein into the cell. The ribonucleotide-protein can be assembled prior to being introduced into the cell as described herein. The components comprising the guide RNA/Cas endonuclease ribonucleotide protein (at least one Cas endonuclease, at least one guide RNA, at least one protein subunit) can be assembled *in vitro* or assembled by any means known in the art prior to being introduced into a cell (targeted for genome modification as described herein).

[0431] Direct delivery of the RGEN ribonucleoprotein, allows for genome editing at a target site in the genome of a cell which can be followed by rapid degradation of the complex, and only a transient presence of the complex in the cell. This transient presence of the RGEN complex may

lead to reduced off-target effects. In contrast, delivery of RGEN components (guide RNA, Cas9 endonuclease) via plasmid DNA sequences can result in constant expression of RGENs from these plasmids which can intensify off target effects (Cradick, T. J. *et al.* (2013) *Nucleic Acids Res* 41:9584-9592; Fu, Y *et al.* (2014) *Nat. Biotechnol.* 31:822-826).

[0432] Direct delivery can be achieved by combining any one component of the guide RNA/Cas endonuclease complex (RGEN), representing the cleavage ready complex described herein, (such as at least one guide RNA, at least one Cas polypeptide, and optionally one additional protein), with a delivery matrix comprising a microparticle (such as but not limited to of a gold particle, tungsten particle, and silicon carbide whisker particle) (see also WO2017070032 published 27 April 2017). The delivery matrix may comprise any one of the components, such as the Cas endonuclease, that is attached to a solid matrix (*e.g.*, a particle for bombardment).

[0433] In some aspects, the guide polynucleotide/Cas polypeptide complex, is a complex wherein the guide RNA and Cas polypeptide forming the guide RNA/Cas polypeptide complex are introduced into the cell as RNA and protein, respectively.

[0434] In some aspects, the guide polynucleotide/Cas polypeptide complex, is a complex wherein the guide RNA and Cas polypeptide protein and the at least one protein subunit of a complex forming the guide RNA/Cas polypeptide complex are introduced into the cell as RNA and proteins, respectively.

[0435] In some aspects, the guide polynucleotide/Cas endonuclease complex, is a complex wherein the guide RNA and Cas endonuclease protein and the at least one protein subunit of a complex forming the guide RNA /Cas endonuclease complex (cleavage ready complex) are preassembled *in vitro* and introduced into the cell as a ribonucleotide-protein complex.

[0436] Protocols for introducing polynucleotides, polypeptides or polynucleotide-protein complexes (PGEN, RGEN) into eukaryotic cells, such as plants or plant cells are known and include microinjection (Crossway *et al.*, (1986) *Biotechniques* 4:320-34 and U.S. Patent No. 6,300,543), meristem transformation (U.S. Patent No. 5,736,369), electroporation (Riggs *et al.*, (1986) *Proc. Natl. Acad. Sci. USA* 83:5602-6, *Agrobacterium*-mediated transformation (U.S. Patent Nos. 5,563,055 and 5,981,840), whiskers mediated transformation (Ainley *et al.* 2013, *Plant Biotechnology Journal* 11:1126-1134; Shaheen A. and M. Arshad 2011 Properties and Applications of Silicon Carbide (2011), 345-358 Editor(s):Gerhardt, Rosario. Publisher: InTech, Rijeka, Croatia. CODEN:69PQBP; ISBN:978-953-307-201-2), direct gene transfer (Paszkowski

et al., (1984) *EMBO J* 3:2717-22), and ballistic particle acceleration (U.S. Patent Nos. 4,945,050; 5,879,918; 5,886,244; 5,932,782; Tomes *et al.*, (1995) "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment" in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods,* ed. Gamborg & Phillips (Springer-Verlag, Berlin); McCabe *et al.*, (1988) *Biotechnology* 6:923-6; Weissinger *et al.*, (1988) *Ann Rev Genet* 22:421-77; Sanford *et al.*, (1987) *Particulate Science and Technology* 5:27-37 (onion); Christou *et al.*, (1988) *Plant Physiol* 87:671-4 (soybean); Finer and McMullen, (1991) *In vitro Cell Dev Biol* 27P:175-82 (soybean); Singh *et al.*, (1998) *Theor Appl Genet* 96:319-24 (soybean); Datta *et al.*, (1990) *Biotechnology* 8:736-40 (rice); Klein *et al.*, (1988) *Proc. Natl. Acad. Sci. USA* 85:4305-9 (maize); Klein *et al.*, (1988) *Biotechnology* 6:559-63 (maize); U.S. Patent Nos. 5,240,855; 5,322,783 and 5,324,646; Klein *et al.*, (1988) *Plant Physiol* 91:440-4 (maize); Fromm *et al.*, (1990) *Biotechnology* 8:833-9 (maize); Hooykaas-Van Slogteren *et al.*, (1984) *Nature* 311:763-4; U.S. Patent No. 5,736,369 (cereals); Bytebier *et al.*, (1987) *Proc. Natl. Acad. Sci. USA* 84:5345-9 (*Liliaceae*); De Wet *et al.*, (1985) in *The Experimental Manipulation of Ovule Tissues,* ed. Chapman *et al.*, (Longman, New York), pp. 197-209 (pollen); Kaeppler *et al.*, (1990) *Plant Cell Rep* 9:415-8) and Kaeppler *et al.*, (1992) *Theor Appl Genet* 84:560-6 (whisker-mediated transformation); D'Halluin *et al.*, (1992) *Plant Cell* 4:1495-505 (electroporation); Li *et al.*, (1993) *Plant Cell Rep* 12:250-5; Christou and Ford (1995) *Annals Botany* 75:407-13 (rice) and Osjoda *et al.*, (1996) *Nat Biotechnol* 14:745-50 (maize via *Agrobacterium tumefaciens*).

[0437] Alternatively, polynucleotides may be introduced into plant or plant cells by contacting cells or organisms with a virus or viral nucleic acids. Generally, such methods involve incorporating a polynucleotide within a viral DNA or RNA molecule. In some examples a polypeptide of interest may be initially synthesized as part of a viral polyprotein, which is later processed by proteolysis *in vivo* or *in vitro* to produce the desired recombinant protein. Methods for introducing polynucleotides into plants and expressing a protein encoded therein, involving viral DNA or RNA molecules, are known, see, for example, U.S. Patent Nos. 5,889,191, 5,889,190, 5,866,785, 5,589,367 and 5,316,931.

[0438] The polynucleotide or recombinant DNA construct can be provided to or introduced into a prokaryotic and eukaryotic cell or organism using a variety of transient transformation methods. Such transient transformation methods include, but are not limited to, the introduction of the polynucleotide construct directly into the plant.

[0439] Nucleic acids and proteins can be provided to a cell by any method including methods using molecules to facilitate the uptake of anyone or all components of a guided Cas system (protein and/or nucleic acids), such as cell-penetrating peptides and nanocarriers. See also US20110035836 published 10 February 2011, and EP2821486A1 published 07 January 2015.

[0440] Other methods of introducing polynucleotides into a prokaryotic and eukaryotic cell or organism or plant part can be used, including plastid transformation methods, and the methods for introducing polynucleotides into tissues from seedlings or mature seeds.

[0441] Stable transformation is intended to mean that the nucleotide construct introduced into an organism integrates into a genome of the organism and is capable of being inherited by the progeny thereof. Transient transformation is intended to mean that a polynucleotide is introduced into the organism and does not integrate into a genome of the organism or a polypeptide is introduced into an organism. Transient transformation indicates that the introduced composition is only temporarily expressed or present in the organism.

[0442] A variety of methods are available to identify those cells having an altered genome at or near a target site without using a screenable marker phenotype. Such methods can be viewed as directly analyzing a target sequence to detect any change in the target sequence, including but not limited to PCR methods, sequencing methods, nuclease digestion, Southern blots, and any combination thereof.

Cells and Plants

[0443] The presently disclosed polynucleotides and polypeptides can be introduced into a cell. Cells include, but are not limited to, human, non-human, animal, mammalian, bacterial, fungal, insect, yeast, non-conventional yeast, and plant cells as well as plants and seeds produced by the methods described herein. Any plant can be used with the compositions and methods described herein, including monocot and dicot plants, and plant elements.

[0444] Examples of monocot plants that can be used include, but are not limited to, corn (*Zea mays*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), wheat (*Triticum* species, for example *Triticum aestivum*, *Triticum monococcum*), sugarcane (*Saccharum spp.*), oats (*Avena*), barley (*Hordeum*), switchgrass (*Panicum virgatum*), pineapple (*Ananas comosus*), banana (*Musa spp.*), palm, ornamentals, turfgrasses, and other grasses.

[0445] Examples of dicot plants that can be used include, but are not limited to, soybean (*Glycine max*), *Brassica* species (for example but not limited to:oilseed rape or Canola) (*Brassica napus, B. campestris, Brassica rapa, Brassica. juncea),* alfalfa *(Medicago sativa),*), tobacco (*Nicotiana tabacum*), *Arabidopsis* (*Arabidopsis thaliana*), sunflower (*Helianthus annuus*), cotton (*Gossypium arboreum, Gossypium barbadense*), and peanut (*Arachis hypogaea*), tomato (*Solanum lycopersicum*), potato (*Solanum tuberosum*.

[0446] Additional plants that can be used include safflower (*Carthamus tinctorius*), sweet potato (*Ipomoea batatus*), cassava (*Manihot esculenta*), coffee (*Coffea spp.*), coconut (*Cocos nucifera*), citrus trees (*Citrus spp.*), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa spp.*), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), vegetables, ornamentals, and conifers.

[0447] Vegetables that can be used include tomatoes (*Lycopersicon esculentum*), lettuce (*e.g., Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus spp.*), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea (*Rhododendron spp.*), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa spp.*), tulips (*Tulipa spp.*), daffodils (*Narcissus spp.*), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum.

[0448] Conifers that may be used include pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*); Douglas fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow cedar (*Chamaecyparis nootkatensis*).

[0449] In some aspects of the disclosure, a fertile plant is a plant that produces viable male and female gametes and is self-fertile. Such a self-fertile plant can produce a progeny plant without the contribution from any other plant of a gamete and the genetic material comprised therein. Other aspects of the disclosure can involve the use of a plant that is not self-fertile because the plant does

not produce male gametes, or female gametes, or both, that are viable or otherwise capable of fertilization.

[0450] The present disclosure finds use in the breeding of plants comprising one or more introduced traits, or edited genomes.

[0451] A non-limiting example of how two traits can be stacked into the genome at a genetic distance of, for example, 5 cM from each other is described as follows: a first plant comprising a first transgenic target site integrated into a first DSB target site within the genomic window and not having the first genomic locus of interest is crossed to a second transgenic plant, comprising a genomic locus of interest at a different genomic insertion site within the genomic window and the second plant does not comprise the first transgenic target site. About 5% of the plant progeny from this cross will have both the first transgenic target site integrated into a first DSB target site and the first genomic locus of interest integrated at different genomic insertion sites within the genomic window. Progeny plants having both sites in the defined genomic window can be further crossed with a third transgenic plant comprising a second transgenic target site integrated into a second DSB target site and/or a second genomic locus of interest within the defined genomic window and lacking the first transgenic target site and the first genomic locus of interest. Progeny are then selected having the first transgenic target site, the first genomic locus of interest and the second genomic locus of interest integrated at different genomic insertion sites within the genomic window. Such methods can be used to produce a transgenic plant comprising a complex trait locus having at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14,15, 16, 17, 19, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31 or more transgenic target sites integrated into DSB target sites and/or genomic loci of interest integrated at different sites within the genomic window. In such a manner, various complex trait loci can be generated.

Cells and Animals

[0452] The presently disclosed polynucleotides and polypeptides can be introduced into an animal cell. Animal cells can include, but are not limited to: an organism of a phylum including chordates, arthropods, mollusks, annelids, cnidarians, or echinoderms; or an organism of a class including mammals, insects, birds, amphibians, reptiles, or fishes. In some aspects, the animal is human, mouse, C. elegans, rat, fruit fly (*Drosophila* spp.), zebrafish, chicken, dog, cat, guinea pig, hamster, chicken, Japanese ricefish, sea lamprey, pufferfish, tree frog (*e.g.*, *Xenopus* spp.), monkey, or chimpanzee. Particular cell types that are contemplated include haploid cells, diploid cells,

reproductive cells, neurons, muscle cells, endocrine or exocrine cells, epithelial cells, muscle cells, tumor cells, embryonic cells, hematopoietic cells, bone cells, germ cells, somatic cells, stem cells, pluripotent stem cells, induced pluripotent stem cells, progenitor cells, meiotic cells, and mitotic cells. In some aspects, a plurality of cells from an organism may be used.

[0453] The Cas polypeptides disclosed herein may be used to edit the genome of an animal cell in various ways. In some aspects,, it may be desirable to delete one or more nucleotides. In another aspect, it may be desirable to insert one or more nucleotides. In some aspects,, it may be desirable to replace one or more nucleotides. In another aspect, it may be desirable to modify one or more nucleotides via a covalent or non-covalent interaction with another atom or molecule.

[0454] Genome modification via a Cas polypeptide may be used to effect a genotypic and/or phenotypic change on the target organism. Such a change is preferably related to an improved phenotype of interest or a physiologically-important characteristic, the correction of an endogenous defect, or the expression of some type of expression marker. In some aspects, the phenotype of interest or physiologically-important characteristic is related to the overall health, fitness, or fertility of the animal, the ecological fitness of the animal, or the relationship or interaction of the animal with other organisms in its environment. In some aspects, the phenotype of interest or physiologically-important characteristic is selected from the group consisting of: improved general health, disease reversal, disease modification, disease stabilization, disease prevention, treatment of parasitic infections, treatment of viral infections, treatment of retroviral infections, treatment of bacterial infections, treatment of neurological disorders (for example but not limited to: multiple sclerosis), correction of endogenous genetic defects (for example but not limited to: metabolic disorders, Achondroplasia, Alpha-1 Antitrypsin Deficiency, Antiphospholipid Syndrome, Autism, Autosomal Dominant Polycystic Kidney Disease, Barth syndrome, Breast cancer, Charcot-Marie-Tooth, Colon cancer, Cri du chat, Crohn's Disease, Cystic fibrosis, Dercum Disease, Down Syndrome, Duane Syndrome, Duchenne Muscular Dystrophy, Factor V Leiden Thrombophilia, Familial Hypercholesterolemia, Familial Mediterranean Fever, Fragile X Syndrome, Gaucher Disease, Hemochromatosis, Hemophilia, Holoprosencephaly, Huntington's disease, Klinefelter syndrome, Marfan syndrome, Myotonic Dystrophy, Neurofibromatosis, Noonan Syndrome, Osteogenesis Imperfecta, Parkinson's disease, Phenylketonuria, Poland Anomaly, Porphyria, Progeria, Prostate Cancer, Retinitis Pigmentosa, Severe Combined Immunodeficiency (SCID), Sickle cell disease, Skin Cancer, Spinal Muscular

Atrophy, Tay-Sachs, Thalassemia, Trimethylaminuria, Turner Syndrome, Velocardiofacial Syndrome, WAGR Syndrome, and Wilson Disease), treatment of innate immune disorders (for example but not limited to: immunoglobulin subclass deficiencies), treatment of acquired immune disorders (for example but not limited to: AIDS and other HIV-related disorders), treatment of cancer, as well as treatment of diseases, including rare or "orphan" conditions, that have eluded effective treatment options with other methods.

[0455] Cells that have been genetically modified using the compositions or methods disclosed herein may be transplanted to a subject for purposes such as gene therapy, *e.g.* to treat a disease, or as an antiviral, antipathogenic, or anticancer therapeutic, for the production of genetically modified organisms in agriculture, or for biological research.

In vitro Polynucleotide Detection, Binding, and Modification

[0456] The compositions disclosed herein may further be used as compositions for use in *in vitro* methods, in some aspects with isolated polynucleotide sequence(s). Said isolated polynucleotide sequence(s) may comprise one or more target sequence(s) for modification. In some aspects, said isolated polynucleotide sequence(s) may be genomic DNA, a PCR product, or a synthesized oligonucleotide.

Compositions

[0457] Modification of a target sequence may be in the form of a nucleotide insertion, a nucleotide deletion, a nucleotide substitution, the addition of an atom molecule to an existing nucleotide, a nucleotide modification, or the binding of a heterologous polynucleotide or polypeptide to said target sequence. The insertion of one or more nucleotides may be accomplished by the inclusion of a donor polynucleotide in the reaction mixture: said donor polynucleotide is inserted into a double-strand break created by said Cas-alpha ortholog polypeptide. The insertion may be via non-homologous end joining or via homologous recombination.

[0458] In some aspects,, the sequence of the target polynucleotide is known prior to modification, and compared to the sequence(s) of polynucleotide(s) that result from treatment with the Cas-alpha ortholog. In some aspects,, the sequence of the target polynucleotide is not known prior to modification, and the treatment with the Cas-alpha ortholog is used as part of a method to determine the sequence of said target polynucleotide.

[0459] In some aspects, the Cas-alpha ortholog may be selected from the group consisting of: an unmodified wild type Cas-alpha ortholog, a functional Cas-alpha ortholog variant, a functional Cas-alpha ortholog fragment, a fusion protein comprising an active or deactivated Cas-alpha ortholog, a Cas-alpha ortholog further comprising one or more nuclear localization sequences (NLS) on the C-terminus or on the N-terminus or on both the N- and C-termini, a biotinylated Cas-alpha ortholog, a Cas-alpha ortholog nickase, a Cas-alpha ortholog endonuclease, a Cas-alpha ortholog further comprising a Histidine tag, , and a mixture of any two or more of the preceding.

[0460] In some aspects, the Cas-alpha ortholog is a fusion protein further comprising a nuclease domain, a transcriptional activator domain, a transcriptional repressor domain, an epigenetic modification domain, a cleavage domain, a nuclear localization signal, a cell-penetrating domain, a translocation domain, a marker, or a transgene that is heterologous to the target polynucleotide sequence or to the cell from which said target polynucleotide sequence is obtained or derived.

[0461] In some aspects, a plurality of Cas-alpha orthologs may be desired. In some aspects, said plurality may comprise Cas-alpha orthologs derived from different source organisms or from different loci within the same organism. In some aspects, said plurality may comprise Cas-alpha orthologs with different binding specificities to the target polynucleotide. In some aspects, said plurality may comprise Cas-alpha orthologs with different cleavage efficiencies. In some aspects, said plurality may comprise Cas-alpha orthologs with different PAM specificities. In some aspects, said plurality may comprise orthologs of different molecular compositions, *i.e.*, a polynucleotide *Cas-alpha* ortholog and a polypeptide Cas-alpha ortholog.

[0462] The guide polynucleotide may be provided as a single guide RNA (sgRNA), a chimeric molecule comprising a tracrRNA, a chimeric molecule comprising a crRNA, a chimeric RNA-DNA molecule, a DNA molecule, or a polynucleotide comprising one or more chemically modified nucleotides.

[0463] The storage conditions of the Cas-alpha ortholog and/or the guide polynucleotide include parameters for temperature, state of matter, and time. In some aspects, the Cas-alpha ortholog and/or the guide polynucleotide is stored at about -80 degrees Celsius, at about -20 degrees Celsius, at about 4 degrees Celsius, at about 20-25 degrees Celsius, or at about 37 degrees Celsius. In some aspects, the Cas-alpha ortholog and/or the guide polynucleotide is stored as a liquid, a frozen liquid, or as a lyophilized powder. In some aspects, the Cas-alpha ortholog and/or the guide

polynucleotide is stable for at least one day, at least one week, at least one month, at least one year, or even greater than one year.

[0464] Any or all of the possible polynucleotide components of the reaction (*e.g.*, guide polynucleotide, donor polynucleotide, optionally a *Cas-alpha* polynucleotide) may be provided as part of a vector, a construct, a linearized or circularized plasmid, or as part of a chimeric molecule. Each component may be provided to the reaction mixture separately or together. In some aspects, one or more of the polynucleotide components are operably linked to a heterologous noncoding regulatory element that regulates its expression.

[0465] The method for modification of a target polynucleotide comprises combining the minimal elements into a reaction mixture comprising: a Cas-alpha ortholog (or variant, fragment, or other related molecule as described above), a guide polynucleotide comprising a sequence that is substantially complementary to, or selectively hybridizes to, the target polynucleotide sequence of the target polynucleotide, and a target polynucleotide for modification. In some aspects, the Cas-alpha ortholog is provided as a polypeptide. In some aspects, the Cas-alpha ortholog is provided as a *Cas-alpha* ortholog polynucleotide. In some aspects, the guide polynucleotide is provided as an RNA molecule, a DNA molecule, an RNA:DNA hybrid, or a polynucleotide molecule comprising a chemically-modified nucleotide.

[0466] The storage buffer of any one of the components, or the reaction mixture, may be optimized for stability, efficacy, or other parameters. Additional components of the storage buffer or the reaction mixture may include a buffer composition, Tris, EDTA, dithiothreitol (DTT), phosphate-buffered saline (PBS), sodium chloride, magnesium chloride, HEPES, glycerol, BSA, a salt, an emulsifier, a detergent, a chelating agent, a redox reagent, an antibody, nuclease-free water, a proteinase, and/or a viscosity agent. In some aspects, the storage buffer or reaction mixture further comprises a buffer solution with at least one of the following components: HEPES, MgCl2, NaCl, EDTA, a proteinase, Proteinase K, glycerol, nuclease-free water.

[0467] Incubation conditions will vary according to desired outcome. The temperature is preferably at least 10 degrees Celsius, between 10 and 15, at least 15, between 15 and 17, at least 17, between 17 and 20, at least 20, between 20 and 22, at least 22, between 22 and 25, at least 25, between 25 and 27, at least 27, between 27 and 30, at least 30, between 30 and 32, at least 32, between 32 and 35, at least 35, at least 36, at least 37, at least 38, at least 39, at least 40, or even greater than 40 degrees Celsius. The time of incubation is at least 1 minute, at least 2 minutes, at

least 3 minutes, at least 4 minutes, at least 5 minutes, at least 6 minutes, at least 7 minutes, at least 8 minutes, at least 9 minutes, at least 10 minutes, or even greater than 10 minutes.

[0468] The sequence(s) of the polynucleotide(s) in the reaction mixture prior to, during, or after incubation may be determined by any method known in the art. In some aspects, modification of a target polynucleotide may be ascertained by comparing the sequence(s) of the polynucleotide(s) purified from the reaction mixture to the sequence of the target polynucleotide prior to combining with the Cas-alpha ortholog.

[0469] Any one or more of the compositions disclosed herein, useful for *in vitro* or *in vivo* polynucleotide detection, binding, and/or modification, may be comprised within a kit. A kit comprises a Cas-alpha ortholog or a polynucleotide *Cas-alpha* ortholog encoding such, optionally further comprising buffer components to enable efficient storage, and one or more additional compositions that enable the introduction of said Cas-alpha ortholog or *Cas-alpha* ortholog to a heterologous polynucleotide, wherein said Cas-alpha ortholog or *Cas-alpha* ortholog is capable of effecting a modification, addition, deletion, or substitution of at least one nucleotide of said heterologous polynucleotide. In an additional aspect, a Cas-alpha ortholog disclosed herein may be used for the enrichment of one or more polynucleotide target sequences from a mixed pool. In an additional aspect, a Cas-alpha ortholog disclosed herein may be immobilized on a matrix for use in *in vitro* target polynucleotide detection, binding, and/or modification.

[0470] A Cas-alpha endonuclease may be attached, associated with, or affixed to a solid matrix for the purposes of storage, purification, and/or characterization. Examples of a solid matrix include, but are not limited to: a filter, a chromatography resin, an assay plate, a test tube, a cryogenic vial, etc. A Cas-alpha endonuclease may be substantially purified and stored in an appropriate buffer solution, or lyophilized.

Methods of Detection

[0471] Methods of detecting the Cas-alpha endonuclease-guide polynucleotide complex bound to the target polynucleotide may include any known in the art, including but not limited to microscopy, chromatographic separation, electrophoresis, immunoprecipitation, filtration, nanopore separation, microarrays, as well as those described below.

[0472] A DNA Electrophoretic Mobility Shift Assay (EMSA): studies proteins binding to known DNA oligonucleotide probes and assesses the specificity of the interaction. The technique is based on the principle that protein-DNA complexes migrate more slowly than free DNA molecules when

subjected to polyacrylamide or agarose gel electrophoresis. Because the rate of DNA migration is retarded upon protein binding, the assay is also called a gel retardation assay. Adding a protein-specific antibody to the binding components creates an even larger complex (antibody-protein-DNA) which migrates even slower during electrophoresis, this is known as a supershift and can be used to confirm protein identities.

[0473] DNA Pull-down Assays use a DNA probe labelled with a high affinity tag, such as biotin, which allows the probe to be recovered or immobilized. A DNA probe can be complexed with a protein from a cell lysate in a reaction similar to that used in the EMSA and then used to purify the complex using agarose or magnetic beads. The proteins are then eluted from the DNA and detected by Western blot or identified by mass spectrometry. Alternatively, the protein may be labelled with an affinity tag or the DNA-protein complex may be isolated using an antibody against the protein of interest (similar to a supershift assay). In this case, the unknown DNA sequence bound by the protein is detected by Southern blotting or through PCR analysis.

[0474] Reporter assays provide a real-time in vivo read-out of translational activity for a promoter of interest. Reporter genes are fusions of a target promoter DNA sequence and a reporter gene DNA sequence which is customized by the researcher and the DNA sequence codes for a protein with detectable properties like firefly/Renilla luciferase or alkaline phosphatase. These genes produce enzymes only when the promoter of interest is activated. The enzyme, in turn, catalyzes a substrate to produce either light or a color change that can be detected by spectroscopic instrumentation. The signal from the reporter gene is used as an indirect determinant for the translation of endogenous proteins driven from the same promoter.

[0475] Microplate Capture and Detection Assays use immobilized DNA probes to capture specific protein-DNA interactions and confirm protein identities and relative amounts with target specific antibodies. Typically, a DNA probe is immobilized on the surface of 96- or 384-well microplates coated with streptavidin. A cellular extract is prepared and added to allow the binding protein to bind to the oligonucleotide. The extract is then removed and each well is washed several times to remove non-specifically bound proteins. Finally, the protein is detected using a specific antibody labelled for detection. This method can be extremely sensitive , detecting less than 0.2pg of the target protein per well. This method may also be utilized for oligonucleotides labelled with other tags, such as primary amines that can be immobilized on microplates coated with an amine-reactive surface chemistry.

[0476] DNA Footprinting is one of the most widely used methods for obtaining detailed information on the individual nucleotides in protein–DNA complexes, even inside living cells. In such an experiment, chemicals or enzymes are used to modify or digest the DNA molecules. • When sequence specific proteins bind to DNA they can protect the binding sites from modification or digestion. This can subsequently be visualized by denaturing gel electrophoresis, where unprotected DNA is cleaved more or less at random. Therefore it appears as a 'ladder' of bands and the sites protected by proteins have no corresponding bands and look like foot prints in the pattern of bands. The foot prints there by identify specific nucleosides at the protein–DNA binding sites.

[0477] Microscopic techniques include optical, fluorescence, electron, and atomic force microscopy (AFM).

[0478] Chromatin immunoprecipitation analysis (ChIP) causes proteins to bind covalently to their DNA targets, after which they are unlinked and characterized separately.

[0479] Systematic Evolution of Ligands by EXponential enrichment (SELEX) exposes target proteins to a random library of oligonucleotides. Those genes that bind are separated and amplified by PCR.

EXAMPLES

[0480] The following are Examples of specific aspects of the disclosure. The Examples are offered for illustrative purposes only, and are not intended to limit the scope of the disclosure in any way. Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for. Table 1 details used in the Examples and described herein.

Table 1: SEQ ID Nos.

SEQ ID NO.	Sequence
1	MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCAEAGNMCLRDLYNYFSMPKED RISSKDLYNAMYHKTKLLHPELPGKVANQIVNHAKDVWKRNAKLIYRNQISMPTYKITTAP IRLQNNIYKLIKNKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDS SKSYTQGQLQIKKDHQGKWYCIIPYTFPTHETVLDPDKVMGVDLGVAKAVYWAFNSSYKR GCIDGGEIEHFRKMIRARRVSIQNQIKHSGDARKGHGRKRALKPIETLSEKEKNFRDTINHRY ANRIVEAAIKQGCGTIQIENLEGIADTTGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQY TSQRCSMCGYIEKTNRSSQAVFECKQCGYGSRTICINCRHVQVSGDVCEECGGIVKKENVN ADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKECGHIQASGNTCEVCGSTNILKPKKIRKAK
2	MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCGEAGNMCLRDLYNYFSMPKEDRISSKD LYNAMYHKTKLLHPGLPGKVKNQIVNHAKDVWKRNAKLIYRNQISMPTYKITTAPIRLQNNIYKLIK NKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYKQGQLQIKKDH QGKWYCIIPYTFPTHEHVLDPDKVMGVDLGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMIRARRVS IQNQIKHSGDARKGHGRKRALKPIETLSEKESNFRDTINFRYANRIVEAAIKQGCGTIQIENLEGIADR TGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYG SRTNCINCRHVQVSGDVCEECGGIVKKENVNADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKEC GHIQASGNTCEVCGSTNILKPKKIRKAK
3	MAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACSKHLKVAAYCTT QVERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIF IKGKGIANASSVEHYLSDVCYTRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSD NFPIPLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISL LLSTQRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIGEKSAWMLNLSIDVPKID KGVDPSIIGGIDVGVKSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRILLKKNRHKRAGHG AKNKLKPITILTEKSERFRKKLIERWACEIADFFIKNKVGTVQMENLESMKRKEDSYFNIRLRG FWPYAEMQNKIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKC NFKENADYNAALNISNPKLKSTKEEP
4	MIKVYRYEIVKPLDLDWKEFGTILRQLQQETRFALNKATQLAWEWMGFSSDYKDNHGEYPK SKDILGYTNVHGYAYHTIKTKAYRLNSGNLSQTIKRATDRFKAYQKEILRGDMSIPSYKRDIP LDLIKENISVNRMNHGDYIASLSLLSNPAKQEMNVKRKISVIIIVRGAGKTIMDRILSGEYQVS ASQIIHDDRKNKWYLNISYDFEPQTRVLDLNKIMGIDLGVAVAVYMAFQHTPARYKLEGGEI ENFRRQVESRRISMLRQGKYAGGARGGHGRDKRIKPIEQLRDKIANFRDTTNHRYSRYIVDM AIKEGCGTIQMEDLTNIRDIGSRFLQNWTYYDLQQKIIYKAEEAGIKVIKIDPQYTSQRCSECG NIDSGNRIGQAIFKCRACGYEANADYNAARNIAIPNIDKIIAESIK

5	MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCGEAGNMCLRDLYNYFSMPKEDRISSKD LYNAMYHKTKLLHPGLPGAVKNQIVNHAKDVWKRNAKLIYRNQISMPTYKITTAPIRLQNNIYKLIK NKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYKQGQLQIKKDH QGKWYCIIPYTFPTHEHVLDPDKVMGVDLGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMIRARRVS IQNQIKHSGDARKGHGRKRALKPIETLSEKESNFRDTINFRYANRIVEAAIKQGCGTIQIENLEGIADR TGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYG SRTNCINCRHVQVSGDVCEECGGIVKKENVNADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKEC GHIQASGNTCEVCGSTNILKPKKIRKAK
6	MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCGEAGNMCLRDLYNYFSMPKEDRISSKD LYNAMYHKTKLLHPGLPGSVKNQIVNHAKDVWKRNAKLIYRNQISMPTYKITTAPIRLQNNIYKLIK NKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYKQGQLQIKKDH QGKWYCIIPYTFPTHEHVLDPDKVMGVDLGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMIRARRVS IQNQIKHSGDARKGHGRKRALKPIETLSEKESNFRDTINFRYANRIVEAAIKQGCGTIQIENLEGIADR TGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYG SRTNCINCRHVQVSGDVCEECGGIVKKENVNADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKEC GHIQASGNTCEVCGSTNILKPKKIRKAK
7	MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCGEAGNMCLRDLYNYFSMPKEDRISSKD LYNAMYHKTKLLHPGLPGKVKNQIVRHAKDVWKRNAKLIYRNQISMPTYKITTAPIRLQNNIYKLIK NKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYKQGQLQIKKDH QGKWYCIIPYTFPTHEHVLDPDKVMGVDLGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMIRARRVS IQNQIKHSGDARKGHGRKRALKPIETLSEKESNFRDTINFRYANRIVEAAIKQGCGTIQIENLEGIADR TGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYG SRTNCINCRHVQVSGDVCEECGGIVKKENVNADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKEC GHIQASGNTCEVCGSTNILKPKKIRKAK
8	MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCGEAGNMCLRDLYNYFSMPKEDRISSKD LYNAMYHKTKLLHPGLPGKVKNQIVKHAKDVWKRNAKLIYRNQISMPTYKITTAPIRLQNNIYKLIK NKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYKQGQLQIKKDH QGKWYCIIPYTFPTHEHVLDPDKVMGVDLGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMIRARRVS IQNQIKHSGDARKGHGRKRALKPIETLSEKESNFRDTINFRYANRIVEAAIKQGCGTIQIENLEGIADR TGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYG SRTNCINCRHVQVSGDVCEECGGIVKKENVNADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKEC GHIQASGNTCEVCGSTNILKPKKIRKAK
9	MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCGEAGNMCLRDLYNYFSMPKEDRISSKD LYNAMYHKTKLLHPGLPGKVKNQIVNHAKDVWKRNAKLIYRNQISMPTYKITTAPIRLKNNIYKLIK NKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYKQGQLQIKKDH QGKWYCIIPYTFPTHEHVLDPDKVMGVDLGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMIRARRVS IQNQIKHSGDARKGHGRKRALKPIETLSEKESNFRDTINFRYANRIVEAAIKQGCGTIQIENLEGIADR TGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYG SRTNCINCRHVQVSGDVCEECGGIVKKENVNADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKEC GHIQASGNTCEVCGSTNILKPKKIRKAK
10	MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCGEAGNMCLRDLYNYFSMPKEDRISSKD LYNAMYHKTKLLHPGLPGKVKHQIVNHAKDVWKRNAKLIYRNQISMPTYKITTAPIRLQNNIYKLIK NKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYKQGQLQIKKDH QGKWYCIIPYTFPTHEHVLDPDKVMGVDLGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMIRARRVS IQNQIKHSGDARKGHGRKRALKPIETLSEKESNFRDTINFRYANRIVEAAIKQGCGTIQIENLEGIADR TGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYG SRTNCINCRHVQVSGDVCEECGGIVKKENVNADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKEC GHIQASGNTCEVCGSTNILKPKKIRKAK
11	MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCGEAGNMCLRDLYNYFSMPKEDRISSKD LYNAMYHKTKLLHPGLPGKVKHGIVNHAKDVWKRNAKLIYRNQISMPTYKITTAPIRLQNNIYKLIK NKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYKQGQLQIKKDH QGKWYCIIPYTFPTHEHVLDPDKVMGVDLGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMIRARRVS IQNQIKHSGDARKGHGRKRALKPIETLSEKESNFRDTINFRYANRIVEAAIKQGCGTIQIENLEGIADR TGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYG SRTNCINCRHVQVSGDVCEECGGIVKKENVNADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKEC GHIQASGNTCEVCGSTNILKPKKIRKAK

12	MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCGEAGNMCLRDLYNYFSMPKEDRISSKD LYNAMYHKTKLLHPGLPGKVKKQIVNHAKDVWKRNAKLIYRNQISMPTYKITTAPIRLQNNIYKLIK NKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYKQGQLQIKKDH QGKWYCIIPYTFPTHEHVLDPDKVMGVDLGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMIRARRVS IQNQIKHSGDARKGHGRKRALKPIETLSEKESNFRDTINFRYANRIVEAAIKQGCGTIQIENLEGIADR TGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYG SRTNCINCRHVQVSGDVCEECGGIVKKENVNADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKEC GHIQASGNTCEVCGSTNILKPKKIRKAK
13	MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCGEAGNMCLRDLYNYFSMPKEDRISSKD LYNAMYHKTKLLHPGLPGKVKKGIVNHAKDVWKRNAKLIYRNQISMPTYKITTAPIRLQNNIYKLIK NKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYKQGQLQIKKDH QGKWYCIIPYTFPTHEHVLDPDKVMGVDLGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMIRARRVS IQNQIKHSGDARKGHGRKRALKPIETLSEKESNFRDTINFRYANRIVEAAIKQGCGTIQIENLEGIADR TGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYG SRTNCINCRHVQVSGDVCEECGGIVKKENVNADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKEC GHIQASGNTCEVCGSTNILKPKKIRKAK
14	MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCGEAGNMCLRDLYNYFSMPKEDRISSKD LYNAMYHKTKLLHPGLPGKVKQQIVNHAKDVWKRNAKLIYRNQISMPTYKITTAPIRLQNNIYKLIK NKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYKQGQLQIKKDH QGKWYCIIPYTFPTHEHVLDPDKVMGVDLGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMIRARRVS IQNQIKHSGDARKGHGRKRALKPIETLSEKESNFRDTINFRYANRIVEAAIKQGCGTIQIENLEGIADR TGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYG SRTNCINCRHVQVSGDVCEECGGIVKKENVNADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKEC GHIQASGNTCEVCGSTNILKPKKIRKAK
15	MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCGEAGNMCLRDLYNYFSMPKEDRISSKD LYNAMYHKTKLLHPGLPGKVKQGIVNHAKDVWKRNAKLIYRNQISMPTYKITTAPIRLQNNIYKLIK NKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYKQGQLQIKKDH QGKWYCIIPYTFPTHEHVLDPDKVMGVDLGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMIRARRVS IQNQIKHSGDARKGHGRKRALKPIETLSEKESNFRDTINFRYANRIVEAAIKQGCGTIQIENLEGIADR TGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYG SRTNCINCRHVQVSGDVCEECGGIVKKENVNADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKEC GHIQASGNTCEVCGSTNILKPKKIRKAK
16	MGESVKAIKLKILDMFLDPECTKQDDNWRKDLSTMSRFCGEAGNMCLRDLYNYFSMPKEDRISSKD LYNAMYHKTKLLHPGLPGKVKNQIVNHAKDVWKRNAKLIYRNQISMPTYKITTAPIRLFNNIYKLIK NKNKYIIDVQLYSKEYSKDSGKGTHRYFLVAVRDSSTRMIFDRIMSKDHIDSSKSYKQGQLQIKKDH QGKWYCIIPYTFPTHEHVLDPDKVMGVDLGVAKAVYWAFNSSYKRGCIDGGEIEHFRKMIRARRVS IQNQIKHSGDARKGHGRKRALKPIETLSEKESNFRDTINFRYANRIVEAAIKQGCGTIQIENLEGIADR TGSKFLKNWPYYDLQTKIVNKAKEHGITVVAINPQYTSQRCSMCGYIEKTNRSSQAVFECKQCGYG SRTNCINCRHVQVSGDVCEECGGIVKKENVNADYNAAKNISTPYIDQIIMEKCLELGIPYRSITCKEC GHIQASGNTCEVCGSTNILKPKKIRKAK
17	MNMSKTTISVKLKIIDLSSEKKEFLDNYFNEYAKATTFCQLRIRRLLRNTHWLGKKEKSSKK WIFESGICDLCGENKELVNEDRNSGEPAKICKRCYNGRYGNQMIRKLFVSTKKREVQENMD IRRVAKLNNTHYHRIPEEAFDMIKAADTAEKRRKKNVEYDKKRQMEFIEMFNDEKKRAARP KKPNERETRYVHISKLESPSKGYTLNGIKRKIDGMGKKIERAEKGLSRKKIFGYQGNRIKLDS NWVRFDLAESEITIPSLFKEMKLRITGPTNVHSKSGQIYFAEWFERINKQPNNYCYLIRKTSS NGKYEYYLQYTYEAEVEANKEYAGCLGVDIGCSKLAAAVYYDSKNKKAQKPIEIFTNPIKK IKMRREKLIKLLSRVKVRHRRRKLMQLSKTEPIIDYTCHKTARKIVEMANTAKAFISMENLE TGIKQKQQARETKKQKFYRNMFLFRKLSKLIEYKALLKGIKIVYVKPDYTSQTCSSCGADK EKTERPSQAIFRCLNPTCRYYQRDINADFNAAVNIAKKALNNTEVVTTLL
18	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVCD CCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHICKTCYSGVAGNMFIRKQMY PNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEY LELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKL KEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEY LKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGC IISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFR

	HKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNT FTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNAD LNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDK
19	MKSFKLKLLPTDEQNVLLNEVFCKWASLCTRMASKGHDKERLAPPDSSGNYFNKTQLNQV NTDVTDHMGALEESASQKERAVEKVKRRLKLISDMLSEPNLRDVSQQKPTTFRPLEWVKEG LLKTKYHTVHYWQKECDKLTKQKERMEKTIEKIKKGKITFKPTKMSLHQNCFSLSFGKGTF SMRPFSDTKRGINLDMLTAPIQPAIGKNDGKSSLRSKEFIARNIENYIIFSIHSQLFGLSRSEELL LNAKKEELVAKRDAMLKKKSDSLSKKIKELEKIVGRKITDSERSEIMSQGGKLSSEKFSEDN SYLKTLKVLAKDIIGREELFRLKKYPIVIRKPLNERKKLKNLKPDEWEYYLQLSYDELEKKE FTPKTIMGIDRGLKHILAIAIYDPVQNKFVKNMLIPNPILGWKWKLRKIKRSIQHMERRIRAQ QNAHVPENQLKKRLKSIENKIDYYYHNVSRQILNLAHDFKSAIVVEDLQNMKQHGRKKSK GLRGLNYALSNFDYGKIMGLVKYKAESENVPLLTVLPAGTSQNCAYCLLYGKEQGNYVRN NVNSKIGKCKLHGEIDADINAARTIAICYHKNINEPKPYGERKTFKRK
20	MISLKLKLLPDEEQKKLLDEMFWKWASICTRVGFGRADKEDLKPPKDAEGVWFSLTQLNQ ANTDINDLREAMKHQKHRLEYEKNRLEAQRDDTQDALKNPDRREISTKRKDLFRPKASVEK GFLKLKYHQERYWVRRLKEINKLIERKTKTLIKIEKGRIKFKATRITLHQGSFKIRFGDKPAFL IKALSGKNQIDAPFVVVPEQPICGSVVNSKKYLDEITTNFLAYSVNAMLFGLSRSEEMLLKAK RPEKIKKKEEKLAKKQSAFENKKKELQKLLGRELTQQEEAIIEETRNQFFQDFEVKITKQYSE LLSKIANELKQKNDFLKVNKYPILLRKPLKKAKSKKINNLSPSEWKYYLQFGVKPLLKQKSR RKSRNVLGIDRGLKHLLAVTVLEPDKKTFVWNKLYPNPITGWKWRRRKLLRSLKRLKRRIK SQKHETIHENQTRKKLKSLQGRIDDLLHNISRKIVETAKEYDAVIVVEDLQSMRQHGRSKG NRLKTLNYALSLFDYANVMQLIKYKAGIEGIQIYDVKPAGTSQNCAYCLLAQRDSHEYKRS QENSKIGVCLNPNCQNHKKQIDADLNAARVIASCYALKINDSQPFGTRKRFKKRTTN
21	MEVQKTVMKTLSLRILRPLYSQEIEKEIKEEKERRKQAGGTGELDGGFYKKLEKKHSEMFSF DRLNLLLNQLQREIAKVYNHAISELYIATIAQGNKSNKHYISSIVYNRAYGYFYNAYIALGIC SKVEANFRSNELLTQQSALPTAKSDNFPIVLHKQKGAEGEDGGFRISTEGSDLIFEIPIPFYEY NGENRKEPYKWVKKGGQKPVLKLILSTFRRQRNKGWAKDEGTDAEIRKVTEGKYQVSQIE INRGKKLGEHQKWFANFSIEQPIYERKPNRSIVGGLDVGIRSPLVCAINNSFSRYSVDSNDVF KFSKQVFAFRRRLLSKNSLKRKGHGAAHKLEPITEMTEKNDKFRKKIIERWAKEVTNFFVK NQVGIVQIEDLSTMKDREDHFFNQYLRGFWPYYQMQTLIENKLKEYGIEVKRVQAKYTSQL CSNPNCRYWNNYFNFEYRKVNKFPKFKCEKCNLEISADYNAARNLSTPDIEKFVAKATKGI NLPEK
22	MKYTKVMRYQIIKPLNAEWDELGMVLRDIQKETRAALNKTIQLCWEYQGFSADYKQIHG QYPKPKDVLGYTSMHGYAYDRLKNEFSKIASSNLSQTIKRAVDKWNSDLKEILRGDRSIPN FRKDCPIDIVKQSTKIQKCNDGYVLSLGLINREYKNELGRKNGVFDVLIKANDKTQQTILER IINGDYTYTASQIINHKNKWFINLTYQFETKETALDPNNVMGVDLGIVYPVYIAFNNSLHRY HIKGGEIERFRRQVEKRKRELLNQGKYCGDGRKGHGYATRTKSIESISDKIARFRDTCNHK YSRFIVDMALKHNCGIIQMEDLTGISKESTFLKNWTYYDLQQKIEYKAREAGIQVIKIEPQY TSQRCSKCGYIDKENRQEQATFKCIECGFKTNADYNAARNIAIPNIDKIIRKTLKMQ
23	MVKVVKIHLISEQFDKAGNRIDYEEVNKILWELQKQTREAKNKTVQLLWEWNNFSSDYVK ASGIYPKAKDIFGYSSVHGQANKELRTKLALNSSNLSTTTMDVCKNFNTYKKEVWKGKRS VPSYKSDQPLDLHKDSIKLIYENNEFYVRLALLKKAEFAKYGFKDGFRFKMQVKDNSTKTI LERCFDEVYKINASKLLYDQKKKKWKLNLSYSFDNKNISELDKEKILGVDVGVNCPLVASV FGDRDRFIIKGGEIEKFRKSVEARRRSMLEQTKYCGDGRIGHGRKKRTEPALNIGDKIARFR DTTNHKYSRALIEYAVKKGCGTIQMEKLTGITSKSDRFLKDWTYYDLQTKIENKAKEVGIN VVYIAPKYTSQRCSKCGYIHKDNRPNQAKFRCLECDFESNADYNASQNIGIKNIDKIIEKDL QKQESEVQVNENK
24	MITVRKIKLTIMGDKDTRNSQYKWIRDEQYNQYRALNMGMTYLAVNDILYMNESGLEIRT IKDLKDCEKDIDKNKKEIEKLTARLEKEQNKKNSSSEKLDEIKYKISLVENKIEDYKLKIVEL NKILEETQKERMDIQKEFKEKYVDDLYQVLDKIPFKHLDNKSLVTQRIKADIKSDKSNGLLK GERSIRNYKRNFPLMTRGRDLKFKYDDNDDIEIKWMEGIKFKVILGNRIKNSLELRHTLHKV IEGKYKICDSSLQFDKNNNLILNLTLDIPIDIVNKKVSGRVVGVDLGLKIPAYCALNDVEYIK KSIGRIDDFLKVRTQMQSRRRRLQIAIQSAKGGKGRVNKLQALERFAEKEKNFAKTYNHFL SSNIVKFAVSNQAEQINMELLSLKETQNKSILRNWSYYQLQTMIEYKAQREGIKVKYIDPYH TSQTCSKCGNYEEGQRESQADFICKKCGYKVNADYNAARNIAMSNKYITKKEESKYYKIK ESMV

25	MNFNKCIKVTLIKCLNYDYRKVKQIIKDFQYKYSKAYNMATNYLYLWDTNSMNLKNLYD TKIVDKELLGKSKGAWIENRMNEIIEGALSNNVAQARQDIINKYNKCKKDGLFKGKVSLPTY KLDSKVIVHNVAYKLRNHNGYFIDIGLLNKGKQKELNVGRFEFQIDKLDGNKKATINKIING EYKQGSAQISISKKGKIELIISYSFDKEEIPVLDNNRILGIDLGITNVATMSVYDSIKDEYDYFS WKTNVIRGKELIAFRQKYYNLRRDISIASKTAGKGRCGHGYKTKMKPVDKVRNRIANFADT YNHKISKYIVEFAVKNRCGIIQMEDLSGATSEVHNKMLKDWSYYDLQQKIEYKAKEQGIEIK KVNPKYTSKRCNNCGCIHEDNRDCKNHQARFECKVCGHGKDTDVNADVNASRNIAIPDIDK IIEETEILHSENKPAS
26	MGVTIKIMKYQILCPMNVDWTIFEKQLRNLTYQVRTISNRTIQQLWEFDALSFDYFKERGTY PTVQDLYGCTQKKIDGYIYHTLQSKYPDIHKGNMSTTLQKIIKTWKSRRNEIRKGEMSIPSFR NRIPIDLHNNSVDIIKEKNGDYIAGISLFSRDFHKENGDVPKGKIFVKLGTQKQKSMKVILDR LINQTYSKGACMIHKYKNKWYLSITYKFNAIKENKFDKELIMGIDMGGINTVYFAFNEGFIR SNIKSDEIKMFNERIRQRRINLLKQSKYCSNSRTGKGRTKRLQPIDVLSNKIAKFRNSTNHKY ANYIVKQCLKHNCGRIQMELLKGISKNDKVLKDWTYFDLQEKIKNQAEIYGIEVIKVVPAY TSQRCSQCGYICKENRCTQAMFECKQCGYKTHADYNAAKNISTYDIENIINKQLAVQSKLH SKKCMEEYIEELGYLD
27	MITVRKLKLTIINDDETKRNEQYKFIRDSQYAQYQGLNLAMSVLTNAYLSANRDIKSDLFK ETQKNLKNSSSIFNDIPFGKGIDSKSSITQKVKQDFSIAIKNGLAGGERNITNYKRTFPLMTRG RDLKFSYKDDCSDEIIIKWVNKIVFKVVIGRKDKNYLELMHTLNKVINGEYKVGQSSIYFDK SNKLILNLTLYIPEKKDDDAINGRTLGVDLGIKYPAYVCLNDDTFIRQHIGESLELSKQREQF RNRRKRLQQQLKNVKGGKGREKKLSALDKVAVCERNFVKTYNHTISKRIVDFAKKNKCEF INLEQLTKDGFDNIILSNWSYYELQNMIKYKADREGIKVRYVNPAYTSQKCSKCGYIDKEN RPTQEKFKCIKCGFELNADHNAAINISRLEE

Example 1: Assaying alterations in Cas-alpha PAM recognition

[0481] In this Example, methods to detect alterations in the protospacer adjacent motif (PAM) recognition of RNA-guided DNA-targeting Cas-alpha endonucleases are described. PAM recognition may also be referred to as target adjacent motif or TAM.

[0482] The PAM specificity of a Cas-alpha polypeptide variant containing one or more amino acid substitutions in its PAM interacting (PI) domain was assayed similarly to that described earlier in Karvelis *et al.* (2020) *Nucleic Acids Res.* 48, 5016-5023 and US patent 10,934,536. Briefly, plasmid DNA expression constructs encoding a Cas-alpha variant were synthesized (GenScript) or generated using site-directed mutagenesis (Kunkel (1985) *Proc Natl Acad Sci U.S.A.* 82, 488-492). An example expression cassette is shown in **FIG. 2**. Next, the Cas-alpha endonuclease variant was *in vitro* transcribed and translated using PURExpress (NEB) according to the manufacture's instruction and complexed with its guide RNA (gRNA). The resulting Cas-alpha-gRNA complex was then diluted 10- or 100-fold with reaction buffer (10 mM Tris-HCl, pH 7.5 at 37°C, 100 mM NaCl, 1 mM DTT, 10 mM $MgCl_2$) and incubated at 37°C with the DNA library. Next, library molecules that supported cleavage were end-repaired, captured by adapter ligation, and enriched by PCR. Sequences and indices required for Illumina sequencing were then appended to the amplicons and single read sequenced on a MiSeq System (Illumina). After deep sequencing,

PAM sequences that supported cleavage were collated, normalized to the original library to remove inherent biases, and the resulting PAM recognition quantified using a position frequency matrix (PFM) (Stormo (2013) *Quant. Biol.* 1, 115-130) and visualized as a Weblogo (Crooks (2004) *Genome Res.* 14, 1188-1190).

Example 2: Cas-alpha variants with altered PAM recognition and expanded targeting range

[0483] In this example, methods for altering the PAM recognition of a Cas-alpha polypeptides to increase the number of genomic targets available for genome modification are described.

[0484] In a first method, amino acids from orthologous Cas-alpha nucleases with different PAM recognition were mapped to and substituted into a Cas-alpha 10 protein engineered to enhance cellular activity (SEQ ID NO: 2) (**FIG. 3**). This was accomplished by first determining (for example, but not limited to, using Cryo-EM (Cheng *et al.* (2015) *Cell.* 161, 438-449)) or predicting (for example ,but not limited to, AlphaFold (Jumper *et al.* (2021) *Nature*. 596, 583-589) the structure of the PI domain of the Cas-alpha orthologs and Cas-alpha 10. PI domains were then superimposed and visually inspected using PyMOL (Schrodinger) for similarity to Cas-alpha 10 and those with similar folds selected for further analysis. Amino acids within the PI domain of orthologs implicated as resulting in different PAM recognition were then incorporated into the respective position of Cas-alpha 10 and assayed as described in Example 1 for shifts in PAM recognition.

[0485] In a second method, variation not readily found in Cas-alpha orthologs was introduced into the PI domain of Cas-alpha 10 (**FIG. 4**). Here, amino acids responsible for Cas-alpha 10 PAM recognition were determined or predicted as described above. Next, novel amino acids were substituted into each of these positions individually and evaluated for perturbations in PAM recognition as described in Example 1.

[0486] In a third method, one or more polypeptide chains of an orthologous Cas-alpha endonuclease with different PAM recognition were used to replace one or more structurally similar polypeptide sequences in Cas-alpha 10 (**FIG. 5**). This is first accomplished by determining or predicting the structure of the Cas-alpha 10 and its orthologs as described above. Then, through close structural inspection exchanging one more respective polypeptide chains in Cas-alpha 10 with those of the ortholog shown to confer different PAM specificity. To minimize structural impact, the length of the polypeptide chain can be adjusted by adding or removing residues based

on the structural predictions of the resulting Cas-alpha 10 variant. Alterations in PAM recognition can then be evaluated as described in Example 1.

[0487] In a fourth method, amino acid or polypeptide chain substitutions that alter PAM specificity identified in methods 1-3 were combined and evaluated for additive shifts to PAM recognition in Cas-alpha 10.

[0488] Utilizing the first method , the 5' PAM recognition of Cas-alpha 10 was shifted from 5'-TTC-3' to 5'-TTTC-3', 5'-YTC-3', and 5'-GTTC-3' and the preference for a C bp at position -1 decreased (**Tables 2-7**). This was accomplished by substituting individual amino acids into the PI domain of Cas-alpha 10 (SEQ ID NO: 2) from Cas-alpha 4 (SEQ ID NO: 3) and Cas-alpha 8 (SEQ ID NO: 4) orthologs that have been shown to recognize different PAMs, 5'-YTTN-3' and 5'-TTTR-3', respectively (Karvelis *et al.*, 2020). **Table 8** summarizes the amino acid substitutions made into Cas-alpha 10 and the resulting change in its PAM specificity.

[0489] Variation not readily found in Cas-alpha orthologs was next introduced into Cas-alpha 10 (using the second method) at positions 65, 66, 68, 69, 72, 76, 85, 88, 89, 92, 125, 126, 127, 146, 147, 148, and 193. Amino acids substitutions at positions 72, 85, 88, 89, 92, and 125 were shown to perturb PAM recognition and, in some cases, resulted in novel specificity (**Tables 9-50**). Since positions 88 and 89 were in close proximity, substitutions were incorporated in combination in some instances. New PAM recognition included 5'-GTTC-3', 5'-GTYC-3', 5'-GTWC-3', 5'-DTHC-3', 5'-GTNC-3', 5'-NNNC-3', 5'-GNNC-3', 5'-TTTC-3', 5'-HCTC-3', 5'-YTC-3', 5'-NTCC-3', 5'-TNC-3', 5'-GTTY-3', 5'-DTTY-3', 5'-GTTT-3', 5'-GTHC-3', and a reduced preference for a T at the -2 PAM position. **Table 51** compiles amino acid substitutions introduced into Cas-alpha 10 and the resulting shift in PAM recognition.

[0490] Amino acid substitutions identified above that resulted in altered PAM specificity were next combined and tested for additive effects. These resulted in additional variants with new PAM recognition that included 5'-NYTY-3', 5'-TTTY-3', 5'-TTYTY-3', 5'-TYTY-3', 5'-TTNN-3', 5'-NNCC-3', 5'-NCCC-3', 5'-NTNN-3', 5'-NCTY-3', 5'-NTHC-3', 5'-NNNN-3', and 5'-NNCN-3' (**Tables 52-74**). Some of the tested amino acid substitutions, for example N92L and N92W, a preference for a G-C bp at PAM position -4 was observed. In examining the predicted three-dimensional structure of Cas-alpha 10, it was reasoned that a K85Q substitution could remove this. To test, K85Q was combined with N92L and N92W. As shown in Tables 75 and 76,

K85Q abrogated the -4 G-C bp. **Table 77** summarizes the amino acid combinations that further altered Cas-alpha 10 PAM specificity.

[0491] Several of the amino acid substitutions described herein that modify PAM recognition were next tested in the context of two additional Cas-alpha 10 variants with enhanced cellular activity (SEQ ID NO:88 and 89) (PCT/US23/68329, filed June 13, 2023). All substitutions examined produced a similar shift in PAM specificity (see Tables 29, 40, 58, 78-83).

[0492] In the context of a genome, several of the PAM variants described herein such as K85S (5'-YTC-3' PAM), K85N (5'-YTC-3' PAM), K85Q+N92L (5'-TTY-3' PAM), K85Q+N92W (5'-TTN-3'), K85N+N92L (5'-YTY-3' PAM), and others increased the number of sites available for genome modification compared to the wildtype 5'-TTC-3' PAM (**FIG. 6**). Variants like K85Q+N92W (5'-TTN-3') and K85N+N92L (5'-YTY-3' PAM) were among the most notable expanding the number of potential target sites beyond what is recognized by the 5'-NGG-3' PAM afforded by the Cas9 protein from *Streptococcus pyogenes* in both the corn and human genomes (**FIG. 6**).

[0493] **Tables 2-7**, **9-50, 52-76,** and **78-83** indicate PAM preferences for Cas-alpha 10 as a position frequency matrix (PFM). PAM positions are numbered backward from the first position of the protospacer target with position -1 being immediately 5' of the first position of the protospacer target. Numbers in brackets [x] represent strong PAM preferences, numbers in slashes /x/ represent weak PAM preferences.

Table 2: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	27%	18%	26%	35%	19%	3%	2%
	A	18%	34%	23%	22%	15%	1%	2%
	T	27%	27%	22%	26%	[57%]	[81%]	11%
	C	28%	21%	28%	18%	9%	14%	[85%]
Consensus		**N**	**N**	**N**	**N**	**T**	**T**	**C**

Table 3: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85A

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	27%	15%	19%	11%	5%	13%	4%
	A	18%	33%	21%	16%	15%	5%	6%
	T	27%	31%	34%	[55%]	[67%]	[77%]	16%
	C	28%	21%	26%	18%	13%	4%	[74%]

Consensus		**N**	**N**	**N**	**T**	**T**	**T**	**C**

Table 4: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85S

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	27%	16%	21%	17%	2%	10%	0%
	A	17%	32%	21%	21%	14%	3%	2%
	T	28%	30%	31%	35%	[43%]	[86%]	8%
	C	28%	21%	27%	27%	[42%]	2%	[90%]
Consensus		**N**	**N**	**N**	**N**	**Y**	**T**	**C**

Table 5: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92R

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	27%	18%	28%	/49%/	6%	8%	3%
	A	17%	36%	24%	/20%/	6%	4%	5%
	T	28%	27%	20%	/21%/	[86%]	[77%]	/28%/
	C	28%	20%	28%	10%	2%	11%	[63%]
Consensus		**N**	**N**	**N**	**D**	**T**	**T**	**C**

Table 6: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92K

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	28%	18%	29%	/48%/	4%	4%	5%
	A	17%	35%	24%	/20%/	6%	2%	7%
	T	27%	27%	20%	/23%/	[89%]	[84%]	/31%/
	C	28%	19%	27%	9%	2%	10%	[57%]
Consensus		**N**	**N**	**N**	**D**	**T**	**T**	**C**

Table 7: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Q125K

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	27%	16%	33%	[80%]	4%	2%	1%
	A	18%	41%	28%	10%	5%	6%	2%
	T	29%	26%	14%	7%	[89%]	[75%]	5%
	C	27%	17%	25%	3%	2%	17%	[91%]
Consensus		**N**	**N**	**N**	**G**	**T**	**T**	**C**

Table 8: Summary of Cas-alpha ortholog amino acids substituted into the PI domain of Cas-alpha 10 and the resulting effect on PAM recognition.

Cas-alpha Ortholog	Ortholog PAM Recognition	Substitution (Ortholog aa->Cas-alpha10)	Resulting Cas-alpha 10 PAM Recognition
Cas-alpha 4	5'-TTTR-3'	A156->K85A (SEQ ID NO:5)	5'-TTTC-3'
Cas-alpha 8	5'-YTTN-3'	S88->K85S (SEQ ID NO:6)	5'-YTC-3'
Cas-alpha 4	5'-TTTR-3'	R163->N92R (SEQ ID NO:7)	5'-TTC-3'*
Cas-alpha 8	5'-YTTN-3'	K96->N92K (SEQ ID NO:8)	5'-TTC-3'*
Cas-alpha 8	5'-YTTN-3'	K129->Q125K (SEQ ID NO:9)	5'-GTTC-3'

*Preference for C decreased and preference for T increased

Table 9: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N88H

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	27%	14%	27%	[65%]	6%	9%	3%
	A	18%	41%	28%	14%	13%	14%	7%
	T	28%	26%	17%	14%	[77%]	[52%]	9%
	C	27%	19%	28%	7%	4%	26%	[81%]
Consensus		N	N	N	G	T	T	C

Table 10: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N88H+Q89G

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	25%	16%	25%	[50%]	9%	9%	4%
	A	20%	38%	29%	17%	18%	19%	7%
	T	28%	24%	21%	25%	[68%]	[34%]	10%
	C	27%	22%	25%	8%	5%	[38%]	[79%]
Consensus		N	N	N	G	T	Y	C

Table 11: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N88K

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	27%	16%	33%	[63%]	8%	16%	10%
	A	17%	41%	29%	16%	12%	12%	8%
	T	27%	22%	13%	15%	[76%]	[43%]	11%
	C	29%	20%	24%	7%	4%	[30%]	[71%]
Consensus		N	N	N	G	T	Y	C

Table 12: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N88K+Q89G

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1

Nucleotide	G	27%	15%	29%	[69%]	4%	10%	9%
	A	18%	43%	28%	16%	7%	9%	4%
	T	28%	25%	17%	10%	[87%]	[66%]	5%
	C	27%	16%	26%	5%	2%	16%	[82%]
Consensus		N	N	N	G	T	T	C

Table 13: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N88Q

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	26%	15%	24%	[68%]	5%	5%	0%
	A	19%	39%	24%	14%	4%	[36%]	1%
	T	29%	29%	24%	12%	[89%]	[41%]	5%
	C	26%	17%	28%	5%	2%	18%	[94%]
Consensus		N	N	N	G	T	W	C

Table 14: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N88Q+Q89G

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	26%	14%	21%	[55%]	7%	8%	1%
	A	20%	41%	25%	18%	5%	/46%/	1%
	T	28%	29%	28%	20%	[87%]	/24%/	4%
	C	25%	16%	26%	7%	2%	/22%/	[93%]
Consensus		N	N	N	G	T	H	C

Table 15: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Q125F

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	27%	17%	26%	[57%]	6%	4%	1%
	A	19%	37%	24%	19%	3%	27%	1%
	T	27%	28%	22%	15%	[88%]	[58%]	7%
	C	27%	17%	28%	9%	2%	11%	[92%]
Consensus		N	N	N	G	T	T	C

Table 16: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72V

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	31%	14%	29%	[54%]	23%	17%	6%
	A	16%	36%	26%	18%	14%	13%	9%
	T	26%	27%	17%	15%	[55%]	/43%/	14%
	C	27%	24%	28%	13%	8%	27%	[71%]
Consensus		N	N	N	G	T	N	C

Table 17: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72E

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	32%	16%	26%	/39%/	29%	17%	12%
	A	13%	34%	29%	22%	21%	16%	13%
	T	27%	26%	18%	20%	/37%/	24%	16%
	C	29%	23%	27%	18%	13%	/43%/	[59%]
Consensus		N	N	N	N	N	N	C

Table 18: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72Q

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	31%	17%	28%	/47%/	26%	21%	8%
	A	15%	35%	27%	20%	15%	16%	14%
	T	27%	24%	16%	15%	/49%/	31%	22%
	C	27%	24%	29%	18%	11%	32%	[56%]
Consensus		N	N	N	N	N	N	C

Table 19: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72T

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	28%	13%	27%	[54%]	26%	15%	6%
	A	16%	37%	27%	18%	12%	12%	11%
	T	28%	27%	19%	15%	[54%]	/39%/	15%
	C	27%	23%	28%	13%	8%	/34%/	[68%]
Consensus		N	N	N	G	T	N	C

Table 20: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72C

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	30%	14%	28%	[54%]	27%	18%	5%
	A	15%	36%	28%	18%	13%	12%	11%
	T	28%	27%	15%	14%	[52%]	/38%/	14%
	C	27%	23%	30%	14%	8%	/31%/	[70%]
Consensus		N	N	N	G	T	N	C

Table 21: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72A

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	32%	15%	29%	[53%]	29%	21%	8%
	A	14%	38%	27%	18%	15%	15%	12%
	T	29%	24%	17%	14%	/45%/	28%	15%

	C	25%	23%	28%	15%	11%	/36%/	[65%]
Consensus		**N**	**N**	**N**	**G**	**N**	**N**	**C**

Table 22: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72S

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	29%	13%	27%	[54%]	25%	21%	9%
	A	15%	37%	27%	17%	13%	14%	8%
	T	27%	26%	17%	16%	[53%]	/36%/	14%
	C	29%	24%	28%	13%	9%	29%	[69%]
Consensus		**N**	**N**	**N**	**G**	**T**	**N**	**C**

Table 23: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72P

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	30%	14%	26%	/49%/	26%	19%	8%
	A	15%	34%	30%	19%	15%	17%	17%
	T	28%	28%	16%	16%	/46%/	27%	17%
	C	27%	25%	28%	15%	13%	/38%/	[58%]
Consensus		**N**	**N**	**N**	**N**	**N**	**N**	**C**

Table 24: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72G

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	28%	13%	26%	[56%]	23%	16%	7%
	A	15%	38%	30%	18%	13%	15%	12%
	T	29%	27%	15%	13%	[56%]	/39%/	19%
	C	27%	22%	30%	14%	9%	30%	[63%]
Consensus		**N**	**N**	**N**	**G**	**T**	**N**	**C**

Table 25: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72D

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	31%	15%	26%	/46%/	29%	15%	8%
	A	13%	37%	28%	22%	16%	15%	12%
	T	30%	23%	18%	15%	/44%/	27%	16%
	C	27%	24%	28%	17%	11%	/43%/	[64%]
Consensus		**N**	**N**	**N**	**N**	**N**	**N**	**C**

Table 26: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72L

		PAM Position

		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	30%	12%	26%	[53%]	22%	13%	7%
	A	18%	38%	27%	17%	12%	11%	9%
	T	29%	28%	20%	16%	[55%]	/45%/	14%
	C	23%	21%	28%	14%	10%	30%	[69%]
Consensus		N	N	N	G	T	N	C

Table 27: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85G

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	29%	14%	17%	10%	6%	14%	6%
	A	19%	34%	22%	16%	12%	9%	8%
	T	26%	32%	39%	[59%]	[66%]	[73%]	24%
	C	26%	20%	22%	16%	16%	5%	[61%]
Consensus		N	N	N	T	T	T	C

Table 28: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85D

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	25%	12%	16%	8%	1%	7%	1%
	A	26%	35%	17%	12%	8%	3%	2%
	T	27%	39%	41%	31%	/33%/	[88%]	7%
	C	22%	15%	26%	/49%/	[58%]	2%	[90%]
Consensus		N	N	N	H	C	T	C

Table 29: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85N

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	29%	13%	20%	13%	3%	2%	1%
	A	15%	36%	20%	22%	9%	2%	3%
	T	30%	32%	33%	37%	[46%]	[92%]	12%
	C	26%	19%	27%	28%	[41%]	4%	[84%]
Consensus		N	N	N	N	Y	T	C

Table 30: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N88D

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	29%	15%	22%	28%	21%	6%	6%
	A	17%	39%	27%	20%	7%	11%	8%
	T	28%	29%	22%	29%	[65%]	23%	22%
	C	26%	18%	29%	24%	8%	[61%]	[65%]
Consensus		N	N	N	N	T	C	C

Table 31: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Q89D

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	27%	11%	21%	/47%/	6%	11%	5%
	A	18%	42%	28%	18%	6%	22%	7%
	T	30%	31%	25%	25%	[83%]	/43%/	10%
	C	25%	16%	27%	10%	5%	25%	[78%]
Consensus		N	N	N	N	T	N	C

Table 32: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92D

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	28%	13%	27%	[64%]	3%	3%	5%
	A	19%	41%	26%	15%	4%	3%	7%
	T	28%	29%	17%	14%	[90%]	[88%]	[44%]
	C	25%	17%	31%	7%	3%	6%	[44%]
Consensus		N	N	N	G	T	T	Y

Table 33: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92C

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	29%	15%	28%	[53%]	3%	3%	4%
	A	18%	39%	26%	20%	3%	3%	8%
	T	28%	29%	19%	18%	[91%]	[87%]	[42%]
	C	25%	17%	27%	8%	2%	6%	[46%]
Consensus		N	N	N	G	T	T	Y

Table 34: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92A

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	28%	15%	28%	[54%]	3%	2%	5%
	A	20%	38%	25%	20%	4%	2%	9%
	T	28%	29%	20%	19%	[91%]	[89%]	[41%]
	C	24%	17%	28%	8%	2%	6%	[46%]
Consensus		N	N	N	G	T	T	Y

Table 35: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92G

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	27%	15%	27%	/45%/	4%	3%	3%

	A	20%	37%	25%	22%	4%	2%	7%
	T	28%	30%	21%	23%	[89%]	[87%]	/36%/
	C	25%	18%	27%	9%	2%	8%	[54%]
Consensus		**N**	**N**	**N**	**D**	**T**	**T**	**C**

Table 36: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92F

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	28%	14%	26%	[52%]	4%	3%	8%
	A	18%	39%	26%	22%	4%	3%	10%
	T	28%	29%	21%	18%	[89%]	[87%]	[48%]
	C	25%	17%	26%	9%	2%	6%	[34%]
Consensus		**N**	**N**	**N**	**G**	**T**	**T**	**Y**

Table 37: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92E

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	28%	15%	28%	[62%]	3%	2%	4%
	A	19%	39%	26%	17%	3%	3%	8%
	T	28%	29%	19%	15%	[92%]	[88%]	[39%]
	C	25%	17%	28%	6%	2%	7%	[48%]
Consensus		**N**	**N**	**N**	**G**	**T**	**T**	**Y**

Table 38: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92H

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	28%	14%	28%	[54%]	3%	3%	7%
	A	19%	38%	26%	20%	4%	3%	9%
	T	29%	30%	20%	19%	[91%]	[87%]	[42%]
	C	25%	17%	27%	7%	2%	7%	[42%]
Consensus		**N**	**N**	**N**	**G**	**T**	**T**	**Y**

Table 39: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92I

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	28%	15%	27%	[56%]	4%	3%	4%
	A	19%	39%	26%	20%	3%	3%	6%
	T	29%	29%	20%	18%	[91%]	[86%]	[45%]
	C	24%	17%	27%	7%	2%	7%	[45%]
Consensus		**N**	**N**	**N**	**G**	**T**	**T**	**Y**

Table 40: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92L

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	28%	15%	27%	[50%]	4%	3%	3%
	A	18%	38%	26%	22%	4%	2%	6%
	T	29%	30%	21%	19%	[91%]	[89%]	[45%]
	C	25%	17%	27%	9%	2%	6%	[46%]
Consensus		N	N	N	G	T	T	Y

Table 41: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92V

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	28%	15%	27%	[58%]	3%	3%	4%
	A	18%	39%	26%	19%	3%	3%	7%
	T	29%	30%	20%	16%	[92%]	[89%]	[51%]
	C	25%	16%	27%	7%	2%	6%	[38%]
Consensus		N	N	N	G	T	T	Y

Table 42: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92W

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	28%	15%	27%	[50%]	6%	3%	10%
	A	18%	38%	25%	21%	5%	4%	13%
	T	29%	29%	22%	20%	[87%]	[85%]	/44%/
	C	26%	18%	26%	9%	2%	8%	34%
Consensus		N	N	N	G	T	T	N

Table 43: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92Q

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	28%	14%	28%	[52%]	4%	3%	6%
	A	18%	39%	25%	20%	4%	3%	8%
	T	28%	30%	20%	20%	[91%]	[86%]	[41%]
	C	25%	17%	27%	8%	2%	8%	[45%]
Consensus		N	N	N	G	T	T	Y

Table 44: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92S

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	28%	15%	28%	/49%/	4%	2%	6%
	A	20%	38%	25%	21%	4%	2%	8%
	T	28%	29%	21%	20%	[90%]	[88%]	[37%]

	C	25%	18%	27%	9%	2%	7%	[50%]
Consensus		**N**	**N**	**N**	**D**	**T**	**T**	**Y**

Table 45: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92P

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	29%	17%	28%	[65%]	2%	3%	10%
	A	19%	40%	27%	17%	4%	2%	4%
	T	26%	28%	17%	12%	[91%]	[91%]	[78%]
	C	25%	15%	28%	6%	2%	3%	7%
Consensus		**N**	**N**	**N**	**G**	**T**	**T**	**T**

Table 46: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92Y

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	28%	15%	28%	[51%]	4%	3%	7%
	A	19%	39%	26%	21%	5%	3%	11%
	T	29%	29%	19%	19%	[90%]	[88%]	[50%]
	C	25%	17%	27%	8%	2%	7%	/32%/
Consensus		**N**	**N**	**N**	**G**	**T**	**T**	**T**

Table 47: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92T

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	27%	14%	27%	/48%/	3%	3%	4%
	A	19%	37%	25%	21%	4%	2%	7%
	T	29%	30%	21%	21%	[91%]	[88%]	/36%/
	C	25%	18%	27%	9%	2%	7%	[54%]
Consensus		**N**	**N**	**N**	**D**	**T**	**T**	**C**

Table 48: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 N92M

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	28%	15%	27%	[51%]	4%	3%	4%
	A	19%	38%	25%	21%	4%	2%	7%
	T	28%	30%	20%	19%	[90%]	[88%]	[46%]
	C	25%	17%	27%	8%	2%	6%	[44%]
Consensus		**N**	**N**	**N**	**G**	**T**	**T**	**Y**

Table 49: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Q125R

		PAM Position

		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	27%	14%	31%	[61%]	8%	8%	8%
	A	18%	42%	28%	16%	10%	13%	9%
	T	29%	23%	18%	13%	[70%]	/43%/	19%
	C	26%	20%	23%	10%	12%	36%	[64%]
Consensus		N	N	N	G	T	H	C

Table 50: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Q125P

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	27%	13%	27%	[66%]	7%	5%	3%
	A	19%	40%	26%	15%	6%	31%	5%
	T	30%	27%	20%	12%	[83%]	/43%/	14%
	C	25%	19%	27%	7%	4%	21%	[77%]
Consensus		N	N	N	G	T	H	C

Table 51: Summary of amino acids not readily identified in Cas-alpha orthologs substituted into the PI domain of Cas-alpha 10 and their effect on PAM recognition

Cas-alpha 10 substitution	Resulting Cas-alpha 10 PAM recognition
N88H (SEQ ID NO:10)	5’-GTTC-3’
N88H+Q89G (SEQ ID NO:11)	5’-GTYC-3’
N88K (SEQ ID NO:12)	5’-GTYC-3’
N88K+Q89G (SEQ ID NO:13)	5’-GTTC-3’
N88Q (SEQ ID NO:14)	5’-GTWC-3’
N88Q+Q89G (SEQ ID NO:15)	5’-GTH*C-3’
Q125F (SEQ ID NO:16)	5’-GTTC-3’
Y72V (SEQ ID NO: 28)	5’-GTNC-3’
Y72E (SEQ ID NO: 29)	5’-NNNC-3’
Y72Q (SEQ ID NO:30)	5’-NNNC-3’
Y72T (SEQ ID NO:31)	5’-GTNC-3’
Y72C (SEQ ID NO:32)	5’-GTNC-3’
Y72A (SEQ ID NO:33)	5’-GNNC-3’
Y72S (SEQ ID NO:34)	5’-GTNC-3’
Y72P (SEQ ID NO:35)	5’-NNNC-3’
Y72G (SEQ ID NO:36)	5’-GTNC-3’
Y72D (SEQ ID NO:37)	5’-NNNC-3’
Y72L (SEQ ID NO:38)	5’-GTNC-3’
K85G (SEQ ID NO:39)	5’-TTTC-3’
K85D (SEQ ID NO:40)	5’-HCTC-3’
K85N (SEQ ID NO:41)	5’-NYTC-3’
N88D (SEQ ID NO:42)	5’-NTCC-3’
Q89D (SEQ ID NO:43)	5’-NTNC-3’
N92D (SEQ ID NO:44)	5’-GTTY-3’

N92C (SEQ ID NO:45)	5’-GTTY-3’
N92A (SEQ ID NO:46)	5’-GTTY-3’
N92G (SEQ ID NO:47)	5’-DTTC+-3’
N92F (SEQ ID NO:48)	5’-GTTY-3’
N92E (SEQ ID NO:49)	5’-GTTY-3’
N92H (SEQ ID NO:50)	5’-GTTY-3’
N92I (SEQ ID NO:51)	5’-GTTY-3’
N92L (SEQ ID NO:52)	5’-GTTY-3’
N92V (SEQ ID NO:53)	5’-GTTY-3’
N92W (SEQ ID NO:54)	5’-GTTN-3’
N92Q (SEQ ID NO:55)	5’-GTTY-3’
N92S (SEQ ID NO:56)	5’-DTTY-3’
N92P (SEQ ID NO:57)	5’-GTTT-3’
N92Y (SEQ ID NO:58)	5’-GTTT-3’
N92T (SEQ ID NO:59)	5’-DTTC+-3’
N92M (SEQ ID NO:60)	5’-GTTY-3’
Q125R (SEQ ID NO:61)	5’-GTHC-3’
Q125P (SEQ ID NO:62)	5’-GTHC-3’

*: Preference for A but T and C also accepted

+: Preference for C but T also accepted

Table 52: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85S+N92L

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	27%	11%	13%	10%	1%	6%	1%
	A	22%	40%	19%	18%	6%	4%	7%
	T	28%	36%	/46%/	/50%/	[53%]	[87%]	[48%]
	C	23%	14%	22%	22%	[40%]	3%	[44%]
Consensus		**N**	**N**	**N**	**N**	**Y**	**T**	**Y**

Table 53: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85S+N92Q

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	27%	9%	12%	9%	2%	6%	4%
	A	21%	40%	20%	16%	6%	6%	10%
	T	29%	36%	/47%/	[55%]	[56%]	[83%]	[41%]
	C	23%	14%	21%	20%	/37%/	6%	[46%]
Consensus		N	N	N	T	T	T	Y

Table 54: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85S+N92C

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	27%	9%	12%	10%	2%	6%	2%
	A	22%	40%	19%	16%	5%	6%	9%
	T	28%	37%	/48%/	[54%]	[55%]	[84%]	[41%]
	C	23%	14%	22%	20%	/39%/	4%	[47%]
Consensus		N	N	N	T	T	T	Y

Table 55: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85S+N92H

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	26%	8%	10%	9%	2%	6%	6%
	A	22%	41%	19%	17%	5%	6%	11%
	T	29%	38%	[52%]	[56%]	[54%]	[82%]	[46%]
	C	23%	13%	19%	19%	[39%]	5%	[37%]
Consensus		N	N	T	T	Y	T	Y

Table 56: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85S+N92A

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	26%	9%	12%	9%	1%	6%	4%
	A	23%	40%	18%	17%	5%	4%	12%
	T	29%	37%	/49%/	[52%]	[52%]	[86%]	[41%]
	C	22%	13%	21%	21%	[42%]	4%	[43%]
Consensus		N	N	N	T	Y	T	Y

Table 57: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85S+N92M

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	26%	10%	13%	11%	1%	6%	2%
	A	22%	40%	19%	18%	6%	5%	8%
	T	29%	36%	/46%/	/50%/	[53%]	[84%]	[50%]

	C	23%	14%	22%	22%	[40%]	4%	[40%]
Consensus		**N**	**N**	**N**	**N**	**Y**	**T**	**Y**

Table 58: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85N+N92L

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	26%	13%	15%	12%	1%	3%	2%
	A	22%	40%	18%	19%	6%	2%	6%
	T	29%	33%	40%	/43%/	/38%/	[92%]	[50%]
	C	24%	15%	27%	26%	[55%]	3%	[42%]
Consensus		**N**	**N**	**N**	**N**	**C**	**T**	**Y**

Table 59: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85N+N92H

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	27%	11%	14%	15%	2%	4%	6%
	A	21%	41%	18%	19%	7%	2%	10%
	T	28%	34%	/42%/	/44%/	/36%/	[89%]	[47%]
	C	24%	14%	26%	22%	[55%]	5%	[37%]
Consensus		**N**	**N**	**N**	**N**	**C**	**T**	**Y**

Table 60: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85N+N92A

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	26%	12%	15%	14%	1%	3%	5%
	A	23%	40%	18%	17%	6%	2%	10%
	T	28%	33%	/41%/	/45%/	/35%/	[92%]	[43%]
	C	24%	15%	27%	24%	[59%]	3%	[42%]
Consensus		**N**	**N**	**N**	**N**	**C**	**T**	**Y**

Table 61: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85N+N92C

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	26%	12%	14%	13%	1%	4%	3%
	A	23%	40%	18%	17%	6%	2%	7%
	T	28%	33%	40%	/46%/	/35%/	[91%]	[46%]
	C	23%	15%	27%	24%	[58%]	3%	[44%]
Consensus		**N**	**N**	**N**	**N**	**C**	**T**	**Y**

Table 62: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85N+N92M

		PAM Position

		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	26%	13%	15%	14%	1%	3%	3%
	A	22%	40%	18%	19%	6%	2%	6%
	T	28%	32%	40%	/42%/	/36%/	[92%]	[52%]
	C	24%	15%	27%	25%	[57%]	2%	[39%]
Consensus		N	N	N	N	C	T	Y

Table 63: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85N+N92Q

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	26%	12%	14%	13%	1%	3%	4%
	A	23%	41%	19%	19%	6%	3%	8%
	T	28%	33%	/42%/	/46%/	/38%/	[90%]	[47%]
	C	23%	14%	26%	22%	[55%]	4%	[41%]
Consensus		N	N	N	N	C	T	Y

Table 64: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85N+N92I

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	26%	12%	15%	12%	1%	3%	2%
	A	22%	42%	18%	18%	5%	3%	6%
	T	28%	32%	/41%/	/44%/	/37%/	[90%]	[52%]
	C	24%	14%	26%	26%	[56%]	4%	[40%]
Consensus		N	N	N	N	C	T	Y

Table 65: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85S+N88D+Q89G

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	31%	12%	16%	13%	7%	12%	20%
	A	12%	39%	30%	21%	16%	38%	17%
	T	28%	29%	31%	/44%/	[65%]	22%	25%
	C	29%	20%	23%	21%	12%	28%	38%
Consensus		N	N	N	N	T	N	N

Table 66: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85S+N88H+Q89G+N92L

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	28%	10%	14%	11%	8%	18%	13%
	A	15%	37%	26%	17%	13%	30%	20%
	T	30%	35%	/41%/	[56%]	[72%]	34%	29%
	C	27%	17%	19%	16%	7%	18%	38%

Consensus		**N**	**N**	**N**	**T**	**T**	**N**	**N**

Table 67: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72A+N88D+Q89D+Q125R

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	31%	15%	25%	/44%/	25%	8%	11%
	A	14%	39%	33%	24%	20%	18%	17%
	T	27%	22%	16%	15%	40%	8%	21%
	C	28%	24%	26%	16%	15%	[66%]	[51%]
Consensus		**N**	**N**	**N**	**N**	**N**	**C**	**C**

Table 68: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85S+N88D+Q89G+N92L

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	31%	12%	18%	13%	7%	14%	20%
	A	12%	39%	29%	22%	14%	34%	17%
	T	28%	28%	29%	/43%/	[67%]	23%	27%
	C	29%	21%	25%	22%	12%	29%	37%
Consensus		**N**	**N**	**N**	**N**	**T**	**N**	**N**

Table 69: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72A+N88D+Q89G+Q125R

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	32%	16%	25%	40%	26%	9%	11%
	A	14%	39%	32%	26%	21%	22%	16%
	T	26%	22%	17%	16%	36%	7%	19%
	C	28%	23%	27%	19%	17%	[62%]	[54%]
Consensus		**N**	**N**	**N**	**N**	**N**	**C**	**C**

Table 70: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85S+N92L+Q125R

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	27%	10%	20%	26%	6%	5%	4%
	A	18%	46%	28%	29%	8%	17%	6%
	T	27%	28%	28%	32%	[82%]	/49%/	/35%/
	C	27%	16%	23%	13%	4%	29%	[55%]
Consensus		**N**	**N**	**N**	**N**	**T**	**H**	**C**

Table 71: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72S+K85D+Q125R+N127R

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	27%	13%	21%	22%	11%	11%	6%
	A	13%	36%	26%	21%	15%	13%	11%
	T	27%	26%	24%	23%	20%	19%	21%
	C	33%	25%	29%	33%	[54%]	[57%]	[63%]
Consensus		N	N	N	N	C	C	C

Table 72: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72A+K85S+Q89D+N92L+Q125R

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	28%	11%	17%	22%	14%	12%	15%
	A	14%	39%	28%	20%	17%	/31%/	11%
	T	26%	29%	29%	/40%/	[62%]	10%	/40%/
	C	32%	22%	27%	18%	7%	/46%/	/34%/
Consensus		N	N	N	N	T	N	N

Table 73: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72C+N88H+Q89G+Q125R

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	33%	16%	27%	38%	22%	16%	14%
	A	11%	37%	33%	25%	23%	25%	21%
	T	22%	19%	13%	16%	/41%/	13%	19%
	C	34%	28%	28%	21%	15%	/46%/	/45%/
Consensus		N	N	N	N	N	N	N

Table 74: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 Y72C+N88D+Q89D+N92W+Q125R

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	31%	16%	26%	/48%/	26%	6%	15%
	A	13%	38%	31%	21%	20%	16%	24%
	T	25%	20%	15%	16%	32%	4%	26%
	C	31%	25%	28%	15%	21%	[73%]	35%
Consensus		N	N	N	N	N	C	N

Table 75: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85Q+N92L

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	27%	17%	22%	17%	1%	3%	2%
	A	17%	36%	23%	24%	3%	3%	4%
	T	27%	26%	26%	27%	[95%]	[91%]	[44%]
	C	29%	21%	29%	33%	1%	4%	[50%]
Consensus		N	N	N	N	T	T	Y

Table 76: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 K85Q+N92W

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	28%	17%	21%	17%	2%	3%	11%
	A	16%	36%	24%	24%	6%	4%	12%
	T	27%	26%	27%	28%	[90%]	[88%]	/40%/
	C	30%	21%	29%	30%	2%	5%	37%
Consensus		N	N	N	N	T	T	N

Table 77: Summary of amino acid combinations substituted into the PI domain of Cas-alpha 10 and their effect on PAM recognition.

Cas-alpha 10 substitution	Resulting Cas-alpha 10 PAM recognition
K85S+N92L (SEQ ID NO:63)	5'-NYTY-3'
K85S+N92Q (SEQ ID NO:64)	5'-TT*TY-3'
K85S+N92C (SEQ ID NO:65)	5'-TT*TY-3'
K85S+N92H (SEQ ID NO:66)	5'-TTYTY-3'
K85S+N92A (SEQ ID NO:67)	5'-TYTY-3'
K85S+N92M (SEQ ID NO:68)	5'-NYTY-3'
K85N+N92L (SEQ ID NO:69)	5'-NC+TY-3'
K85N+N92H (SEQ ID NO:70)	5'-NC+TY-3'
K85N+N92A (SEQ ID NO:71)	5'-NC+TY-3'
K85N+N92C (SEQ ID NO:72)	5'-NC+TY-3'
K85N+N92M (SEQ ID NO:73)	5'-NC+TY-3'
K85N+N92Q (SEQ ID NO:74)	5'-NC+TY-3'
K85N+N92I (SEQ ID NO:75)	5'-NC+TY-3'
K85S+N88D+Q89G (SEQ ID NO:76)	5'-NTNN-3'
K85S+N88H+Q89G+N92L (SEQ ID NO:77)	5'-TTNN-3'
Y72A+N88D+Q89D+Q125R (SEQ ID NO:78)	5'-NNCC-3'
K85S+N88D+Q89G+N92L (SEQ ID NO:79)	5'-NTNN-3'
Y72A+N88D+Q89G+Q125R (SEQ ID NO:80)	5'-NNCC-3'
K85S+N92L+Q125R (SEQ ID NO:81)	5'-NTHC-3'
Y72S+K85D+Q125R+N127R (SEQ ID NO:82)	5'-NCCC-3'
Y72A+K85S+Q89D+N92L+Q125R (SEQ ID NO:83)	5'-NTNN-3'
Y72C+N88H+Q89G+Q125R (SEQ ID NO:84)	5'-NNNN-3'
Y72C+N88D+Q89D+N92W+Q125R (SEQ ID NO:85)	5'-NNCN-3'

K85Q+N92L (SEQ ID NO:86)	5'-NTTY-3'
K85Q+N92W (SEQ ID NO:87)	5'-NTTN-3'

*: Preference for T but C also accepted; +: Preference for C but T also accepted

Table 78: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 engineered nuclease 2 with K85N

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	28%	15%	22%	17%	5%	6%	2%
	A	18%	32%	21%	21%	12%	3%	4%
	T	27%	30%	29%	34%	[42%]	[83%]	15%
	C	28%	23%	29%	28%	[41%]	8%	[79%]
Consensus		**N**	**N**	**N**	**N**	**Y**	**T**	**C**

Table 79: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 engineered nuclease 2 with N92L

		PAM Position						
		-7	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**
Nucleotide	G	28%	17%	27%	/48%/	6%	4%	3%
	A	17%	36%	23%	21%	5%	3%	6%
	T	25%	26%	20%	19%	[86%]	[84%]	[37%]
	C	29%	21%	29%	11%	3%	9%	[54%]
Consensus		**N**	**N**	**N**	**N**	**T**	**T**	**Y**

Table 80: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 engineered nuclease 2 with K85N+N92L

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	28%	14%	19%	16%	2%	5%	2%
	A	17%	38%	19%	20%	7%	3%	7%
	T	26%	30%	32%	36%	[41%]	[88%]	[43%]
	C	29%	18%	30%	28%	[51%]	4%	[48%]
Consensus		N	N	N	N	Y	T	Y

Table 81: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 engineered nuclease 3 with K85N

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	28%	16%	21%	17%	4%	5%	2%
	A	19%	32%	20%	21%	12%	2%	3%
	T	26%	29%	29%	33%	[42%]	[86%]	11%
	C	28%	23%	29%	29%	[42%]	6%	[84%]
Consensus		N	N	N	N	Y	T	C

Table 82: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 engineered nuclease 3 with N92L

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	28%	17%	27%	[53%]	4%	3%	3%
	A	17%	37%	23%	21%	4%	2%	4%
	T	26%	27%	20%	17%	[90%]	[88%]	[38%]
	C	29%	20%	30%	9%	2%	7%	[55%]
Consensus		N	N	N	G	T	T	Y

Table 83: Protospacer adjacent motif (PAM) preferences for Cas-alpha 10 engineered nuclease 3 with K85N+N92L

		PAM Position						
		-7	-6	-5	-4	-3	-2	-1
Nucleotide	G	27%	14%	19%	16%	1%	5%	2%
	A	18%	38%	18%	21%	6%	2%	5%
	T	27%	30%	32%	36%	[40%]	[90%]	[43%]
	C	28%	18%	31%	28%	[52%]	3%	[49%]

Consensus		**N**	**N**	**N**	**N**	**Y**	**T**	**Y**

Example 3: Rescuing activity of Cas-alpha variants with altered PAM recognition

[0494] In this example, methods for improving DNA target binding and/or cleavage activity of Cas-alpha 10 variants with modified PAM recognition are described.

[0495] In a first method, protein directed evolution is used to improve DNA target binding and/or cleavage of a Cas-alpha 10 PAM variant. This can be accomplished using a toxic gene disruption assay or other similar method wherein one or more genes encoding a compound that slow or prevent cell growth and/or result in cell death is used as a way to select for protein variants with improved enzymatic activity (for example but not limited to *ccdB*, *codA*, or sacB) (Bernard and Couturier (1992) *Journal of Molecular Biology*. 226, 735-745, Mahan *et al*. (2004) *Biochemistry*. 43, 8957-8964, Gay *et al*. (1985) *Journal of Bacteriology*. 164, 918-921, Chen and Zhao (2005) *Nucleic Acids Research*. 33, e154). For this, one or more Cas-alpha 10 DNA target site(s) is selected within or in the vicinity of the toxic gene(s) in a way to disrupt its expression upon DNA target binding and/or cleavage. Cas-alpha 10 amino acid substitution(s) that permit improved DNA target affinity and/or cleavage at the new PAM permit cell survival and are then selected for.

[0496] In a second method, the above approach can also be used to select against undesirable PAM recognition. In this case, one or more DNA target(s) comprising the unwanted PAM can be placed within or near the Cas-alpha 10 gene such that variants that bind and/or cleave the site are selected against.

CLAIMS

What is claimed:

1. A non-naturally occurring Cas-alpha 10 polypeptide comprising an amino acid amino acid sequence having at least 90% sequence identity to SEQ ID NO: 2, wherein the amino acid sequence comprises a mutation relative to SEQ ID NO: 2, wherein the mutation comprises a K85S mutation; a K85A mutation; a N92R mutation; a N92K mutation; a Q125K mutation; a N88H mutation; a N88K mutation; a N88Q mutation; a Q125F mutation; a Y72V mutation; a Y72E mutation; a Y72Q mutation; a Y72T mutation; a Y72C mutation; a Y72A mutation; a Y72S mutation; a Y72P mutation; a Y72G mutation; a Y72D mutation; a Y72L mutation; a K85G mutation; a K85D mutation; a K85N mutation; a N88D mutation; a Q89D mutation; a N92D mutation; a N92C mutation; a N92A mutation; a N92G mutation; a N92F mutation; a N92E mutation; a N92H mutation; a N92I mutation; a N92L mutation; a N92V mutation; a N92W mutation; a N92Q mutation; a N92S mutation; a N92P mutation; a N92Y mutation; a N92T mutation; a N92M mutation; a Q125R mutation; or a Q125P mutation.

2. The non-naturally occurring Cas-alpha 10 polypeptide of claim 1, wherein the amino acid sequence comprises one or more of the following combination of mutations relative to SEQ ID NO: 2: a combination of a K85Q mutation and a N92L mutation; a combination of a N88H mutation and a Q89G mutation; a combination of a N88K mutation and a Q89G mutation; a combination of a N88Q mutation and a Q89G mutation; a combination of a K85S mutation and a N92L; a combination of a K85S mutation and a N92Q mutation; a combination of a K85S mutation and a N92C mutation; a combination of a K85S mutation and a N92H mutation; a combination of a K85S mutation and a N92A mutation; a combination of a K85S mutation and a N92M mutation; a combination of a K85N mutation and a N92L mutation; a combination of a K85N mutation and a N92H; a combination of a K85N mutation and a N92A mutation; a combination of a K85N mutation and a N92C mutation; a combination of a K85N mutation and a N92M mutation; a combination of a K85N mutation and a N92Q mutation; a combination of a K85N mutation and a N92I mutation; a combination of a K85S mutation, a N88D mutation, and a Q89G mutation; a combination of a K85S mutation, a N88H mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89D mutation, and a

Q125R mutation; a combination of a K85S mutation, a N88D mutation, a Q89G mutation, and a N92L mutation; a combination of a Y72A mutation, a N88D mutation, a Q89G mutation, and a Q125R mutation; a combination of a K85S mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72S mutation, a K85D mutation, a Q125R mutation, and a N127R mutation; a combination of a Y72A mutation, a K85S mutation, a Q89D mutation, a N92L mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88H mutation, a Q89G mutation, and a Q125R mutation; a combination of a Y72C mutation, a N88D mutation, a Q89D mutation, a N92W mutation, and a Q125R mutation; or a combination of a K85Q and a N92W mutation.

3. The non-naturally occurring Cas-alpha 10 polypeptide of claim 1 or claim 2, wherein the amino acid sequence comprises a PAM interacting (PI) domain comprising amino acids from S63 to I196.

4. The non-naturally occurring Cas-alpha 10 polypeptide of any one of claims 1-3, wherein the amino acid sequence comprises the K85S mutation.

5. The non-naturally occurring Cas-alpha 10 polypeptide of any one of claims 1-3, wherein the amino acid sequence comprises the combination of the K85Q mutation and the N92L mutation.

6. The non-naturally occurring Cas-alpha 10 polypeptide of any one of claims 1-3, wherein the amino acid sequence comprises the combination of the K85N mutation and the N92L mutation.

7. The non-naturally occurring Cas-alpha 10 polypeptide of any one of claims 1-6, wherein the Cas-alpha 10 polypeptide has endonuclease activity.

8. The non-naturally occurring Cas-alpha 10 polypeptide of any one of claims 1-6, wherein the Cas-alpha 10 polypeptide is a deactivated Cas-alpha 10 endonuclease complexed to a deaminase.

9. The non-naturally occurring Cas-alpha 10 polypeptide of any one of claims 1-6, wherein the Cas-alpha 10 polypeptide has nickase activity.

10. The non-naturally occurring Cas-alpha 10 polypeptide of any one of claims 1-6 or claim 9, wherein the Cas-alpha 10 polypeptide is complexed to a reverse transcriptase or operably associated with a reverse transcriptase.

11. The non-naturally occurring Cas-alpha 10 polypeptide of any one of claims 1-6, wherein the Cas-alpha 10 polypeptide is complexed to a heterologous protein domain through a linker, the heterologous protein domain having methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, reverse transcriptase activity, RNA aptamer binding activity, or nucleic acid binding activity.

12. The non-naturally occurring Cas-alpha 10 polypeptide of any one of claims 1-11 comprising an amino acid amino acid sequence having at least 90% sequence identity to any one of SEQ ID Nos. 6, 5, 7-16, or 28-87.

13. A non-naturally occurring Cas-alpha 10 polypeptide comprising a PAM interacting (PI) domain, wherein the PI domain recognizes a PAM sequence comprising 5'-YTC-3', 5'-TTTC-3', 5'-GTTC-3', 5'-GTYC-3', 5'-GTWC-3', 5'-DTHC-3', 5'-GTNC-3', 5'-NNNC-3', 5'-GNNC-3', 5'-HCTC-3', 5'-NTCC-3', 5'-TNC-3', 5'-GTTY-3', 5'-DTTY-3', 5'-GTTT-3', 5'-GTHC-3', 5'-NYTY-3', 5'-TTTY-3', 5'-TTYTY-3', 5'-TYTY-3', 5'-TTNN-3', 5'-NNCC-3', 5'-NCCC-3', 5'-NTNN-3', 5'-NCTY-3', 5'-NTHC-3', 5'-NNNN-3', or 5'-NNCN-3', wherein N = A, C, G, or T, Y = T or C, D = G, A, or T, W = A or T, and H = A, T, or C.

14. The non-naturally occurring Cas-alpha 10 polypeptide of claim 13, wherein the Cas-alpha 10 polypeptide has endonuclease activity.

15. The non-naturally occurring Cas-alpha 10 polypeptide of claim 13, wherein the Cas-alpha 10 polypeptide is a deactivated Cas-alpha 10 endonuclease complexed to a deaminase.

16. The non-naturally occurring Cas-alpha 10 polypeptide of claim 13, wherein the Cas-alpha 10 polypeptide has nickase activity.

17. The non-naturally occurring Cas-alpha 10 polypeptide of claim 13 or claim 16, wherein the Cas-alpha 10 polypeptide is complexed to a reverse transcriptase or operably associated with a reverse transcriptase.

18. The non-naturally occurring Cas-alpha 10 polypeptide of claim 13, wherein the Cas-alpha 10 polypeptide is complexed to a heterologous protein domain through a linker, and wherein the heterologous protein domain has methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, reverse transcriptase activity, RNA aptamer binding activity, or nucleic acid binding activity.

19. A synthetic composition comprising:

(a) a Cas-alpha 10 polypeptide having DNA binding activity, the non-naturally occurring Cas-alpha 10 polypeptide according to any one of claims 1-18; and

(b) at least one guide polynucleotide comprising a region of complementarity to a target polynucleotide, wherein the target polynucleotide is heterologous to the source of the Cas-alpha 10 polypeptide,

wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide and forms a complex with the at least one guide polynucleotide, and wherein the complex binds to the target polynucleotide.

20. A method of editing a target polynucleotide in a cell, the method comprising:

(a) providing to the cell a Cas-alpha 10 polypeptide according to any one of claims 1-18, wherein the Cas-alpha 10 polypeptide recognizes a PAM sequence on the target polynucleotide;

(b) providing the cell with at least one guide polynucleotide comprising a region of complementarity to the target polypeptide, wherein the Cas-alpha 10 polypeptide forms a complex with the guide polynucleotide and the complex binds the target polynucleotide; and

(c) introducing at least one nucleotide modification in the target polynucleotide via the complex, wherein the target polynucleotide is heterologous to the Cas-alpha 10 polypeptide.

21. The method of claim 20, further comprising providing the cell with a donor DNA molecular or a polynucleotide modification template.

22. The method of claim 20, wherein the cell is derived or obtained from an animal, a fungus, or a plant.

23. The method of claim 22, wherein the plant is a dicot or a monocot.

24. The method of claim 22, wherein the plant is maize, soybean, cotton, wheat, canola, oilseed rape, sorghum, rice, rye, barley, millet, oats, sugarcane, turfgrass, switchgrass, alfalfa, sunflower, tobacco, peanut, potato, Arabidopsis, safflower, or tomato.

25. An animal, a fungus, or a cell thereof comprising the non-naturally occurring Cas-alpha 10 polypeptide according to any one of claims 1-18.

26. A plant or plant cell comprising the non-naturally occurring Cas-alpha 10 polypeptide according to any one of claims 1-18.

27. A method for altering protospacer adjacent motif (PAM) specificity of a target Cas-alpha 10 polypeptide, the method comprising :

(a) comparing the PAM interacting (PI) domain of an orthologous Cas-alpha polypeptide with the PI domain of the target Cas-alpha 10 polypeptide;

(b) selecting one or more amino acids and/or one or more polypeptide chains from the PI domain of the orthologous Cas-alpha polypeptide;

(c) incorporating the one or more amino acids and/or the one or more polypeptide chains selected from the PI domain of the orthologous Cas-alpha polypeptide into one or more

structurally similar positions of the target Cas-alpha 10 polypeptide resulting in a modified target Cas-alpha 10 polypeptide; and

(d) determining PAM recognition of the modified target Cas-alpha 10 polypeptide.

28. The method of claim 27, wherein the orthologous Cas-alpha polypeptide has different PAM recognition than the target Cas-alpha 10 polypeptide.

29. The method of claim 27, wherein the Cas-alpha 10 polypeptide comprises an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 2.

30. The method of any one of claims 27-29, wherein the orthologous Cas-alpha polypeptide is Cas-alpha 1, Cas-alpha 2, Cas-alpha 3, Cas-alpha 4, Cas-alpha 5, Cas-alpha 6, Cas-alpha 7, Cas-alpha 8, or Cas-alpha 11.

FIG. 1

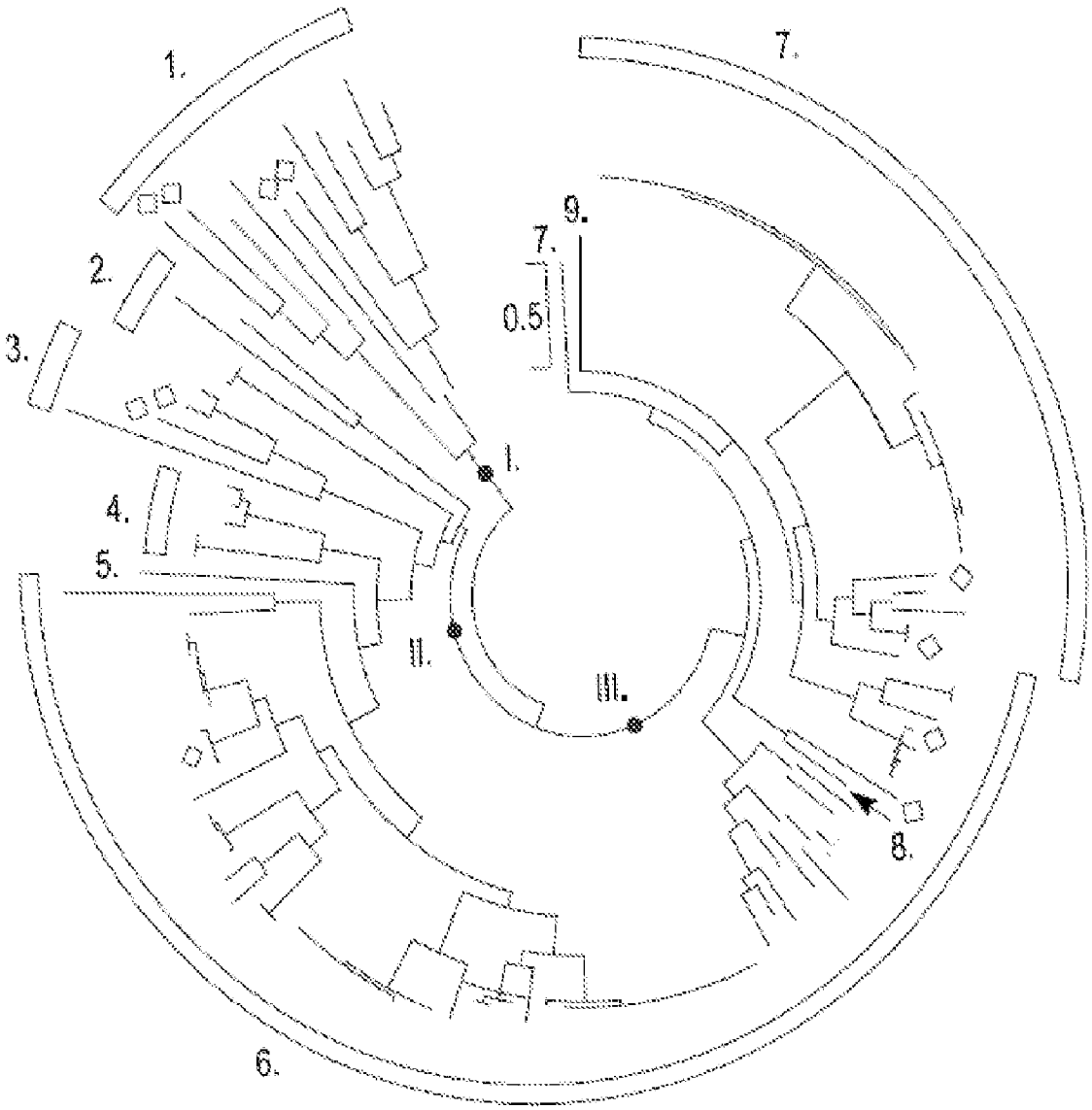


FIG. 2

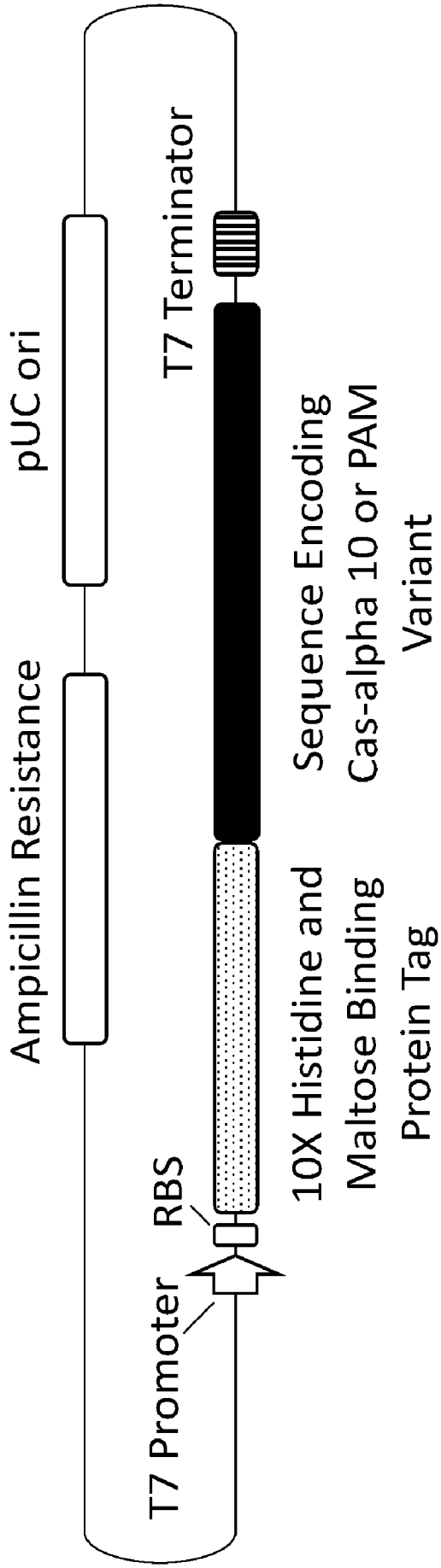

Ampicillin Resistance
pUC ori
T7 Terminator
Sequence Encoding Cas-alpha 10 or PAM Variant
10X Histidine and Maltose Binding Protein Tag
RBS
T7 Promoter

FIG. 3

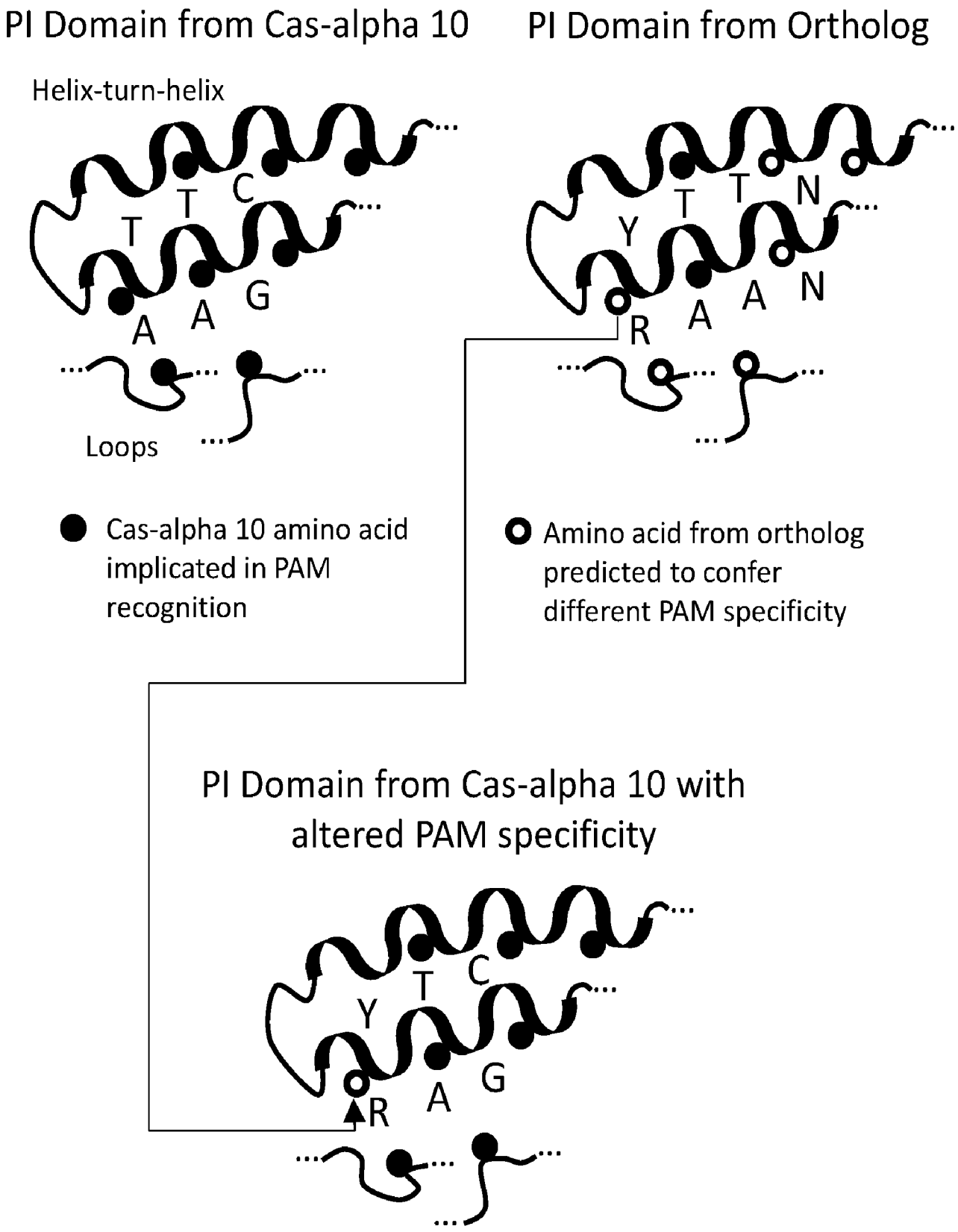
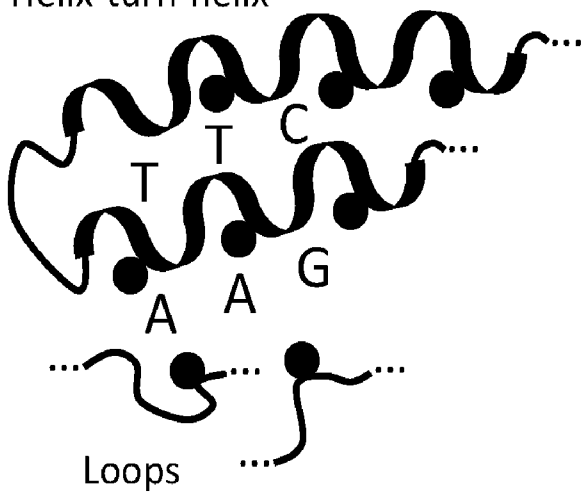

PI Domain from Cas-alpha 10
PI Domain from Ortholog
Helix-turn-helix
T C
T
A A G
T T N
Y
R A A N
Loops
Cas-alpha 10 amino acid implicated in PAM recognition
Amino acid from ortholog predicted to confer different PAM specificity
PI Domain from Cas-alpha 10 with altered PAM specificity
T C
Y
R A G

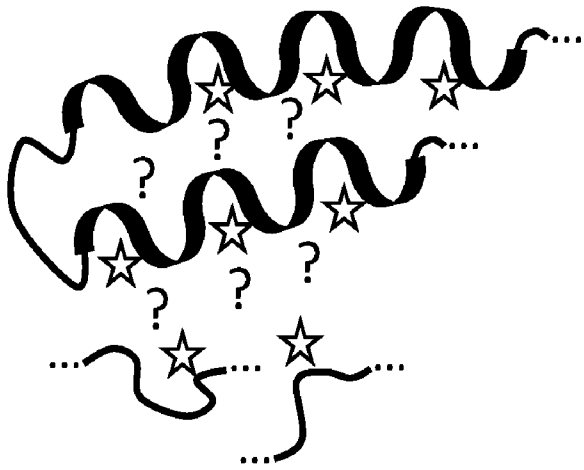
FIG. 4

PI Domain from Cas-alpha 10

Helix-turn-helix



Novel Variation Introduced



● Cas-alpha 10 amino acid implicated in PAM recognition

☆ All possible amino acids inserted and individually tested for perturbations in PAM specificity

PI Domain from Cas-alpha 10 with Altered PAM specificity

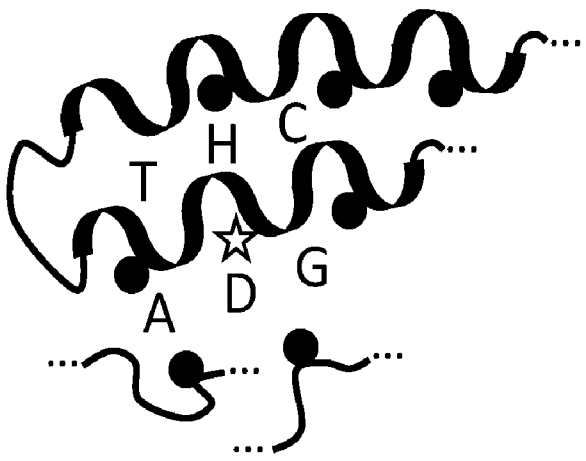


FIG. 5

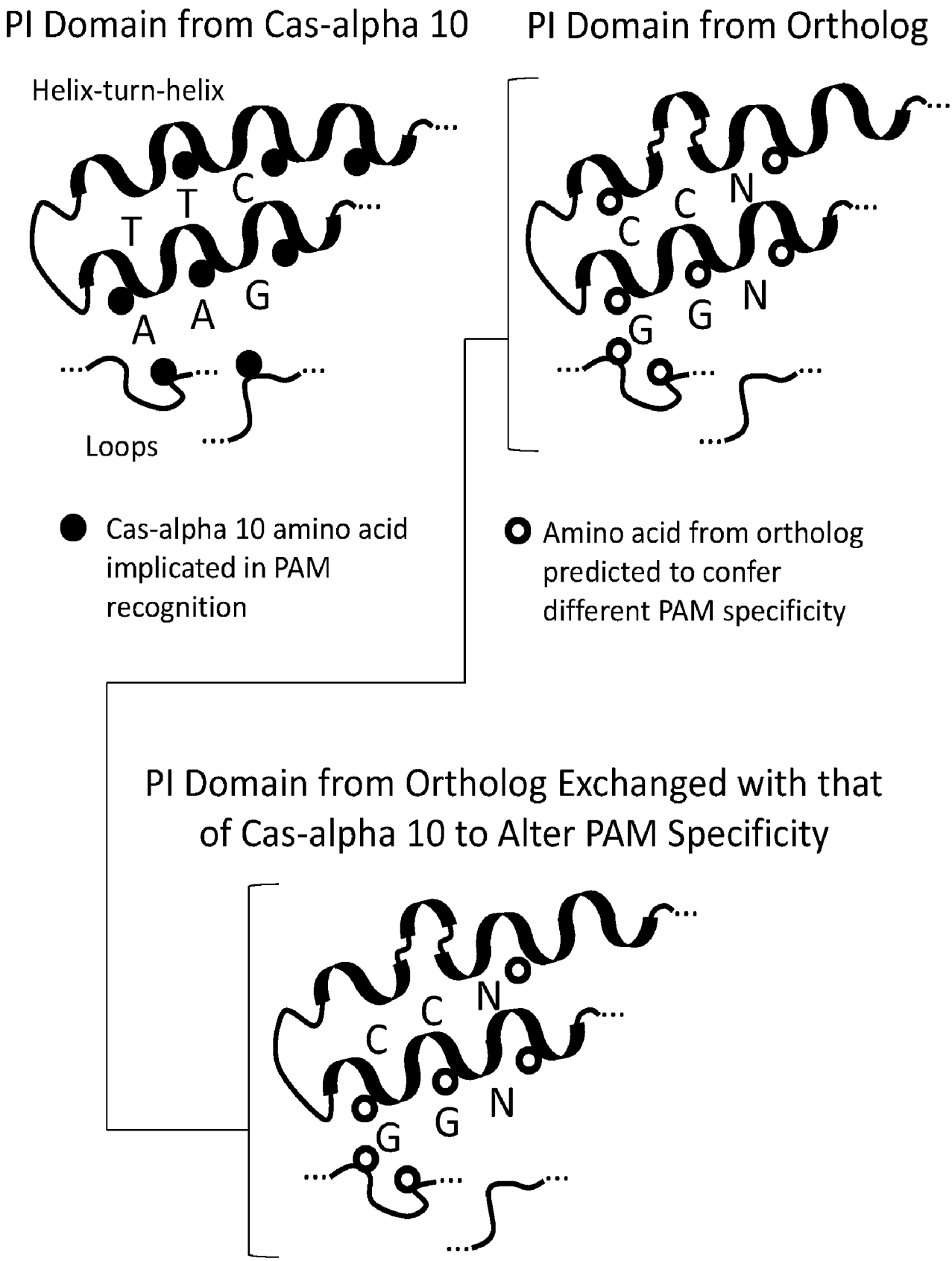
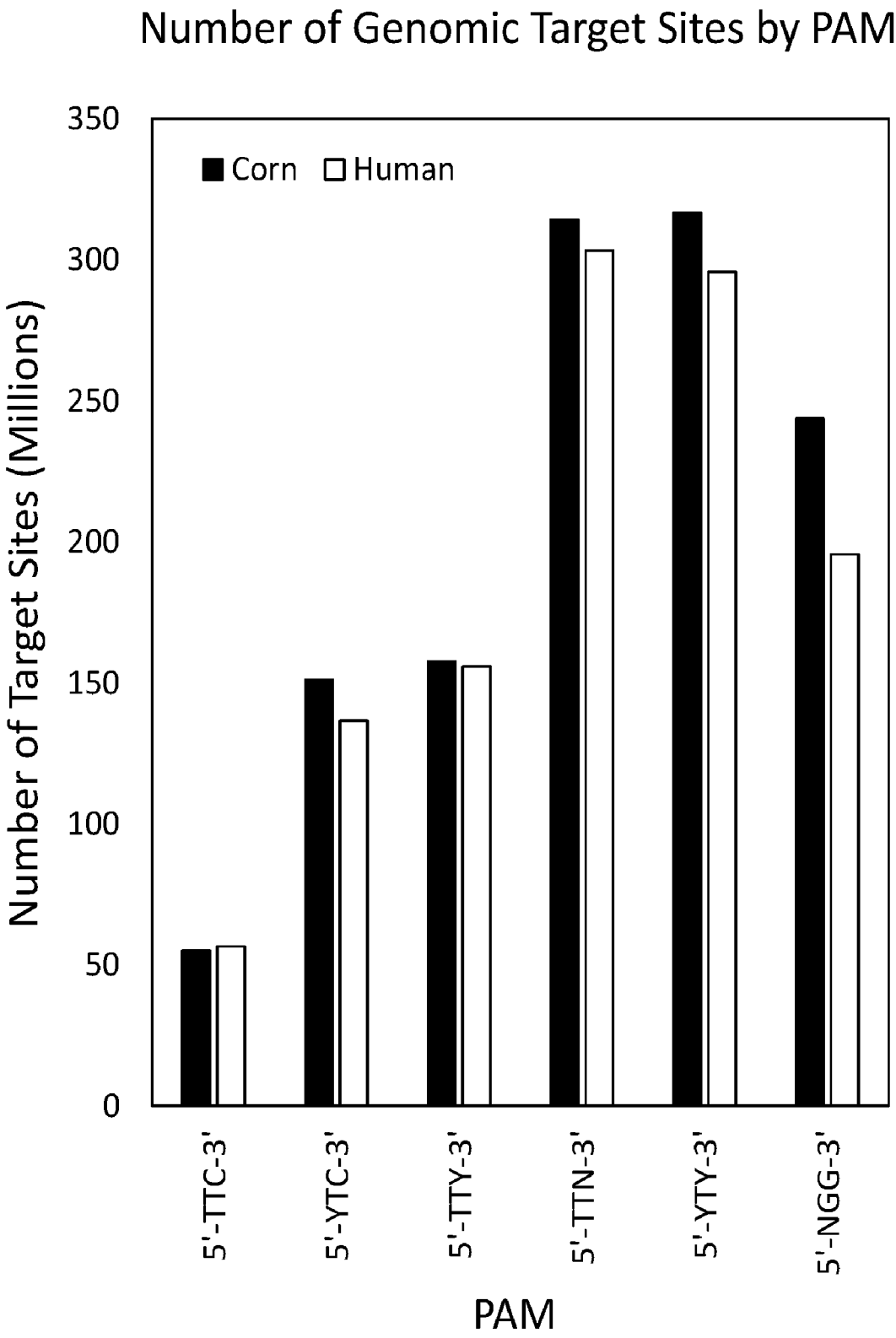

PI Domain from Cas-alpha 10
PI Domain from Ortholog
Helix-turn-helix
T
C
T
A
A
G
Loops
C
C
N
G
G
N
Cas-alpha 10 amino acid implicated in PAM recognition
Amino acid from ortholog predicted to confer different PAM specificity
PI Domain from Ortholog Exchanged with that of Cas-alpha 10 to Alter PAM Specificity
C
C
N
G
G
N

FIG. 6


Number of Genomic Target Sites by PAM
Corn
Human
Number of Target Sites (Millions)
350
300
250
200
150
100
50
0
5'-TTC-3'
5'-YTC-3'
5'-TTY-3'
5'-TTN-3'
5'-YTY-3'
5'-NGG-3'
PAM

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2024/020512

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/22
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, CHEM ABS Data, Sequence Search, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2022/082179 A2 (PIONEER HI BRED INT [US]) 21 April 2022 (2022-04-21) abstract paragraph [0493] - paragraph [0494] paragraph [0566] - paragraph [0568] claims 1-19,24-29; sequence 135 ----- -/--	1-12, 19-26

[x] Further documents are listed in the continuation of Box C.

[x] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
7 June 2024	16/08/2024

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Gurdjian, Didier

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WANG YUJUE ET AL: "Guide RNA engineering enables efficient CRISPR editing with a miniature Syntrophomonas palmitatica Cas12f1 nuclease", CELL REPORTS, vol. 40, no. 13, 1 September 2022 (2022-09-01), page 111418, XP093146198, US ISSN: 2211-1247, DOI: 10.1016/j.celrep.2022.111418 Retrieved from the Internet: URL:https://www.sciencedirect.com/science/article/pii/S2211124722012591?via%3Dihub> abstract; figure 1	1-12, 19-26
Y	WO 2020/088450 A1 (UNIV CHINA AGRICULTURAL [CN]) 7 May 2020 (2020-05-07) abstract paragraph [0164] - paragraph [0166]; claims 1-14,49,50	1-12, 19-26
Y	WU ZHAOWEI ET AL: "Programmed genome editing by a miniature CRISPR-Cas12f nuclease", NATURE CHEMICAL BIOLOGY, NATURE PUBLISHING GROUP US, NEW YORK, vol. 17, no. 11, 2 September 2021 (2021-09-02), pages 1132-1138, XP037599543, ISSN: 1552-4450, DOI: 10.1038/S41589-021-00868-6 [retrieved on 2021-09-02] abstract; figure 1f	1-12, 19-26
Y	NGUYEN GIANG T. ET AL: "Miniature CRISPR-Cas12 endonucleases - Programmed DNA targeting in a smaller package", CURRENT OPINION IN STRUCTURAL BIOLOGY, vol. 77, 1 December 2022 (2022-12-01), page 102466, XP093165005, GB ISSN: 0959-440X, DOI: 10.1016/j.sbi.2022.102466 Retrieved from the Internet: URL:https://pdf.sciencedirectassets.com/272019/1-s2.0-S0959440X22X00067/1-s2.0-S0959440X22001452/main.pdf> abstract; figure 2f	1-12, 19-26

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2024/020512

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MATTIELLO LUCIA ET AL: "Molecular and Computational Strategies to Increase the Efficiency of CRISPR-Based Techniques", FRONTIERS IN PLANT SCIENCE, vol. 13, 31 May 2022 (2022-05-31), XP093170331, CH ISSN: 1664-462X, DOI: 10.3389/fpls.2022.868027 Retrieved from the Internet: URL:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9194676/pdf/fpls-13-868027.pdf> abstract page 12, left-hand column, paragraph 2 -----	1-12, 19-26
Y	COLLIAS DAPHNE ET AL: "CRISPR technologies and the search for the PAM-free nuclease", NATURE COMMUNICATIONS, vol. 12, no. 1, 22 January 2021 (2021-01-22), pages 1-12, XP093099053, DOI: 10.1038/s41467-020-20633-y Retrieved from the Internet: URL:https://www.nature.com/articles/s41467-020-20633-y.pdf> abstract -----	1-12, 19-26

1

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2024/020512

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. [X] forming part of the international application as filed.

   b. [ ] furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)).

      [ ] accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. [ ] With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/020512

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

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

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

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

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

see additional sheet

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

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

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

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims;; it is covered by claims Nos.:
1-12, 19-26

Remark on Protest

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

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

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-12, 19-26

A non-naturally occurring Cas-alpha 10 polypeptide comprising an amino acid amino acid sequence having at least 90% sequence identity to SEQ ID NO: 2, wherein the amino acid sequence comprises a mutation relative to SEQ ID NO: 2, wherein the mutation comprises a K85S mutation; a K85A mutation; a N92R mutation; a N92K mutation; a Q125K mutation; a N88H mutation; a N88K mutation; a N88Q mutation; a Q125F mutation; a Y72V mutation; a Y72E mutation; a Y72Q mutation; a Y72T mutation; a Y72C mutation; a Y72A mutation; a Y72S mutation; a Y72P mutation; a Y72G mutation; a Y72D mutation; a Y72L mutation; a K85G mutation; a K85D mutation; a K85N mutation; a N88D mutation; a Q89D mutation; a N92D mutation; a N92C mutation; a N92A mutation; a N92G mutation; a N92F mutation; a N92E mutation; a N92H mutation; a N92I mutation; a N92L mutation; a N92V mutation; a N92W mutation; a N92Q mutation; a N92S mutation; a N92P mutation; a N92Y mutation; a N92T mutation; a N92M mutation; a Q125R mutation; or a Q125P mutation, a corresponding synthetic composition, a method of editing a target polynucleotide, and aninal, fungus or plant comprising it.

---

2. claims: 13-18

A non-naturally occurring Cas-alpha 10 polypeptide comprising a PAM interacting (PI) domain, wherein the PI domain recognizes a PAM sequence comprising 5'-YTC-3', 5'-TTTC-3', 5'-GTTC-3', 5'-GTYC-3', 5'-GTWC-3', 5'-DTHC-3', 5'-GTNC-3', 5'-NNNC-3', 5'-GNNC-3', 5'-HCTC-3', 5'-NTCC-3', 5'-TNC-3', 5'-GTTY-3', 5'-DTTY-3', 5'-GTTT-3', 5'-GTHC-3', 5'-NYTY-3', 5'-TTTY-3', 5'-TTYTY-3', 5'-TYTY-3', 5'-TTNN-3', 5'-NNCC-3', 5'-NCCC-3', 5'-NTNN-3', 5'-NCTY-3', 5'-NTHC-3', 5'-NNNN-3', or 5'-NNCN-3', wherein N = A, C, G, or T, Y = T or C, D = G, A, or T, W = A or T, and H = A, T, or C.

---

3. claims: 27-30

A method for altering protospacer adjacent motif (PAM) specificity of a target Cas-alpha 10 polypeptide, the method comprising :(a) comparing the PAM interacting (PI) domain of an orthologous Cas-alpha polypeptide with the PI domain of the target Cas-alpha 10 polypeptide;(b) selecting one or more amino acids and/or one or more polypeptide chains from the PI domain of the orthologous Cas-alpha polypeptide;(c) incorporating the one or more amino acids and/or the one or more polypeptide chains selected from the PI domain of the orthologous Cas-alpha polypeptide into one or more structurally similar positions of the target Cas-alpha 10 polypeptide resulting in a modified target Cas-alpha 10

polypeptide; and(d) determining PAM recognition of the modified target Cas-alpha 10 polypeptide.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2024/020512

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2022082179 A2	21-04-2022	AU 2021362238 A1	11-05-2023
		CA 3197672 A1	21-04-2022
		CN 116391038 A	04-07-2023
		EP 4229194 A2	23-08-2023
		JP 2023545493 A	30-10-2023
		KR 20230082683 A	08-06-2023
		US 2023392135 A1	07-12-2023
		WO 2022082179 A2	21-04-2022
WO 2020088450 A1	07-05-2020	AU 2019372642 A1	17-06-2021
		BR 112021007994 A2	26-10-2021
		CA 3118251 A1	07-05-2020
		CN 111757889 A	09-10-2020
		CN 113106081 A	13-07-2021
		CN 113136375 A	20-07-2021
		EP 3875469 A1	08-09-2021
		IL 282746 A	30-06-2021
		JP 7216877 B2	02-02-2023
		JP 2022512982 A	07-02-2022
		KR 20210129033 A	27-10-2021
		PH 12021550904 A1	29-11-2021
		SG 11202104347U A	28-05-2021
		US 2021395784 A1	23-12-2021
		WO 2020088450 A1	07-05-2020