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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
01 November 2018 (01.11.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/197520 A1

(51) International Patent Classification:
C07K 14/005 (2006.01) *C12N 5/10* (2006.01)
C07K 14/315 (2006.01) *C12N 15/63* (2006.01)

(21) International Application Number:
PCT/EP2018/060510

(22) International Filing Date:
24 April 2018 (24.04.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/488,981	24 April 2017 (24.04.2017)	US
62/488,969	24 April 2017 (24.04.2017)	US
62/510,914	25 May 2017 (25.05.2017)	US
62/510,896	25 May 2017 (25.05.2017)	US

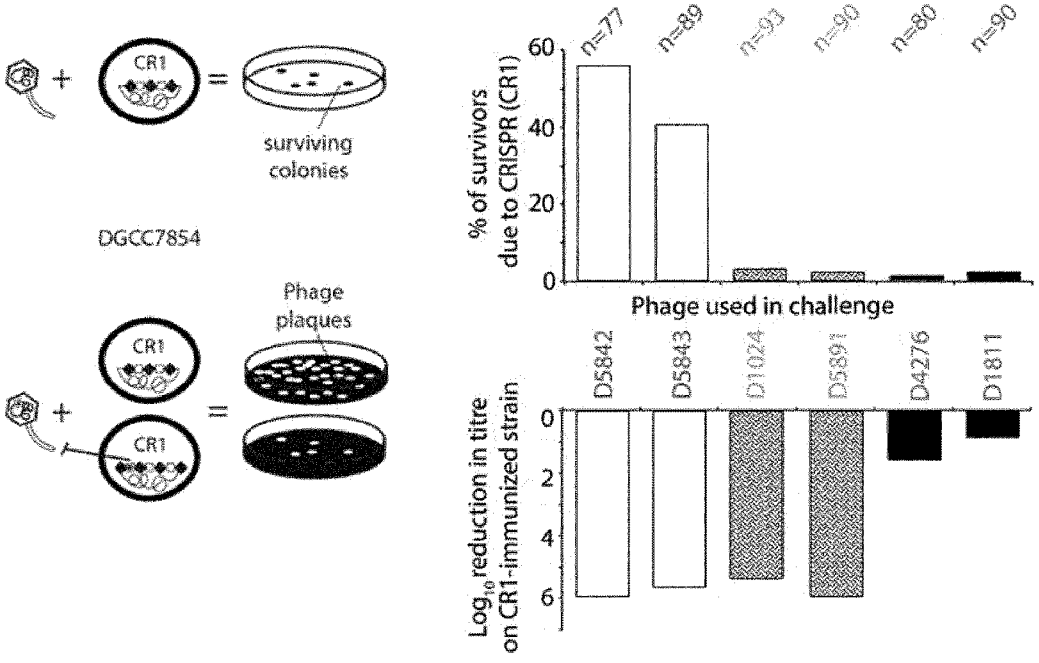
(71) Applicants: **DUPONT NUTRITION BIOSCIENCES APS** [DK/DK]; Langebrogade 1, 1411 Copenhagen K (DK). **UNIVERSITÉ LAVAL (CANADA)** [CA/CA]; Vice-rectorat à la recherche et à la création, 2320 rue des Bibliothèques, Québec, G1V 0A6 (CA).

(72) Inventors: **FREMAUX, Christophe**; c/o Danisco France SAS, Patent Department, 20, rue Brunel, 75917 Paris (FR). **HORVATH, Philippe**; c/o Danisco France SAS, Patent Department, 20, rue Brunel, 75917 Paris (FR). **ROMERO, Dennis A**; c/o Danisco USA Inc., Four New Century Parkway, New Century, Kansas 66301-0026 (US). **HYNES, Alexander**; c/o Université Laval (Canada), Vice-rectorat à la recherche et à la création, 2320 rue des Bibliothèques, Québec, G1V 0A6 (CA). **MOINEAU, Sylvain**; c/o Université Laval (Canada), Vice-rectorat à la recherche et à la création, 2320 rue des Bibliothèques, Québec, G1V 0A6 (CA). **ROUSSEAU, Geneviève**; c/o Université Laval (Canada), Vice-rectorat à la recherche et à la création, 2320 rue des Bibliothèques, Québec, G1V 0A6 (CA). **YOUNG, Joshua**; c/o Pioneer Hi-Bred International, Inc., 7250 NW 62ND AVENUE, PO BOX 552, JOHNSTON, Iowa 50131-0552 (US). **LEMAY, Marie-Laurence**; c/o Université Laval (Canada), Vice-rectorat à la recherche et à la création, 2320 rue des Bibliothèques, Québec, G1V 0A6 (CA).

(74) Agent: **MALLALIEU, Catherine** et al.; D Young & Co LLP, 120 Holborn, London EC1N 2DY (GB).

(54) Title: METHODS AND COMPOSITIONS OF ANTI-CRISPR PROTEINS FOR USE IN PLANTS

FIG. 1



(57) Abstract: Methods and compositions are provided for the use of anti-CRISPR (ACR) proteins in plants, including modulation of Cas endonuclease activity, improvement of frequency of homologous recombination, control of Cas endonuclease activity during various cell cycles, spatial and/or temporal regulation of Cas endonuclease activity in plants, usage in gene activation or repression, as well as reduction of off-target polynucleotide cleavage.

[Continued on next page]

WO 2018/197520 A1

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

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

Published:

— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with information concerning one or more priority claims considered void (Rule 26bis.2(d))*
— *with sequence listing part of description (Rule 5.2(a))*

METHODS AND COMPOSITIONS OF ANTI-CRISPR PROTEINS FOR USE IN PLANTS

[0001] This application claims the benefit of U.S. Provisional Application No. 62/488,981 filed on 24 April 2017, and of U. S. Provisional Application No. 62/510,914 filed on 25 May 2017, and of PCT Application No. PCT/EP2018/060481 filed on 24 April 2018 which claims the benefit of U.S. Provisional Patent Application No. 62/488,969 filed on 24 April 2017, and of U.S. Provisional Application No. 62/510,896 filed on 25 May 2017, all of which are incorporated herein in their entirety by reference.

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

[0002] The official copy of the sequence listing is submitted electronically via EFS-Web as an ASCII formatted sequence listing with a file named NB41268WOPCT_SequenceListing_ST25.TXT created on 24 April 2018 and having a size of 525,189 bytes and is filed concurrently with the specification. The sequence listing comprised in this ASCII formatted document is part of the specification and is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

[0003] The disclosure relates to the field of molecular biology, in particular, to compositions and methods relating to anti-CRISPR (ACR) proteins compositions and methods of use in plants.

BACKGROUND

[0004] Recombinant DNA technology has made it possible to modify (edit) specific endogenous chromosomal sequences and/or insert DNA sequences at targeted genomic locations, thus altering the organism's phenotype. 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) or transcription activator-like effector nucleases (TALENs), or homing meganucleases, are available for producing targeted genome perturbations, but these systems tends to have a low

specificity and employ designed nucleases that need to be redesigned for each target site, which renders them costly and time-consuming to prepare. Recently, genome-editing tools have been developed from bacterial and archaeal CRISPR systems that offer improved programmability to address a wider array of target sequences as well as improved specificity and efficiency in some applications.

[0005] Although CRISPR-derived systems offer many benefits over previous gene-editing tools, compositions and methods are still needed that can further improve these benefits.

SUMMARY OF INVENTION

[0006] As described herein, methods and compositions are provided for the identification, characterization, and utilization of anti-CRISPR (ACR) proteins in plants, including modulation of Cas endonuclease activity, improvement of frequency of homologous recombination, control of Cas endonuclease activity during various cell cycles, spatial and/or temporal regulation of Cas endonuclease activity in plants, usage in gene activation or repression, as well as reduction of off-target polynucleotide cleavage.

[0007] In one aspect, a method for modulating the activity of a Cas endonuclease with a target polynucleotide in a plant cell is provided, comprising providing to said cell a Cas endonuclease, a guide RNA capable of binding to the target polynucleotide in the plant cell, and an anti-CRISPR polypeptide capable of reducing the activity of said Cas endonuclease in said plant cell.

[0008] In one aspect, a method for modulating the activity of a Cas endonuclease with a target polynucleotide in a plant cell is provided, comprising providing to said cell a Cas endonuclease, a guide RNA capable of binding to the target polynucleotide in the plant cell, and an anti-CRISPR polypeptide capable of reducing the activity of said Cas endonuclease in said plant cell, wherein the activity selected from the group consisting of: target polynucleotide binding, target polynucleotide nicking, target polynucleotide double-strand-break creation, and target polynucleotide modification.

[0009] In one aspect, a method for modulating the activity of a Cas endonuclease with a target polynucleotide in a plant cell is provided, comprising providing to said cell a Cas endonuclease, a guide RNA capable of binding to the target polynucleotide in the plant cell, and an anti-CRISPR polypeptide capable of reducing the activity of said Cas endonuclease in said

plant cell, wherein the activity selected from the group consisting of: target polynucleotide binding, target polynucleotide nicking, target polynucleotide double-strand-break creation, and target polynucleotide modification, wherein said target polynucleotide modification is selected from the group consisting of: insertion of at least one nucleotide, deletion of at least one nucleotide, substitution of at least one nucleotide, and chemical alteration of at least one nucleotide.

[0010] In one aspect, a method for modulating the activity of a Cas endonuclease with a target polynucleotide in a plant cell is provided, comprising providing to said cell a Cas endonuclease, a guide RNA capable of binding to the target polynucleotide in the plant cell, and an anti-CRISPR polypeptide capable of reducing the activity of said Cas endonuclease in said plant cell, wherein the Cas endonuclease lacks the ability to nick or cleave a target polynucleotide.

[0011] In one aspect, a method for modulating the activity of a Cas endonuclease with a target polynucleotide in a plant cell is provided, comprising providing to said cell a Cas endonuclease, a guide RNA capable of binding to the target polynucleotide in the plant cell, and an anti-CRISPR polypeptide capable of reducing the activity of said Cas endonuclease in said plant cell, wherein said activity is decreased as compared to an isoline plant cell comprising said Cas endonuclease and guide RNA but not comprising said anti-CRISPR polypeptide.

[0012] In one aspect, a method for modulating the activity of a Cas endonuclease with a target polynucleotide in a plant cell is provided, comprising providing to said cell a Cas endonuclease, a guide RNA capable of binding to the target polynucleotide in the plant cell, and an anti-CRISPR polypeptide, wherein the activity of said Cas endonuclease in said plant cell is abolished during at least one timepoint, in at least one tissue or cell type, or during at least one phase of the cell or plant life cycle.

[0013] In one aspect, a method for modulating the activity of a Cas endonuclease with a target polynucleotide in a plant cell is provided, comprising providing to said cell a Cas endonuclease, a guide RNA capable of binding to the target polynucleotide in the plant cell, and an anti-CRISPR polypeptide capable of reducing the activity of said Cas endonuclease in said plant cell, wherein the Cas endonuclease is a Type II-A Cas endonuclease.

[0014] In one aspect, a method for modulating the activity of a Cas endonuclease with a target polynucleotide in a plant cell is provided, comprising providing to said cell a Cas

endonuclease, a guide RNA capable of binding to the target polynucleotide in the plant cell, and an anti-CRISPR polypeptide capable of reducing the activity of said Cas endonuclease in said plant cell, wherein the Cas endonuclease is Cas9.

[0015] In one aspect, a method for modulating the activity of a Cas endonuclease with a target polynucleotide in a plant cell is provided, comprising providing to said cell a Cas endonuclease, a guide RNA capable of binding to the target polynucleotide in the plant cell, and an anti-CRISPR polypeptide capable of reducing the activity of said Cas endonuclease in said plant cell, wherein the Cas endonuclease is Cpf1.

[0016] In one aspect, a method for modulating the activity of a Cas endonuclease with a target polynucleotide in a plant cell is provided, comprising providing to said cell a Cas endonuclease, a guide RNA capable of binding to the target polynucleotide in the plant cell, and an anti-CRISPR (ACR) polypeptide capable of reducing the activity of said Cas endonuclease in said plant cell, wherein the ACR has an amino acid sequence 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% 85, and 99-374%, 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 125, at least 125, between 125 and 150, at least 150, between 150 and 175, at least 175, between 175 and 200, or at least 200 contiguous amino acids of a sequence selected from the group consisting of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, and 375-650.

[0017] In one aspect, a method for modulating the activity of a Cas endonuclease with a target polynucleotide in a plant cell is provided, comprising providing to said cell a Cas endonuclease, a guide RNA capable of binding to the target polynucleotide in the plant cell, and an anti-CRISPR (ACR) polynucleotide capable of reducing the activity of said Cas endonuclease in said plant cell, wherein the ACR has a polynucleotide sequence 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% 85, and 99-374%, 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 250, between 250 and 500, at least 500, between 500 and 600, or at least 600 contiguous nucleotides of a sequence selected from the group consisting of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, and 99-374.

[0018] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide.

[0019] In one aspect, a method is provided for increasing the ratio of on-target polynucleotide cleavage activity to off-target polynucleotide cleavage activity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide.

[0020] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the specificity is increased by at least 5%, at least 6%, at least 7%, at least 8%, at least 9%, at least 10%, or even greater than 10%, greater than 15%, greater than 20%, or greater than 25% compared to the cleavage ratio of the Cas endonuclease in a sample lacking said ACR polypeptide.

[0021] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein said method is performed during a specific cell cycle.

[0022] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein said method is performed during meiosis.

[0023] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein said method is performed during mitosis.

[0024] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein said method is performed during a specific stage of the plant's development.

[0025] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein said method is performed during a specific stage of the plant's development, wherein said stage is selected from the group consisting of: growth, reproductive, vegetative, and senescence.

[0026] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein said method is performed during a specific time point.

[0027] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein said method is performed in a specific tissue or cell type of the plant, in some embodiments selected from the group consisting of: 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, single cells, protoplasts, embryos, and callus tissue.

[0028] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the Cas endonuclease and the guide RNA are both provided as polynucleotides.

[0029] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together

a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the guide RNA does not solely comprise ribonucleic acids.

[0030] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the Cas endonuclease is provided as a protein and the guide polynucleotide is provided as an RNA molecule.

[0031] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR is provided as a polynucleotide encoding a polypeptide.

[0032] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR is provided as a polypeptide.

[0033] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR is provided concurrently with either the Cas endonuclease or the guide polynucleotide.

[0034] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR is provided prior to the introduction of the Cas endonuclease or the guide polynucleotide.

[0035] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR is provided after the introduction of the Cas endonuclease or the guide polynucleotide.

[0036] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together

a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein a polynucleotide encoding the ACR is pre-integrated into the genome of the cell or organism.

[0037] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein a polynucleotide encoding the Cas endonuclease is pre-integrated into the genome of the cell or organism.

[0038] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the expression or activity of the ACR is inducible.

[0039] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the expression or activity of the ACR is inducible, wherein induction is in response to a condition selected from the group consisting of: temperature, presence or absence of an exogenously-applied molecule, activation or inhibition of an endogenous gene, light, cell cycle, organism phase, tissue or cell type, and environmental stress.

[0040] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR protein comprises a coiled-coil motif.

[0041] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR protein comprises a heptad repeat pattern of amino acids in the pattern of "hxxhcxc", wherein h = a hydrophobic amino acid, c = a charged amino acid, and x = any amino acid.

[0042] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR has an amino acid sequence 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% 85, and 99-374%, 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 125, at least 125, between 125 and 150, at least 150, between 150 and 175, at least 175, between 175 and 200, or at least 200 contiguous amino acids of a sequence selected from the group consisting of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, and 375-650.

[0043] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polynucleotide, wherein the ACR has a polynucleotide sequence 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% 85, and 99-374%, 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 250, between 250 and 500, at least 500, between 500 and 600, or at least 600 contiguous nucleotides of a sequence selected from the group consisting of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, and 99-374.

[0044] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the Cas endonuclease is a Type II-A Cas endonuclease.

[0045] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together

a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the Cas endonuclease is Cas9.

[0046] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the Cas endonuclease is Cpf1.

[0047] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the Cas endonuclease lacks the ability to nick or cleave a target polynucleotide.

[0048] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing to the target polynucleotide a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide.

[0049] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the specificity is increased by at least 5%, at least 6%, at least 7%, at least 8%, at least 9%, at least 10%, or even greater than 10%, greater than 15%, greater than 20%, or greater than 25% compared to the cleavage ratio of the Cas endonuclease in a sample lacking said ACR polypeptide.

[0050] In one aspect, a method is provided increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein said method is performed during a specific cell cycle.

[0051] In one aspect, a method is provided increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein said method is performed during meiosis.

[0052] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas

endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein said method is performed during mitosis.

[0053] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein said method is performed during a specific stage of the plant's development.

[0054] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein said method is performed during a specific stage of the plant's development, wherein said stage is selected from the group consisting of: growth, reproductive, vegetative, and senescence.

[0055] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein said method is performed during a specific time point.

[0056] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein said method is performed in a specific tissue or cell type of the plant, in some embodiments selected from the group consisting of: 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, single cells, protoplasts, embryos, and callus tissue.

[0057] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the Cas endonuclease and the guide RNA are both provided as polynucleotides.

[0058] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the guide RNA does not solely comprise ribonucleic acids.

[0059] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the Cas endonuclease is provided as a protein and the guide polynucleotide is provided as an RNA molecule.

[0060] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR is provided as a polynucleotide encoding a polypeptide.

[0061] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR is provided as a polypeptide.

[0062] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR is provided concurrently with either the Cas endonuclease or the guide polynucleotide.

[0063] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR is provided prior to the introduction of the Cas endonuclease or the guide polynucleotide.

[0064] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR is provided after the introduction of the Cas endonuclease or the guide polynucleotide.

[0065] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein a polynucleotide encoding the ACR is pre-integrated into the genome of the cell or organism.

[0066] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas

endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein a polynucleotide encoding the Cas endonuclease is pre-integrated into the genome of the cell or organism.

[0067] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the expression or activity of the ACR is inducible.

[0068] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the expression or activity of the ACR is inducible, wherein induction is in response to a condition selected from the group consisting of: temperature, presence or absence of an exogenously-applied molecule, activation or inhibition of an endogenous gene, light, cell cycle, organism phase, tissue or cell type, and environmental stress.

[0069] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR protein comprises a coiled-coil motif.

[0070] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR protein comprises a heptad repeat pattern of amino acids in the pattern of "hxxhcxc", wherein h = a hydrophobic amino acid, c = a charged amino acid, and x = any amino acid.

[0071] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the ACR has an amino acid sequence 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% 85, and 99-374%, 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 125, at least 125, between 125 and 150, at least 150, between 150 and 175, at least 175, between 175 and 200, or at least 200 contiguous amino acids of a sequence selected from the group consisting of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, and 375-650.

[0072] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polynucleotide, wherein the ACR has a polynucleotide sequence 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% 85, and 99-374%, 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 250, between 250 and 500, at least 500, between 500 and 600, or at least 600 contiguous nucleotides of a sequence selected from the group consisting of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, and 99-374.

[0073] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the Cas endonuclease is a Type II-A Cas endonuclease.

[0074] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the Cas endonuclease is Cas9.

[0075] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas

endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the Cas endonuclease is Cpf1.

[0076] In one aspect, a method is provided for increasing the efficiency of homologous recombination of a target polynucleotide in a plant cell, comprising introducing together a Cas endonuclease, a guide polynucleotide, and an anti-CRISPR (ACR) polypeptide, wherein the Cas endonuclease lacks the ability to nick or cleave a target polynucleotide.

[0077] In one aspect, a plant cell comprising a Cas endonuclease and an ACR molecule is provided.

[0078] In one aspect, a plant cell comprising a Cas endonuclease and an ACR molecule is provided, wherein said ACR molecule is provided as a polynucleotide by a phage or virus.

[0079] In one aspect, a plant cell comprising a Cas endonuclease, a guide RNA, and an ACR molecule is provided.

[0080] In one aspect, a plant cell comprising a heterologous Cas endonuclease, a guide RNA, and an ACR protein is provided, wherein the guide RNA is capable of binding to a target polynucleotide in the plant's genome.

[0081] In one aspect, a plant cell comprising a Cas endonuclease and an ACR molecule is provided, wherein the plant cell is obtained or derived from a plant selected from the group consisting of: maize, rice, sorghum, rye, barley, wheat, millet, oats, sugarcane, turfgrass, switchgrass, soybean, canola, alfalfa, sunflower, cotton, tobacco, peanut, potato, tobacco, *Arabidopsis*, vegetable, and safflower.

[0082] In one aspect, a plant cell comprising a Cas endonuclease and an ACR molecule is provided, wherein the ACR has an amino acid sequence 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% 85, and 99-374%, 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 125, at least 125, between 125 and 150, at least 150, between 150 and 175, at least 175, between 175 and 200, or at least 200 contiguous amino acids of a sequence selected from the group consisting of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30,

32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, and 375-650.

[0083] In one aspect, a plant cell comprising a Cas endonuclease and an ACR molecule is provided, wherein the ACR has a polynucleotide sequence 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% 85, and 99-374%, 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 250, between 250 and 500, at least 500, between 500 and 600, or at least 600 contiguous nucleotides of a sequence selected from the group consisting of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, and 99-374.

[0084] In one aspect, plant cell comprising a recombinant construct comprising a polynucleotide sequence encoding an ACR protein, operably linked to a heterologous regulatory expression element is provided.

[0085] In one aspect, plant cell comprising a recombinant construct comprising a polynucleotide sequence encoding an ACR protein, operably linked to a heterologous regulatory expression element is provided.

[0086] In one aspect, plant cell comprising a recombinant construct comprising a polynucleotide sequence encoding an ACR protein, operably linked to a heterologous regulatory expression element is provided, wherein said plant cell is selected from the group consisting of: maize, rice, sorghum, rye, barley, wheat, millet, oats, sugarcane, turfgrass, switchgrass, soybean, canola, alfalfa, sunflower, cotton, tobacco, peanut, potato, tobacco, *Arabidopsis*, vegetable, and safflower

[0087] In one aspect, plant cell comprising a recombinant construct comprising a polynucleotide sequence encoding an ACR protein, operably linked to a heterologous regulatory expression element is provided, wherein the heterologous regulatory expression element is inducible in response to a condition selected from the group consisting of: temperature, presence

or absence of an exogenously-applied molecule, activation or inhibition of an endogenous gene, light, cell cycle, organism phase, tissue or cell type, and environmental stress.

[0088] In one aspect, plant cell comprising a recombinant construct comprising a polynucleotide sequence encoding an ACR protein, operably linked to a heterologous regulatory expression element is provided, wherein the ACR protein comprises a coiled-coil motif.

[0089] In one aspect, plant cell comprising a recombinant construct comprising a polynucleotide sequence encoding an ACR protein, operably linked to a heterologous regulatory expression element is provided, wherein the ACR protein comprises a heptad repeat pattern of amino acids in the pattern of "hxxhcxc", wherein h = a hydrophobic amino acid, c = a charged amino acid, and x = any amino acid.

[0090] In one aspect, plant cell comprising a recombinant construct comprising a polynucleotide sequence encoding an ACR protein, operably linked to a heterologous regulatory expression element is provided, wherein the ACR protein comprises an amino acid sequence 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% 85, and 99-374%, 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 125, at least 125, between 125 and 150, at least 150, between 150 and 175, at least 175, between 175 and 200, or at least 200 contiguous amino acids of a sequence selected from the group consisting of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, and 375-650.

[0091] In one aspect, plant cell comprising a recombinant construct comprising a polynucleotide sequence encoding an ACR protein, operably linked to a heterologous regulatory expression element is provided, wherein the polynucleotide sequence encoding the ACR protein 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% 85, and 99-374%, 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 250, between 250 and 500, at least 500, between 500 and 600, or at least 600 contiguous nucleotides of a sequence selected from the group consisting of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, and 99-374.

[0092] In one aspect, a method is provided for characterizing the activity of an anti-CRISPR protein, comprising: (a) obtaining a bacterial host cell comprising a recombinant construct having a CRISPR system having a targeting sequence capable of targeting a genomic target sequence in a virulent phage; (b) introducing a construct comprising a promoter functional in the bacterial host cell operably linked to a polynucleotide encoding a polypeptide to be assayed for anti-CRISPR activity; (c) challenging the bacterial host with the virulent phage; and (d) identifying one or more bacterial colonies having a phage titre substantially similar to a bacterial cell lacking the recombinant construct encoding the CRISPR system having the targeting sequence capable of targeting a genomic target sequence in the virulent phage challenged with the virulent phage.

[0093] In one aspect, a method is provided for identifying an anti-CRISPR protein, comprising: (a) obtaining a first bacterial host cell comprising a recombinant construct having a Type II-A CRISPR system having a targeting sequence capable of targeting a genomic target sequence in a first virulent phage; (b) challenging the first bacterial host with the virulent phage; (c) obtaining a second bacterial host cell comprising a recombinant construct having a Type II-A CRISPR system having a targeting sequence capable of targeting a genomic target sequence in a second virulent phage; (d) challenging the second bacterial host with the second virulent phage; (e) identifying one or more bacterial colonies of the first bacterial host cell having a phage titre substantially similar to a bacterial cell lacking the recombinant construct encoding the CRISPR system having the targeting sequence capable of targeting a genomic target sequence in the first virulent phage challenged with the first virulent phage; (f) identifying one more bacterial colonies of the second bacterial host cell having a phage titre substantially different than a bacterial cell lacking the recombinant construct encoding the CRISPR system having the targeting sequence capable of targeting a genomic target sequence in the second virulent phage challenged with the second virulent phage; (g) sequencing the genomes of the first and second

virulent phages; (h) identifying one or more gene(s) that is(are) present in the first virulent phage but not the second virulent phage; (i) obtaining a third bacterial host cell comprising a recombinant construct having a CRISPR system having a targeting sequence capable of targeting a genomic target sequence in the first virulent phage; (j) introducing a construct comprising a promoter functional in the third bacterial host cell operably linked to a polynucleotide identical to the gene of (h); (k) challenging the bacterial host with the first virulent phage; and (l) identifying one or more bacterial colonies of the third bacterial host cell having a phage titre substantially similar to a bacterial cell lacking the recombinant construct encoding the CRISPR system having the targeting sequence capable of targeting a genomic target sequence in the first virulent phage challenged with the first virulent phage.

[0094] In one aspect, a method is provided for identifying an anti-CRISPR protein, comprising: (a) obtaining a phage that displays virulence against a bacterium comprising a CRISPR; (b) sequencing the genome of the phage; and (c) identifying at least one contiguous polynucleotide of at least 100 bases 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% 85, and 99-374%, 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 a sequence selected from the group consisting of: SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71,73, 75, 77, 79, 81, 83, 85, and 99-374.

[0095] In one aspect, a method is provided for modulating the activity of a Cas endonuclease with a target polynucleotide in a cell, comprising providing an anti-CRISPR polypeptide to the cell, wherein the anti-CRISPR polypeptide modulates the activity of the Cas endonuclease in the cell, wherein the concentration ratio of Cas endonuclease to anti-CRISPR polypeptide is in the range of 1:1000 to 1:100, 1:100 to 1:10, 1:10 to 1:1, 1:1 to 10:1, 10:1 to 100:1, 100:1 to 1000:1, or any concentration ratio between 1:1000 and 1000:1.

[0096] In one aspect, a method is provided for increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a cell, comprising introducing an anti-

CRISPR (ACR) polypeptide to the cell, wherein the ACR polypeptide interacts with the Cas endonuclease, wherein the anti-CRISPR polypeptide modulates the activity of the Cas endonuclease in the cell, wherein the concentration ratio of Cas endonuclease to anti-CRISPR polypeptide is in the range of 1:1000 to 1:100, 1:100 to 1:10, 1:10 to 1:1, 1:1 to 10:1, 10:1 to 100:1, 100:1 to 1000:1, or any concentration ratio between 1:1000 and 1000:1.

[0097] In one aspect, a method is provided for increasing site-specific homologous recombination frequency of a donor polynucleotide in a cell, comprising introducing to the cell an anti-CRISPR (ACR) polypeptide to increase the homologous recombination of the donor polynucleotide by a polynucleotide-guided Cas endonuclease, wherein the anti-CRISPR polypeptide modulates the activity of the Cas endonuclease in the cell, wherein the concentration ratio of Cas endonuclease to anti-CRISPR polypeptide is in the range of 1:1000 to 1:100, 1:100 to 1:10, 1:10 to 1:1, 1:1 to 10:1, 10:1 to 100:1, 100:1 to 1000:1, or any concentration ratio between 1:1000 and 1000:1.

[0098] In one aspect, a cell is provided, wherein the cell comprises a Cas endonuclease and an ACR protein, wherein the concentration ratio of Cas endonuclease to anti-CRISPR polypeptide is in the range of 1:1000 to 1:100, 1:100 to 1:10, 1:10 to 1:1, 1:1 to 10:1, 10:1 to 100:1, 100:1 to 1000:1, or any concentration ratio between 1:1000 and 1000:1; wherein the cell optionally further comprises a heterologous polynucleotide.

[0099] In any of the methods or compositions described herein, the Cas endonuclease may be heterologous to the cell. In any of the methods or compositions described herein, the ACR may be heterologous to the cell. In any of the methods or compositions described herein, the Cas endonuclease and the ACR may be heterologous to the cell and/or to each other.

[0100] In any aspect, the specificity of the Cas endonuclease that is modulated by the ACR may be selected from the group consisting of: cleavage specificity, nicking specificity, binding specificity, or target recognition specificity.

[0101] Any of methods and compositions herein may comprise any of the sequences, motifs, or other features of an ACR described in PCT Application No. PCT/EP2018/060481, herein incorporated by reference in its entirety.

BRIEF DESCRIPTION OF THE DRAWINGS AND THE SEQUENCE LISTING

[0102] 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. The sequence descriptions and sequence listing attached hereto comply with the rules governing nucleotide and amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §§1.821 and 1.825. The sequence descriptions in the sequence listing comprise the three letter codes for amino acids as defined in 37 C.F.R. §§ 1.821 and 1.825, which are incorporated herein by reference.

[0103] **FIG. 1** provides one embodiment of a schematic for the discovery of virulent phages impeding CRISPR-based immunity. (Top) When a virulent phage is used to challenge a bacterium, phage-resistant survivors can be isolated. Six virulent phages infecting *Streptococcus thermophilus* DGCC7854 generated differing frequencies of CRISPR-immune survivors. (Bottom) When comparing these same phages plated on the phage-sensitive wild-type strain DGCC7854 and a CRISPR-immunized mutant targeting a sequence conserved in all six phages, a large reduction in phage titre was expected. Phage D4276 and D1811 suffered a much smaller reduction in titre than the other four related phages. Phage names and associated data are divided according to CRISPR-interacting phenotypes; permissive (white, D5842 and D5843), impeded adaptation (fractal pattern, D1024 and D5891), and restrictive (black, D4276).

[0104] **FIGs. 2A-2D** provide a representative anti-CRISPR activity of StAcrIIA in *Streptococcus thermophilus*. **FIG. 2A**: Genes cloned from the CRISPR restrictive (black capsid) phage D4276 were expressed in the DGCC7854-derived immunized strain, and the resulting transformants were assayed for increased sensitivity to the permissive (white capsid) phage D5842. **FIG. 2B**: Titre of the restrictive (black) cos-type phage D4276 and permissive (white) cos-type phage D5842 on the naïve DGCC7854 or its CR1-immune derivative, carrying either the empty vector pNZ123 or the vector expressing StAcrIIA (pNZAcr). Each column depicts the average of three biological replicates, each of three technical replicates. **FIG. 2C**:Titre of the permissive (white) pac-type phage 2972 on the naïve DGCC7710, a CR1-immune mutant or a CR3-immune mutant carrying either the empty vector pNZ123 or the vector expressing StAcrIIA (pNZAcr). Each column depicts the average of three biological replicates, each of three technical replicates. In FIG. 2B and FIG. 2C, error bars represent the standard deviation, and an asterisk denotes a difference ($p<0.001$) from all other data, while no other strain differed from any other

($p>0.5$) as determined by one-way ANOVA and Tukey HSD test. Figure 2D: Number and characterization of survivors following a phage 2972 challenge of DGCC7710, carrying either the empty vector pNZ123 or the vector expressing StAcrIIA (pNZAcr). The single CR1 acquisition detected in the presence of the ACR targeted the plasmid and not the phage. CR3 acquisitions targeted the phage, as expected. All cells maintained an intact stAcrIIA.

[0105] **FIGs. 3A-3B** provide one embodiment of an anti-CRISPR gene and protein. **FIG. 3A**: Genomic context of the gene encoding the ACR in all phage homologues, as well as the closest non-phage homologue (using blastP). The ACR homologue is centered (filled with dots and labelled as ACR D4276_028; Sfi21_p24; V442_gp51; HMPREF 2991_02915; and 01205p56, respectively), and identified by a locus tag. Gene function, where known, is annotated. ORFs are dark if their predicted protein product shares 50% amino acid identity with another in the dataset consisting of the full phage genomes, and white if they do not. A solid line connects these similar protein products, and the line is dotted where it passes through a protein that does not meet the similarity criterion. HTH annotations and likelihood were obtained from Helix-Turn-Helix motif prediction, and are displayed only if they have at least 50% likelihood. **FIG. 3B**: Protein alignment of all phage ACR homologues, as well as the closest non-phage homologue (using blastP), highlighting only differences. A dash (-) indicates the residue is absent from the protein. Bold residues (KQRREYAQEMDRLEKAFENLD and ENKLDKIIEKIDKL) are predicted coiled-coil motifs (pCoils, >90% confidence) from the consensus, while an "E" (extended) and "H" (helical) indicator above the residue are the consensus predictions obtained from jnet, jhmm and jpssm.

[0106] **FIGs. 4A-4B** provide schematic representation of one example of anti-CRISPR activity against SpCas9. **FIG. 4A**: Generating an immunized *Lactococcus lactis* MG1363 strain. pL2Cas9 contains the SpCas9 module from pCas9 on a pTRKL2 vector backbone. A spacer targeting orf44 of *L. lactis* virulent phage p2 was cloned-in to create a phage-targeting SpCas9. **FIG. 4B**: The titre of virulent phage p2 on its host *L. lactis* MG1363 carrying either pL2Cas9 or pL2Cas9 targeting the phage (pL2Cas9-44), and either the empty vector pNZ123 or the vector expressing StAcrIIA (pNZAcr). The titre was assayed by spot test, and each bar represents an average of three biological replicates, each of three technical replicates. Only plaques with typical morphology were counted, although a secondary morphology of tiny plaques occasionally appeared when plated on pL2Cas9-44 with pNZ123. While these plaques were not

reliably countable, the maximum threshold at which they appeared is depicted by a patterned box. Error bars represent the standard deviation, and an asterisk denotes a difference ($p<0.001$) from all other data, while no other strain differed from any other ($p>0.5$) as determined by one-way ANOVA and Tukey HSD test.

[0107] **FIGs. 5A-5B** provide a characterization of the "tiny-plaque" phenotype. When phage p2 is plated on its host carrying pL2Cas9-44, its titre is greatly reduced and a tiny plaque phenotype emerges (see Figure 1). **FIG. 5A**: By ECOI assays (also depicted by following every downward-pointing arrow in FIG. 5B), where the phage is passaged on the restrictive host carrying pL2Cas9-44 then plated on a permissive host, we determined that $1/30^{th}$ of phage particles were still able to infect the restrictive host and release wild type (orange) progeny. The pL2Cas9-44 system was determined, in this way, to be "leaky". **FIG. 5B**: (Top 2 rows) A phage population is mixed, containing some pre-existing mutants in *orf44* able to bypass the CRISPR-Cas system (top phage depicted on left). We attributed all of the large plaques (0.01% of the population) in FIG. 1 to pre-existing mutants able to replicate effectively on the restrictive host. When sequenced, all of these plaques contained pure populations mutated in the region targeted by pL2Cas9-44. (Bottom 2 rows) In contrast, the vast majority of phages are wild type, and will be replicating by occasionally 'leaking' through the restrictive strain; a typical burst of ~100 phages would result in only 3 productive infections. This phage replication, due to the leaky system, does not result in visible plaques as the effective burst size (burst size * ECOI) is too low. If plated on an indicator strain, however, it is clear that wild type phage (bottom phage depicted on left) plaques exist, they are simply not visible on the restrictive host. (middle row) The small plaques observed in FIG. 1 were found to also be mutated in the orf44 region (14/15 tested were clearly mutated, although all 15 plaques tested yielded signal consistent with a mixed population containing wild-type phages), and those mutations were indistinguishable from those observed in the large plaques. In fact, when a tiny plaque was passaged again on the restrictive strain, it resulted in large plaques containing pure populations of mutant phages. We attribute the tiny-plaque phenotype to mutations arising in the 'invisible' wild- plaques after several rounds of replication through the leaky pL2Cas9-44 system. Once a mutant emerges, it can replicate efficiently and begins to form a visible plaque – but the delay in generating that mutant constrains the plaque size. This is consistent with the mixed populations observed in the sequence data.

[0108] **FIG. 6** provides a dendrogram of additional, more distant ACR orthologs identified in genomes from *Enterococcus faecalis*, *Enterococcus cecorum*, *Lactobacillus saerimneri*, and *Granulicatella* sp.

[0109] **FIGs. 7A and 7B** provides a representative anti-CRISPR activity of the Acr2 protein (SEQ ID NO:28) in *Streptococcus thermophilus*. **FIG. 7A** Titer of the restrictive (black) *cos*-type phage D1811 and permissive (white) *cos*-type phage D5842 on the naïve DGCC7854 or its CR1-immune derivative, carrying either the empty vector pNZ123 or the vector expressing the *acr2* gene (SEQ ID NO:27) (pNZAcr). Each column depicts the average of three biological replicates, each of three technical replicates. **FIG. 7B** Titer of the permissive (white) *pac*-type phage 2972 on the naïve DGCC7710, a CRISPR1-immune mutant or a CRISPR3-immune mutant carrying either the empty vector pNZ123 or the vector expressing the *acr2* gene (pNZAcr). Each column depicts the average of three biological replicates, each of three technical replicates. In (A) and (B), error bars represent the standard deviation, and an asterisk denotes a difference ($p<0.001$) from all other data, while no other strain differed from any other ($p>0.5$) as determined by one-way ANOVA and Tukey HSD test.

[0110] **FIG. 8** depicts one non-limiting depiction for using a Cas endonuclease to restore function, or abolish function in an ACR gene. In one embodiment, an ACR recombinant gene expression cassette is designed to be non-functional, then following Cas endonuclease expression and RNA guided cleavage or after a sufficient time has passed for site-specific cleavage, converted into a functional expression cassette.

[0111] **FIG. 9** provides one non-limiting depiction for using a Cas endonuclease with an ACR protein and another coding region. In one embodiment, the gene encoding an out-of-frame ACR protein may be combined with other genes (such as, but not limited to, a selectable marker) in a polycistron separated by sequences encoding 'self-cleaving' 2A peptides. Then, following restoration of the ACR ORF, the other genes in the multicistronic expression cassette can also be converted into a functional state.

[0112] **FIG. 10** provides one non-limiting depiction of modulating Cas endonuclease activity in a cell during particular cell cycle(s). In one embodiment, inactivating Cas9 during G1 when HR repair is inactive and permitting Cas9 re-activation during S and G2 when HR repair machinery is expressed and active may increase the frequency of homologous recombination in a cell.

[0113] The following anti-CRISPR gene sequences and anti-CRISPR protein sequences are disclosed as representative, but not limiting, examples in this application:

[0114] **SEQID NO:1** / anti-CRISPR gene isolated from bacteriophage O1205

[0115] ATGGCATACGGAAAAAGCAGATACAACTCATATAGGAAACGCAGTTT
CAATAGAAGCGATAAGCAACGTAGAGAATACGCACAAGAAATGGATAGATTAGAA
CAAACATTTGAAAAACTTGATGGTTGGTATCTATCTAGCATGAAAGATAGTGCGTAT
AAAGATTTCGGAAAATACGAAATTCGCTTATCAAATCATTCAGCAGACAACAAATA
TCATGACCTAGAAAATGGTCGTTTAATTGTTAATGTCAAAGCAAGTAAATTGAAATT
CGTTGATATCAAATGTTACTATAAGGGATTTAAGACAAAGAAGGATGTAATCTAA

[0116] **SEQID NO:2** / anti-CRISPR protein encoded by SEQ ID NO:1
MAYGKSRYNSYRKRSFNRSDKQRREYAQEMDRLEQTFEKLDGWYLSSMKDSAYKDFG
KYEIRLSNHSADNKYHDLENGRLIVNVKASKLKFVDIKCYYKGFKTKKDVI

[0117] **SEQID NO:3** / anti-CRISPR gene isolated from Bacteriophage Sfi21
ATGGCATACGGAAAAAGTAGATATAACTCATATAGAAAGCGCAGTTTTAACAGAAG
TAATAAGCAACGTAGAGAATACGCACAAGAAATGGATAGATTAGAGAAAGCTTTCG
AAAATCTTGACGGATGGTATCTATCTAGCATGAAAGATAGTGCGTACAAAGATTTCG
GAAAATACGAAATTCGCTTATCAAATCATTCAGCAGACAACAAATATCATGACCTA
GAAAATGGTCGTTTAATTGTTAATGTCAAAGCAAGTAAATTGAACTTCGTTGATATC
ATCGAGAACAAACTTGATAAAATCATTGAGAAGATTGATACTCTTGATTTAGATAAG
TACAGATTCATTAATGCTACTAAATTGGAACGTGATATCAAATGCTACTATAAAGGC
TATAAGACAAAGAAGGATGTAATCTAA

[0118] **SEQID NO:4** / anti-CRISPR protein encoded by SEQ ID NO:3
MAYGKSRYNSYRKRSFNRSNKQRREYAQEMDRLEKAFENLDGWYLSSMKDSAYKDF
GKYEIRLSNHSADNKYHDLENGRLIVNVKASKLNFVDIIENKLDKIIEKIDTLDLDKYRFI
NATKLERDIKCYYKGYKTKKDVI

[0119] **SEQID NO:5** / anti-CRISPR gene isolated from Bacteriophage TP-778L
ATGGCATACGGAAAAAGCAGATACAACTCATATAGAAAACGTAGTTTCAACATAAG
TGACACAAAGCGTAGGGAATATGCAAAAGAAATGGAGAAATTAGAACAAGCATTTG
AAAAGCTAGATGGTTGGTATCTATCTAGCATGAAGGATAGTGCATACAAGGATTTTG
GAAAATACGAAATCCGCTTATCAAATCATTCAGCAGACAATAAATATCATGACCTA
GAAAATGGTCGTTTAATTGTTAATGTTAAAGCAAGTAAATTGAACTTCGTTGATATC

ATCGAAAACAAACTTGATAAAATCATCGAGAAGATTGATAAGCTTGATTTAGATAA
GTACAGATTTATTAACGCTACTAGAATGGAGCATGACATTAAATGCTACTATAAAGG
ATTTAAGACAAAGAAAGATGTAATCTAA

[0120] **SEQID NO:6** / anti-CRISPR protein encoded by SEQ ID NO:5

MAYGKSRYNSYRKRSFNISDTKRREYAKEMEKLEQAFEKLDGWYLSSMKDSAYKDFG
KYEIRLSNHSADNKYHDLENGRLIVNVKASKLNFVDIIENKLDKIIEKIDKLDLDKYRFIN
ATRMEHDIKCYYKGFKTKKDVI

[0121] **SEQID NO:7** / anti-CRISPR gene isolated from the genome of *Streptococcus* sp. HMSC072D07

ATGGCATTTGGCAAGAACAGATACAATCCATACAGGAAACGTAGTTTTAATCGTAGT
GATAAACAATGTCGAGAGTATGCTCAGGCAATGGACGAACTAGAACAAGCCTTTGA
GGAACTTGATGGATGGCACTTATCTAGTATGATGGATAGTGCTTATAAGAATTTTGA
AAAGTACCAGGTTCGCCTATCAAATCATTCAGCAGACAACCAATATCATGACTTAGA
AAATGGTTACTTGATTGTCAATGTTAAAGCAAGTAAATTGAACTTTGTCGATATTAT
CGAAAATAAATTGGATAAGATTTTAGAGAAAGTAGACAAGCTTGATCTTGATAAGT
ATAGGTTTATCAATGCGACCAATCTGGAACATGATATTAAATGTTATCTCAAAGGCT
ATAAGACGAAAAAAGACGTGATTTAA

[0122] **SEQID NO:8** / anti-CRISPR protein encoded by SEQ ID NO: 7

MAFGKNRYNPYRKRSFNRSDKQCREYAQAMDELEQAFEELDGWHLSSMMDSAYKNF
EKYQVRLSNHSADNQYHDLENGYLIVNVKASKLNFVDIIENKLDKILEKVDKLDLDKYR
FINATNLEHDIKCYLKGYKTKKDVI

[0123] **SEQID NO:9** / anti-CRISPR gene isolated from Bacteriophage D4276

ATGGCATACGGAAAAAGTAGATATAACTCATATAGAAAGCGCAGTTTTAACAGAAG
TAATAAGCAACGTAGAGAATACGCACAAGAAATGGATAGATTAGAGAAAGCTTTCG
AAAATCTTGACGGATGGTATCTATCTAGCATGAAAGACAGTGCTTACAAGGATTTTG
GGAAATACGAAATTCGCTTATCAAATCATTCGGCAGACAACAAATATCACGACTTA
GAAAACGGTCGTTTAATTGTTAATATTAAAGCTAGTAAATTGAATTTCGTTGATATC
ATCGAGAATAAGCTTGATAAAATAATCGAGAAGATTGATAAGCTTGATTTAGATAA
GTACCGATTCATCAATGCGACCAACCTAGAGCATGATATCAAATGCTATTACAAGGG
GTTTAAAACGAAAAAGGAGGTAATCTAA

[0124] **SEQID NO:10** / anti-CRISPR protein encoded by SEQ ID NO:9

MAYGKSRYNSYRKRSFNRSNKQRREYAQEMDRLEKAFENLDGWYLSSMKDSAYKDF
GKYEIRLSNHSADNKYHDLENGRLIVNIKASKLNFVDIIENKLDKIIEKIDKLDLDKYRFI
NATNLEHDIKCYYKGFKTKKEVI

[0125] **SEQID NO:11** / anti-CRISPR gene isolated from Bacteriophage D1126

ATGGCATACGGAAAAAGCAGATACAATTCATATAGGAAGCGAAACTTCTCTATAAG
CGACAATCAGCGTAGGGAATATGCTAAAAAAAATGAAGGAGTTAGAACAAGCGTTTG
AAAACCTTGACGGATGGTATCTATCTAGCATGAAAGATAGTGCGTACAAAGATTTCG
GAAAATACGAAATTCGCTTATCAAATCATTCAGCAGACAATAGATATCATGACCTAG
AAAATGGTCGCTTAATCGTTAATGTTAAAGCTAGTAAATTGAACTTCGTTGATATCA
TCGAGAATAAACTTGGTAAAATCATTGAGAAGATTGATACTCTTGATTTAGATAAGT
ACAGATTCATTAATGCTACTAAATTGGAACGTGATATCAAATGCTACTATAAAGGCT
ATAAGACAAAGAAGGATGTAATCTAA

[0126] **SEQID NO:12** / anti-CRISPR protein encoded by SEQ ID NO:11

MAYGKSRYNSYRKRNFSISDNQRREYAKKMKELEQAFENLDGWYLSSMKDSAYKDFG
KYEIRLSNHSADNRYHDLENGRLIVNVKASKLNFVDIIENKLGKIIEKIDTLDLDKYRFIN
ATKLERDIKCYYKGYKTKKDVI

[0127] **SEQID NO:13** / anti-CRISPR gene isolated from Bacteriophage D4250

ATGGCATACGGAAAAAGTAGATATAACTCATATAGAAAACGCAGTTTCAACAGAAG
CGATAAACAGCGTAGAGAATACGCACAAGCAATGGAAGAATTAGAGCAAGCATTTG
AAAACTTTGATGATTGGTATCTATCAAGCATGAAAGACAGTGCTTACAAGGATTTTG
GGAAATACGAAATTCGCTTATCAAATCATTCGGCAGACAACAAATATCACGACTTA
GAAAACGGTCGTTTAATTGTTAATATTAAAGCTAGTAAATTGAATTTCGTTGATATC
ATCGAGAATAAGCTTGATAAAATAATCGAGAAGATTGATAAGCTTGATTTAGATAA
GTACCGATTCATCAATGCGACCAACCTAGAGCATGATATCAAATGCTATTACAAGGG
GTTTAAAACGAAAAAGGAGGTAATCTAA

[0128] **SEQID NO:14** / anti-CRISPR protein encoded by SEQ ID NO:13

MAYGKSRYNSYRKRSFNRSDKQRREYAQAMEELEQAFENFDDWYLSSMKDSAYKDFG
KYEIRLSNHSADNKYHDLENGRLIVNIKASKLNFVDIIENKLDKIIEKIDKLDLDKYRFIN
ATNLEHDIKCYYKGFKTKKEVI

[0129] **SEQID NO:15** / anti-CRISPR gene isolated from Bacteriophage D4252

ATGGCATACGGAAAAAGCAGATACAACTCATATAGAAAGCGCAGTTTTAACAGAAG
TGATAAGCAACGTAGAGAATACGCTAAAAAAAATGAAGGAGTTAGAACAAGCGTTTG
AAAACCTTGATGGTTGGTATCTATCGAGCATGAATGACAGTGCTTATAAAAATTTTG
GCAAATATGAAGTTCGATTGTCAAATCATTCGGCAGATAATAAATATCACGACATAG
AAAACGGTCGTTTAATTGTTAATGTTAAAGCTAGTAAATTGAATTTCGTTGATATCA
TCGAGAACAAGCTTGATAAAATAATCGAGAAGATTGATAAGCTTGATTTAGATAAG
TACCGATTCATCAACGCTACCAATCTAGAGCATAATATTAAATGCTATTACAAGGGA
TTTAAGACAAAGAAGGATGTAATATAA

[0130] **SEQID NO:16** / anti-CRISPR protein encoded by SEQ ID NO:15

MAYGKSRYNSYRKRSFNRSDKQRREYAKKMKELEQAFENLDGWYLSSMNDSAYKNF
GKYEVRLSNHSADNKYHDIENGRLIVNVKASKLNFVDIIENKLDKIIEKIDKLDLDKYRFI
NATNLEHNIKCYYKGFKTKKDVI

[0131] **SEQID NO:17** / anti-CRISPR gene isolated from Bacteriophage D4598

ATGGCATACGGAAAAAGTAGATATAACTCATATAGAAAACGCAGTTTCAACAGAAG
CGATAAACAGCGTGGAGAATACGCACAAGCAATGGAAGAATTAGAGCAAGCATTTG
AAAACTTTGATGATTGGTATCTATCAAGCATGAAAGACAGTGCTTACAAGGATTTTG
GGAAATACGAAATTCGCTTATCAAATCATTCGGCAGACAATAAATATCATGACCTAG
AAAATGGTCGCTTAATCGTTAATGTTAAAGCTAGTAAATTGAACTTCGTCGATATCA
TCGAGAATAAAATCGATAAAATCATTGAGAAGATTGATAAGCTTGATTTAGATAAG
TACCGATTCATCAACGCTACCAACCTAGAGCATGATATCAAATGTTATTACAAGGGA
TTTAAGACAAAAAAAGGATGTAATCTAA

[0132] **SEQID NO:18** / anti-CRISPR protein encoded by SEQ ID NO:17

MAYGKSRYNSYRKRSFNRSDKQRGEYAQAMEELEQAFENFDDWYLSSMKDSAYKDF
GKYEIRLSNHSADNKYHDLENGRLIVNVKASKLNFVDIIENKIDKIIEKIDKLDLDKYRFI
NATNLEHDIKCYYKGFKTKKDVI

[0133] **SEQID NO:19** / anti-CRISPR gene isolated from the genome of a *Streptococcus mutans* strain

ATGGCATTTGGAAAAAGAAGATATAACTCGTATCGTAAACGCAGTTTTAATAGAAG
TGATAAGCAACGTCGAGAATATGCACAAGCAATGGAAGAACTTGAACAAACATTTG
AAAATCTTGAAGGTTGGAATTTATCAAGCATGAAAGATAGTGCTTATAAAGATTATG
ATAAATATGAAGTTCGACTTTCAAATCATTCAGCTGATAATCAATATCATAACTTAC

AAGATGGTAAATTAATCATCAATATCAAAGCTAGTAAAATGAATTTTGTTTGGATTA
TAGAAAATAAACTTGATGCAATTCTTGAAAAAGTAAATAAGTTAGACCTTAGCAAA
TACAGATTTATTAATGCTACAAGTTTAGATCATGATATCAAATGTTATTACAAAAAT
TATAAAACAAAGAAAGATGTAATTTAA

[0134] **SEQID NO:20** / anti-CRISPR protein encoded by SEQ ID NO:19

MAFGKRRYNSYRKRSFNRSDKQRREYAQAMEELEQTFENLEGWNLSSMKDSAYKDYD
KYEVRLSNHSADNQYHNLQDGKLIINIKASKMNFVWIIENKLDAILEKVNKLDLSKYRFI
NATSLDHDIKCYYKNYKTKKDVI

[0135] **SEQID NO:21** / anti-CRISPR gene isolated from the genome of a *Streptococcus mutans* strain

ATGGCATTTGGAACAAGAAGATATAATTCATATCGTAAACGCAGTTTTAATAGAAGT
GATAAGCAACGTCGAGAATATGCACAAGCAATGGAAGAACTTGAACAAACATTTGA
AAATCTTGAAGATTGGAATTTGTCGAGCATGAAAGATAGTGCTTATAAAGATTATGA
TAAATATGAAGTTCGACTTTCAAATCATTCAGCTGATAATCAATATCATAACTTACA
AGATGGTAAATTAATCATCAATATCAAAGCTAGTAAAATGAATTTTGTTTGGATTAT
AGAAAATAAACTTGATGCAATTCTTGAAAAAGTAAATAAGTTAGACCTTAGCAGAT
ACAGATTTATTAATGCTACAAATTTAGAACATGATATCAAATGTTATTACAAAAATT
ATAAAACAAAGAAAGATGTAATTTAA

[0136] **SEQID NO:22** / anti-CRISPR protein encoded by SEQ ID NO:21

MAFGTRRYNSYRKRSFNRSDKQRREYAQAMEELEQTFENLEDWNLSSMKDSAYKDYD
KYEVRLSNHSADNQYHNLQDGKLIINIKASKMNFVWIIENKLDAILEKVNKLDLSRYRFI
NATNLEHDIKCYYKNYKTKKDVI

[0137] **SEQID NO:23** / anti-CRISPR gene isolated from the genome of a *Streptococcus mutans* strain

ATGGCATTTGGAACAAGAAGATATAATTCATATCGTAAACGCAGTTTTAATAGAAGT
GATAAGCAACGTCGAGAATATGCACAAGCAATGGAAGAACTTGAACAAACATTTGA
AAATCTTGAAGATTGGAATTTGTCGAGCATGAAAGATAGTGCTTATAAAGATTATGA
TAAATATGAAGTTAGACTTTCAAATCATTCAGCTGATAATCAATATCATAACTTACA
AGATGGTAAATTAATCATCAATATCAAAGCTAGTAAAATGAATTTTGTTTGGATTAT
AGAAAATAAACTTGATGTAATTCTTGAAAAAGTAAATAAGTTAGACCTTAGCAAAT

ACAGATTTATTAATGCTACAAGTTTAGATCATGATATCAAATGTTATTACAAAAATT
ATAAAACAAAGAAAGATGTAATCTAA

[0138] **SEQID NO:24** / anti-CRISPR protein encoded by SEQ ID NO:23

MAFGTRRYNSYRKRSFNRSDKQRREYAQAMEELEQTFENLEDWNLSSMKDSAYKDYD
KYEVRLSNHSADNQYHNLQDGKLIINIKASKMNFVWIIENKLDVILEKVNKLDLSKYRFI
NATSLDHDIKCYYKNYKTKKDVI

[0139] **SEQID NO:25** / anti-CRISPR gene isolated from the genome of a *Streptococcus mutans* strain

ATGGCATTTGGAACAAGAAGATATAATTCATATCGTAAACGCAATTTTAATAGAAGT
GATAAACAACGTCGAGAATATGCACAAGCAATGGAAGAACTTGAACAAACATTTGA
AAATCTTGAAGATTGGAATTTGTCGAGCATGAAAGATAGTGCTTATAAAGATTATGA
TAAATTTGAAGTTCGACTTTCAAATCATTCAGCTGATAATCAATATCATAACTTACA
AGATGGTAAATTAATCATCAATATCAAAGCTAGTAAAATGAATTTTGTTTGGATTAT
AGAAAATAAACTTGATGCAATTCTTGAAAAGGTAAATAAGTTAGACCTTAGCAAAT
ACAGATTTATTAATGCTACAAGTTTAGATCATGATATCAAATGTTATTACAAAAATT
ATAAAACAAAAAAAAGATGTAATTTAA

[0140] **SEQID NO:26** / anti-CRISPR protein encoded by SEQ ID NO:25

MAFGTRRYNSYRKRNFNRSDKQRREYAQAMEELEQTFENLEDWNLSSMKDSAYKDYD
KFEVRLSNHSADNQYHNLQDGKLIINIKASKMNFVWIIENKLDAILEKVNKLDLSKYRFI
NATSLDHDIKCYYKNYKTKKDVI

[0141] **SEQID NO:27** / anti-CRISPR gene isolated from Bacteriophage D1811

ATGAAAATAAATGACGACATCAAAGAGTTAATTTTAGAATATATGAGCCGTTACTTC
AAATTCGAGAACGACTTTTATAAACTGCCAGGCATCAAGTTCACTGATGCAAATTGG
CAGAAGTTCAAAAATGGAGGCACTGACATTGAGAAGATGGGGGCGGCACGAGTAA
ACGCCATGCTCGACTGCCTATTCGACGATTTCGAGCTTGCTATGATTGGCAAGGCTC
AAACTAATTATTACAATGATAATTCACTAAAGATGAACATGCCATTTTACACTTACT
ATGACATGTTCAAAAAGCAGCAACTTCTAAAATGGCTTAAAAATAACCGTGATGAT
GTCATCGGCGGAACTGGTAGGATGTACACAGCAAGTGGTAATTACATTGCTAACGCT
TATTTAGAGGTGGCATTAGAATCGAGCTCGCTTGGTAGTGGCTCTTACATGCTTCAA
ATGAGGTTTAAAGACTATTCAAAAGGTCAAGAACCTATTCCGTCAGGTCGTCAGAAT
CGACTTGAATGGATTGAAAACAATCTCGAAAACATTCGATAA

[0142] **SEQID NO:28** / anti-CRISPR protein encoded by SEQ ID NO:27

MKINDDIKELILEYMSRYFKFENDFYKLPGIKFTDANWQKFKNGGTDIEKMGAARVNA
MLDCLFDDFELAMIGKAQTNYYNDNSLKMNMPFYTYYDMFKKQQLLKWLKNNRDDV
IGGTGRMYTASGNYIANAYLEVALESSSLGSGSYMLQMRFKDYSKGQEPIPSGRQNRLE
WIENNLENIR

[0143] **SEQID NO:29** / anti-CRISPR gene isolated from Bacteriophage D1024

ATGAAAATAAATGACGACATCAAAGAGTTAATTTTAGAATATATGAGCCGTTACTTC
AAATTCGAGAACGACTTTTATAAACTGCCAGGCATCAAGTTCACTGATGCAAATTGG
CAGAAGTTCAAAAATGGAGGCACTGACATTGAGAAGATGGGGGCGGCACGAGTAA
ATGCCATGCTTTCCTGCCTATTCGAGGATTTTGAGCTTGCAATGATTGGCAAGGCTC
AAACTAATTATTACATTGATAACTCACTTAAATTGAACATGCCATTTTACGCTTACTA
TGACATGTTCAAAAAGCAACTTCTTATAAATTGGCTTAAAAATAACCGTGATGATGT
CATCTGCGGAACTGGTAGGATGTACACAGCAAGTGGTAATTACATTGCTAACGCTTA
TTTAGAGGTGGCATTAGAATCTAGCCGTCTGGGTGGTGGTGAGTACATGTTGCAAAT
GCGTTTTAAAAATTATTCAAGAAGTCAAGAACCTATTCCGTCTGGTCGTCAGAATCG
ACTTGAATGGATTGAAAACAATCTTGAAAACATTCGATAA

[0144] **SEQID NO:30** / anti-CRISPR protein encoded by SEQ ID NO:29

MKINDDIKELILEYMSRYFKFENDFYKLPGIKFTDANWQKFKNGGTDIEKMGAARVNA
MLSCLFEDFELAMIGKAQTNYYIDNSLKLNMPFYAYYDMFKKQLLINWLKNNRDDVIC
GTGRMYTASGNYIANAYLEVALESSRLGGGEYMLQMRFKNYSRSQEPIPSGRQNRLEW
IENNLENIR

[0145] **SEQID NO:31** / anti-CRISPR gene isolated from Bacteriophage D4530

ATGAAAATCAATAATGACATCAAAGAGCTAATTTTGGAATATGTAAGTCGCTATTTT
AAATTTGAAAACGACTTCTACAAATTGCCAGGCATCAAATTCACTGATGCAAATTGG
CAAAAATTCAAGAATGGAGATACTTCCATCGAGAAGATGGGGGCAGCACGAGTAAA
TGCCATGCTTGACTGCCTGTTCGAAGATTTTGAGCTTGCAATGATTGGCAAGGCTCA
AACTAATTATTACAATGATAATTCACTAAAGATGAACATGCCATTTTACACTTACTA
TGACATGTTCAAAAAGCAGCAACTTCTAAAATGGCTTAAAAATAACCGTGATGATGT
CATCTGCGGAACTGGTAGGATGTACACAGCAAGTGGTAATTACATTGCTAACGCTTA
TTTAGAAATTGCGTTAGAATCTAGCCGTCTGGGTAGTGGCTCTTACATGCTTCAAAT

GAGATTCAAAGACTATTCAAGAAGTCAAGAACCTATTCCGTCTGGTCGCCAAAATA
GACTAGAATGGATTGAGAGCAACTTGGAAAACATTCGATAA

[0146] **SEQID NO:32** / anti-CRISPR protein encoded by SEQ ID NO:31

MKINNDIKELILEYVSRYFKFENDFYKLPGIKFTDANWQKFKNGDTSIEKMGAARVNAM
LDCLFEDFELAMIGKAQTNYYNDNSLKMNMPFYTYYDMFKKQQLLKWLKNNRDDVIC
GTGRMYTASGNYIANAYLEIALESSRLGSGSYMLQMRFKDYSRSQEPIPSGRQNRLEWI
ESNLENIR

[0147] **SEQID NO:33** / anti-CRISPR gene isolated from Bacteriophage D2759

ATGAAAATCAATAATGACATCAAAGAGCTAATTTTGGAATATGTAAGTCGCTATTTT
AAATTTGAAAACGACTTCTACAAATTGCCAGGCATCAAATTCACTGATGCAAATTGG
CAAAAATTCAAGAATGGAGATACTTCCATCGAGAAGATGGGGGCAGCACGAGTAAA
TGCCATGCTTGACTGCCTATTCGACGATTTTGAGCTTGCTTTGATTGGCAAGGCTCAA
ACTAATTATTACATTGATAACTCACTTAAATTGAACATGCCATTTTACGCTTACTATG
ACATGTTCAAAAAGCAGCAACTTCTAAAATGGCTTAAAAATAACCGTGATGATGTC
ATCTGCGGAACTGGTAGGATGTACACAGCAAGTGGTAATTACATTGCTAACTCTTAT
TTAGAGGTAGCGTTAGAATCTAGCCGTCTGGGTAGTGGCTCTTACATGCTTCAAATG
AGATTCAAAGACTATTCAAGAAGTCAAGAACCTATTCCATCTGGTCGCCAAAATAG
ACTAGAATGGATTGAGAGCAACTTGGAAAACATTCGATAA

[0148] **SEQID NO:34** / anti-CRISPR protein encoded by SEQ ID NO:33

MKINNDIKELILEYVSRYFKFENDFYKLPGIKFTDANWQKFKNGDTSIEKMGAARVNAM
LDCLFDDFELALIGKAQTNYYIDNSLKLNMPFYAYYDMFKKQQLLKWLKNNRDDVICG
TGRMYTASGNYIANSYLEVALESSRLGSGSYMLQMRFKDYSRSQEPIPSGRQNRLEWIE
SNLENIR

[0149] **SEQID NO:35** / anti-CRISPR gene isolated from Bacteriophage D1297

ATGAAAATCAATAATGACATCAAAGAGCTAATTTTGGAATATGTAAGTCGCTATTTT
AAATTTGAAAACGACTTCTACAAATTGCAAGGCATCAAATTCACTGATGCAAATTGG
CAAAAATTCAAGAATGGAGATACTTCCATCGAGAAGATGGGGGCAGCACGAGTAAA
TGCCATGCTTGACTGCCTATTCGACGATTTTGAGCTTGCTTTGATTGGCAAGGCTCAA
CAAGAATACTATTCGGATAATTCCTTGAAATTGAACATGCCATTTTACGCTTACTAT
GACATGTTCAAAAAGCAGCAACTTCTAAAATGGCTTAAAAATAACCGTGATGATGT
CATCTGCGGAACTGGTAGGATGTACACAGCAAGTGGTAATTACATTGCTAACTCTTA

TTTAGAGGTAGCGTTAGAATCTAGCCGTCTGGGTAGTGGCTCTTACATGCTTCAAAT
GAGATTCAAAGACTATTCAAGAAGTCAAGAACCTATTCCGTCAGGTCGCAAAAACC
GACTTGAGTGGATTGAAAACAATCTGGAAAATATTCGATAA

[0150] **SEQID NO:36** / anti-CRISPR protein encoded by SEQ ID NO:35

MKINNDIKELILEYVSRYFKFENDFYKLQGIKFTDANWQKFKNGDTSIEKMGAARVNA
MLDCLFDDFELALIGKAQQEYYSDNSLKLNMPFYAYYDMFKKQQLLKWLKNNRDDVI
CGTGRMYTASGNYIANSYLEVALESSRLGSGSYMLQMRFKDYSRSQEPIPSGRKNRLEW
IENNLENIR

[0151] **SEQID NO:37** / anti-CRISPR gene isolated from Bacteriophage M5728

ATGAAAATCAATAATGACATCAAAGAGCTAATTTTGGAATATGTAAGTCGCTATTTT
AAATTTGAAAACGACTTCTACAAATTGCCAGGCATCAAATTCACTGATGCAAATTGG
CAAAAATTCAAGAATGGAGATACTTCCATCGAGAAGATGGGGGCAGCACGAGTAAA
TGCCATGCTTGACTGCCTATTCGACGATTTTGAGCTTGCTTTTATTGGCAAGGCTCAA
CAAGAATACTATTCGGATAATTCCTTGAAATTGAACATGCCATTTTACGCTTACTAT
GACATGTTCAAAAAGCAGCAACTTCTAAAATGGCTTAAAAATAACCGTGATGATGT
CATCTGCGGAACTGGTAGGATGTACACAGCAAGTGGTAATTACATTGCTAACTCTTA
TTTAGAGGTAGCGTTAGAATCTAGCCGTCTGGGTAGTGGCTCTTACATGCTTCAAAT
GAGATTCAAAGACTATTCAAGAAGTCAAGAACCTATTCCATCTGGTCGCCAAAATA
GACTAGAATGGATTGAAAACAATCTTGAGAATATTCGATAA

[0152] **SEQID NO:38** / anti-CRISPR protein encoded by SEQ ID NO:37

MKINNDIKELILEYVSRYFKFENDFYKLPGIKFTDANWQKFKNGDTSIEKMGAARVNAM
LDCLFDDFELAFIGKAQQEYYSDNSLKLNMPFYAYYDMFKKQQLLKWLKNNRDDVIC
GTGRMYTASGNYIANSYLEVALESSRLGSGSYMLQMRFKDYSRSQEPIPSGRQNRLEWI
ENNLENIR

[0153] **SEQID NO:39** / anti-CRISPR gene isolated from Bacteriophage D4419

ATGAAAATCAATAATGACATCAAAGAGCTAATTTTGGAATATGTAAGTCGCTATTTT
AAATTTGAAAACGACTTCTACAAATTGCCAGGCATCAAATTCACTGATGCAAATTGG
CAAAAATTCAAGAATGGAGATACTTCCATCGAGAAGATGGGGGCAGCACGAGTAAA
TGCCATGCTTGACTGCCTATTCGACGATTTTGAGCTTGCTTTGATTGGCAAGGCTCAA
CAAGAATACTATTCGGATAATTCCTTGAAATTGAACATGCCATTTTACGCTTACTAT
GACATGTTCAAAAAGCAGCAACTTCTAAAATGGCTTAAAAATAACCGTGATGATGT

CATCTGCGGAACTGGTAGGATGTACACAGCAAGTGGTAATTACATTGCTAACTCTTA
TTTAGAGGTAGCGTTAGAATCTAGCCGTCTGGGTAGTGGCTCTTACATGCTTCAAAT
GAGATTCAAAGACTATTCAAGAAGTCAAGAACCTATTCCATCTGGTCGCCAAAATA
GACTAGAATGGATTGAGAGCAACTTGGAAAACATTCGATAA

[0154] **SEQID NO:40** / anti-CRISPR protein encoded by SEQ ID NO:39

MKINNDIKELILEYVSRYFKFENDFYKLPGIKFTDANWQKFKNGDTSIEKMGAARVNAM
LDCLFDDFELALIGKAQQEYYSDNSLKLNMPFYAYYDMFKKQQLLKWLKNNRDDVIC
GTGRMYTASGNYIANSYLEVALESSRLGSGSYMLQMRFKDYSRSQEPIPSGRQNRLEWI
ESNLENIR

[0155] **SEQID NO:41** / anti-CRISPR gene isolated from Bacteriophage D5891

ATGAAAATCAATAATGACATCAAAGAGCTAATTTTGGAATATGTAAGTCGCTATTTT
AAATTTGAAAACGACTTCTACAAATTGCCAGGCATCAAATTCACTGATGCAAATTGG
CAAAAATTCAAGAATGGAGATACTTCCATCGAGAAGATGGGGGCAGCACGAGTAAA
TGCCATGCTTGACTGCCTATTCGACGATTTTGAGCTTGCTTTTATTGGCAAGGCTCAA
CAAGAATACTATTCGGATAATTCCTTGAAATTGAACATGCCATTTTACGCTTACTAT
GACATGTTCAAAAAGCAGCAACTTCTAAAATGGCTTAAAAATAACCGTGATGATGT
CATCTGCGGAACTGGTAGGATGTACACAGCAAGTGGTAATTACATTGCTAACTCTTA
TTTAGAGGTAGCGTTAGAATCTAGCCGTCTGGGTAGTGGCTCTTACATGCTTCAAAT
GAGATTCAAAGACTATTCAAGAAGTCAAGAACCTATTCCATCTGGTCGCCAAAATA
GACTAGAATGGATTGAGAGCAACTTGGAAAACATTCGATAA

[0156] **SEQID NO:42** / anti-CRISPR protein encoded by SEQ ID NO:41

MKINNDIKELILEYVSRYFKFENDFYKLPGIKFTDANWQKFKNGDTSIEKMGAARVNAM
LDCLFDDFELAFIGKAQQEYYSDNSLKLNMPFYAYYDMFKKQQLLKWLKNNRDDVIC
GTGRMYTASGNYIANSYLEVALESSRLGSGSYMLQMRFKDYSRSQEPIPSGRQNRLEWI
ESNLENIR

[0157] **SEQID NO:43** / anti-CRISPR gene isolated from Bacteriophage ALQ13.2

ATGAAAATCAATAATGACATCAAAGAGCTAATTTTGGAATATGTAAGTCGCTATTTT
AAATTTGAAAACGACTTCTACAAATTGCCAGGCATCAAATTCACTGATGCAAATTGG
CAAAAATTCAAGAACGGAGATACTTCCATCGAGAAGATGGGGGCAGCACGAGTAAA
TGCCATGCTTGACTGCCTATTCGACGATTTTGAGCTTGCTTTGATTGGCAAGGCTCAA
CAAGAATACTATTCGGATAATTCCTTGAAATTGAACATGCCATTTTACGCTTACTAT

GACATGTTAAAAAAGCAGCAACTTCTAAAATGGCTTAAAAATAACCGTGATGATGT
CATCTGCGGAACTGGTAGGATGTACACAGCAAGTGGTAATTACATTGCTAACTCTTA
TTTAGAGGTAGCGTTAGAATCTAGCCGTCTGGGTAGTGGCTCTTACATGCTTCAAAT
GAGATTCAAAGACTATTCAAGAAGTCAAGAACCTATTCCATCTGGTCGCCAAAATA
GACTAGAATGGATTGAGAGCAACTTGGAAAACATTCGATGA

[0158] **SEQID NO:44** / anti-CRISPR protein encoded by SEQ ID NO:43

MKINNDIKELILEYVSRYFKFENDFYKLPGIKFTDANWQKFKNGDTSIEKMGAARVNAM
LDCLFDDFELALIGKAQQEYYSDNSLKLNMPFYAYYDMLKKQQLLKWLKNNRDDVIC
GTGRMYTASGNYIANSYLEVALESSRLGSGSYMLQMRFKDYSRSQEPIPSGRQNRLEWI
ESNLENIR

[0159] **SEQID NO:45** / anti-CRISPR gene isolated from Bacteriophage D802

ATGAAAATCAATAATGACATCAAAGAGCTAATTTTGGAATATGTAAGTCGCTATTTT
AAATTCGAGAACGACTTTTACAGATTGCCTGGCATCAAATTCACTGATGCCAACTGG
CAAAAATTCAAGAATGGAGGCACTGCCATTGAGAAGATGGGAGCAGCACGAGTTAA
TGCCATGCTTTCCTGCCTATTCGAGGATTTTGAGCTTGCAATGATTGGCAAGGCTCA
ATATGAATACTATTCGGATAATTCCTTGAAATTGAACATGCCATTTTACGCTTACTAT
GACATGTTCAAAAAGCAGCAACTTCTAAAATGGCTTAAAAATAACCGTGATGATGT
CATCGGCGGAACTGGTAGAATGTACACGTCAAGCGGTAGTTACATTGCTAACGCTTA
TTTAGAAATTGCGTTAGAATCTAGCCGTCTGGGTAGTGGCTCTTACATGCTTCAAAT
GAGATTCAAAGACTATTCAAGAAGTCAAGAACCTATTCCGTCTGGTCGCCAAAATA
GACTTGAATGGATTGAGAGCAACTTGGAAAACATTCGATAA

[0160] **SEQID NO:46** / anti-CRISPR protein encoded by SEQ ID NO:45

MKINNDIKELILEYVSRYFKFENDFYRLPGIKFTDANWQKFKNGGTAIEKMGAARVNA
MLSCLFEDFELAMIGKAQYEYYSDNSLKLNMPFYAYYDMFKKQQLLKWLKNNRDDVI
GGTGRMYTSSGSYIANAYLEIALESSRLGSGSYMLQMRFKDYSRSQEPIPSGRQNRLEWI
ESNLENIR

[0161] **SEQID NO:47** / anti-CRISPR gene isolated from Bacteriophage 73

ATGAAAATCAATAATGACATCAAAGAGCTAATTTTGGAATATGTAAGTCGCTATTTT
AAATTTGAAAACGACTTCTACAAATTACCTGGCATCAAATTCACTGATGCCAACTGG
CAAAAATTCAAGAATGGAGATACTTCCATCGAGAAGATGGGGGCAGCACGAGTAAA
TGCCATGCTTGACTGCCTATTCGAGGATTTTGAGCTTGCAATGATTGGCAAGGCTCA

AACTAATTATTACATTGATAACTCACTTAAATTGAACATGCCATTTTACGCTTACTAT
GACATGTTCAAAAAGCAGCAACTTCTAAAATGGCTTAAAAATAACCGTGATGATGT
CATCGGCGGAACTGGTAGGATGTACACAGCAAGTGGTAATTACATTGCTAACGCTTA
TTTAGAGGTGGCATTAGAATCGAGCTCGCTTGGTAGTGGCTCTTACATGCTTCAAAT
GAGGTTTAAAGACTATTCAAAAGGTCAAGAACCTATTCCGTCAGGTCGTCAGAATCG
ACTTGAATGGATTGAAAACAATCTCGAAAACATTCGATAA

[0162] **SEQID NO:48** / anti-CRISPR protein encoded by SEQ ID NO:47

MKINNDIKELILEYVSRYFKFENDFYKLPGIKFTDANWQKFKNGDTSIEKMGAARVNAM
LDCLFEDFELAMIGKAQTNYYIDNSLKLNMPFYAYYDMFKKQQLLKWLKNNRDDVIG
GTGRMYTASGNYIANAYLEVALESSSLGSGSYMLQMRFKDYSKGQEPIPSGRQNRLEWI
ENNLENIR

[0163] **SEQID NO:49** / anti-CRISPR gene isolated from Bacteriophage DT1

ATGAAAATCAATAATGACATCAAAGAGCTAATTTTGGAATATGTAAGTCGCTATTTT
AAATTTGAAAACGACTTCTACAAATTACCTGGCATCAAATTCACTGATGCCAACTGG
CAAAAATTCAAGAATGGAGAAACTTCAATCGAAAAAATGGGAGCAGCACGAGTTAA
TGCCATGCTTTCATGCCTATTCGAGGATTTTGAGCTTGCAATGATTGGCAAGGCTCA
AACTAATTATTACATTGATAACTCACTTAAATTGAACATGCCATTTTACGCTTACTAT
GACATGTTCAAAAAGCAACTTCTTATAAATTGGCTTAAAAATAACCGTGATGATGTC
ATCGGCGGAACTGGTAGGATGTACACAGCAAGTGGTAATTACATTGCTAACGCTTAT
TTAGAGGTGGCATTAGAATCAAGCTCGCTTGGTAGTGGCTCTTACATGATCCAAATG
AGGTTTAAAGACTATTCAAAAGGTCAAGAACCTATTCCGTCAGGTCGCAAAAACCG
ACTTGAGTGGATTGAAAACAATCTGGAAAATATTCGATAA

[0164] **SEQID NO:50** / anti-CRISPR protein encoded by SEQ ID NO:49

MKINNDIKELILEYVSRYFKFENDFYKLPGIKFTDANWQKFKNGETSIEKMGAARVNAM
LSCLFEDFELAMIGKAQTNYYIDNSLKLNMPFYAYYDMFKKQLLINWLKNNRDDVIGG
TGRMYTASGNYIANAYLEVALESSSLGSGSYMIQMRFKDYSKGQEPIPSGRKNRLEWIE
NNLENIR

[0165] **SEQID NO:51** / anti-CRISPR gene isolated from Bacteriophage D1427

ATGAAAATCAATAATGACATCAAAGAGCTAATTTTGGAATATGTAAGTCGCTATTTT
AAATTTGAAAACGACTTCTACAAATTACCTGGCATCAAATTCACTGATGCCAACTGG
CAAAAATTCAAGAATGGAGAAACTTCAATCGAAAAAATGGGAGCAGCACGAGTTAA

TGCCATGCTTTCATGCCTATTCGAGGATTTTGAGCTTGCAATGATTGGCAAGGCTCA
AACTAATTATTACATTGATAACTCACTTAAATTGAACATGCCATTTTACGCTTACTAT
GACATGTTCAAAAAGCAACTTCTTATAAATTGGCTTAAAAATAACCGTGATGATGTC
ATCGGCGGAACTGGTAGGATGTACACAGCAAGTGGTAATTACATTGCTAACGCTTAT
TTAGAGGTGGCATTAGAATCAAGCTCGCTTGGTAGTGGCTCTTACATGATCCAAATG
AGGTTTAAAGACTATTCAAAAGGTCAAGAACCTATTCCGTCAGGTCGCCAAAATAG
ACTAGAATGGATTGAGAGCAACTTGGAAAACATTCGATAA

[0166] **SEQID NO:52** / anti-CRISPR protein encoded by SEQ ID NO:51

MKINNDIKELILEYVSRYFKFENDFYKLPGIKFTDANWQKFKNGETSIEKMGAARVNAM
LSCLFEDFELAMIGKAQTNYYIDNSLKLNMPFYAYYDMFKKQLLINWLKNNRDDVIGG
TGRMYTASGNYIANAYLEVALESSSLGSGSYMIQMRFKDYSKGQEPIPSGRQNRLEWIE
SNLENIR

[0167] **SEQID NO:53** / anti-CRISPR gene isolated from Bacteriophage N1162

ATGAAAATCAATAATGACATCAAAGAGCTAATTTTGGAATATGTAAGTCGCTATTTT
AAATTTGAAAACGACTTCTACAAATTACCTGGCATCAAATTCACTGATGCCAACTGG
CAAAAATTCAAGAATGGAGATACTTCCATCGAGAAGATGGGGGCAGCACGAGTAAA
TGCCATGCTTGACTGCCTATTCGAGGATTTTGAGCTTGCAATGATTGGCAAGGCTCA
AACTAATTATTACATTGATAACTCACTTAAATTGAACATGCCATTTTACGCTTACTAT
GACATGTTCAAAAAGCAGCAACTTCTAAAATGGCTTAAAAATAACCGTGATGATGT
CATCGGCGGAACTGGTAGGATGTACACATCAACCGGTAATTACATTGCTAACGCTTA
TTTAGAAATTGCGTTAGAATCTAGCCGTCTGGGTAGTGGCTCTTACATGATCCAAAT
GAGGTTTAAAGACTATTCAAAAGGTCAAGAACCTATTCCGTCTGGTCGTCAGAATCG
ACTTGAATGGATTGAAAACAATCTGGAAAATATTCGATAA

[0168] **SEQID NO:54** / anti-CRISPR protein encoded by SEQ ID NO:53

MKINNDIKELILEYVSRYFKFENDFYKLPGIKFTDANWQKFKNGDTSIEKMGAARVNAM
LDCLFEDFELAMIGKAQTNYYIDNSLKLNMPFYAYYDMFKKQQLLKWLKNNRDDVIG
GTGRMYTSTGNYIANAYLEIALESSRLGSGSYMIQMRFKDYSKGQEPIPSGRQNRLEWIE
NNLENIR

[0169] **SEQID NO:55** / anti-CRISPR gene isolated from Bacteriophage D1018

ATGAAAATCAATAATGATATCAAAGAGCTAATTTTGGAATATGTAAGTCGCTATTTT
AAATTTGAAAACGACTTCTACAAATTGCCAGACATCAAGTTCACAGATGCTAATTGG

CAAAAATTTAAGAATGGAGAAACTTCAATCGAAAAAATGGGAGCAGCACGAGTAA
ATGCCATGCTTGACTGCCTATTCGAGGATTTTGAGCTTGCAATGATTGGCAAGGCTC
AAACTAATTATTACATTGATAACTCACTTAAATTGAACATGCCATTTTACGCTTACTA
TGACATGTTCAAAAAGCAGCAACTTCTAAAATGGCTTAAAAATAACCGTGATGATGT
CATCGGCGGAACTGGTAGGATGTACACATCAACCGGTAATTACATTGCTAACGCTTA
TTTAGAAATTGCGTTAGAATCTAGCCGTCTGGGTAGTGGCTCTTACATGATCCAAAT
GAGGTTTAAAGACTATTCAAAAGGTCAAGAACCTATTCCGTCTGGTCGTCAGAATCG
ACTTGAATGGATTGAAAACAATCTGGAAAATATTCGATAA

[0170] **SEQID NO:56** / anti-CRISPR protein encoded by SEQ ID NO:55

MKINNDIKELILEYVSRYFKFENDFYKLPDIKFTDANWQKFKNGETSIEKMGAARVNAM
LDCLFEDFELAMIGKAQTNYYIDNSLKLNMPFYAYYDMFKKQQLLKWLKNNRDDVIG
GTGRMYTSTGNYIANAYLEIALESSRLGSGSYMIQMRFKDYSKGQEPIPSGRQNRLEWIE
NNLENIR

[0171] **SEQID NO:57** / anti-CRISPR gene isolated from Bacteriophage D3577

ATGAAAATAAACAACGATATCAAAGAGCTAATTTTGGAATACGCTAAACGTTATTTC
AAGTTTGAAAACGACTTCTACAAACTGCCAGACATCAAATTCACTGATGCCAACTGG
CAAAAATTTAAGAATGGAGAAACTTCCATCGAAAAAATGGGAGCAGCACGAGTTAA
TGCCATGCTTTCCTGCCTGTTCGACGATTTTGAGCTTGCTATGATTGGCAAGGCTCAA
ACTAATTATTACAATGATAACTCACTTAAATTGAACATGCCATTTTACGCTTACTATG
ACATGTTCAAAAAGCAGCAACTTCTAAAATGGCTTAAAAATAACCGTGATGATATC
ATCTGCGGAACTGGTAGAATGTACACTTCAAGAGGTAGTTACATTGCTAACGCTTAT
TTAGAGGTAGCGTTAGAATCAAGCTTGCTTGGTAGTGGCTCTTACATGCTTCAAATG
AGGTTCAAAGACTATTCAAAAAGTCAAGAACCTATTCCATCTGGTCGTCAGAATCGA
CTTGAATGGATTGAGAGCAACTTGGAAAATATTCGATAA

[0172] **SEQID NO:58** / anti-CRISPR protein encoded by SEQ ID NO:57

MKINNDIKELILEYAKRYFKFENDFYKLPDIKFTDANWQKFKNGETSIEKMGAARVNAM
LSCLFDDFELAMIGKAQTNYYNDNSLKLNMPFYAYYDMFKKQQLLKWLKNNRDDIICG
TGRMYTSRGSYIANAYLEVALESSLLGSGSYMLQMRFKDYSKSQEPIPSGRQNRLEWIE
SNLENIR

[0173] **SEQID NO:59** / anti-CRISPR gene isolated from Bacteriophage CHPC577

ATGAAAATAAACAACGATATCAAAGAGCTAATTTTGGAATATGGAAGTCGCTATTTT
AAATTTGAAAACGACTTCTACAAACTGCCTGGCATCAAGTTCACTGATGCTAATTGG
CAAAAATTCAAAAATGGTGATACTTTAATCGAAAAAATGGGGGCAGCACGAGTAAA
TGCCATGCTTGACTGCCTGTTCGACGATTTTGAGCTTGCTATGATTGGCAAGGCTCA
AACTAATTATTACAATGATAATTCCTTGAAATTGAACATGCCATTTTACGCTTACTAT
GACATGTTCAAAAAGCAACAGCTTATACATTGGCTCAAAAACAACCGTGATGACAT
CGTAGGCGGAACTGGTAGACTGTACACTTCAAGCGGTAGTTACATTGCTAACGCTTA
TTTAGAAATTGCATTAGAATCGAGCTCGCTTGGTAGTGGCTCTTACATGCTTCAAAT
GAGATTCAAAAACTATTCAAAAAGTCAAGAACCTATTCCATCTGGTCGCCAGAATCG
ACTTGAATGGATTGAAAACAATCTTGAGAATATTCGATAA

[0174] **SEQID NO:60** / anti-CRISPR protein encoded by SEQ ID NO:59

MKINNDIKELILEYGSRYFKFENDFYKLPGIKFTDANWQKFKNGDTLIEKMGAARVNAM
LDCLFDDFELAMIGKAQTNYYNDNSLKLNMPFYAYYDMFKKQQLIHWLKNNRDDIVG
GTGRLYTSSGSYIANAYLEIALESSSLGSGSYMLQMRFKNYSKSQEPIPSGRQNRLEWIE
NNLENIR

[0175] **SEQID NO:61** / anti-CRISPR gene isolated from Bacteriophage D4237

ATGAAAATAAATAACGACATCAAAGAATTAATTTTAGAATATATGAGCCGTTACTTC
AAATTCGAAAACGACTTCTACAAATTGCCAGACATCAAGTTCACAGATGCTAATTGG
CAAAAATTTAAGAATGGAGAAACTTCAATCGAAAAAATGGGAGCAGCACGAGTTAA
TGCCATGCTCAACTGCCTATTCGAAGATTTTGAGCTTGCTATGATTGGCAAGGCTCA
AATTAATTATTACAATGATAACTCACTTAAAATGAACATGCCATTTTACGCTTACTAT
GATATGTTCAAAAAACAACAGCTTCTAAAATGGCTTAAAGATCACCATGATGACATC
ATCGGAGGAGCTGGCAGAATGTACACATCAACCGGTAGTTACATTGCTAATGCTTAT
TTAGAGGTAGCGTTAGAATCAAGCTCGCTTGGTGATGGTGAGTACATGTTGCAAATG
CGTTTTAAAAATTATTCACGAAGTCAAGAACCTATTCCGTCAGGTCGCAAAAACCGA
CTTGAGTGGATTGAAAACAATCTGGAAAATATTCGATAA

[0176] **SEQID NO:62** / anti-CRISPR protein encoded by SEQ ID NO:61

MKINNDIKELILEYMSRYFKFENDFYKLPDIKFTDANWQKFKNGETSIEKMGAARVNAM
LNCLFEDFELAMIGKAQINYYNDNSLKMNMPFYAYYDMFKKQQLLKWLKDHHDDIIG
GAGRMYTSTGSYIANAYLEVALESSSLGDGEYMLQMRFKNYSRSQEPIPSGRKNRLEWI
ENNLENIR

[0177] **SEQID NO:63** / anti-CRISPR gene isolated from Bacteriophage 9874

ATGAAAATAAATGACGACATCAAAGAATTAATTTTAGAATATATGAGCCGTTACTTC
AAATTCGAGAACGACTTCTACAAATTGCCTGACATCAAATTCACTGATGCCAACTGG
CAAAAATTCAAAAATGGAGATACTTCCATCGAGAAGATGGGGGCAGCACGAGTAAA
TGCCATGCTTGACTGCCTATTCGAAGATTTCGAACTTGCCATGATTGGCAAGGCTCA
ACAAGAATACTATTTGGATAATTCACTAAAGATGAACATGCCATTTTACGCTTATTA
TGATATGTTCAAGAAAAAACAGCTCGTCAAATGGCTTAAAGATCACCATGATGACA
TCCTAGGCGGAACTGGTAGGATGTACACTTCAGACGGTAGTTACATTGCTAACTCTT
ATTTAGAGGTAGCGTTAGAATCTAGCCGTCTGGGTAGTGGCTCTTACATGCTTCAAA
TGAGATTCAAAGACTATTCAAGAAGTCAAGAACCCATTCCGTCAGGTCGCAAAAAC
CGACTTGAGTGGATTGAAAACAATCTGGAAAATATTCGATAA

[0178] **SEQID NO:64** / anti-CRISPR protein encoded by SEQ ID NO:63

MKINDDIKELILEYMSRYFKFENDFYKLPDIKFTDANWQKFKNGDTSIEKMGAARVNA
MLDCLFEDFELAMIGKAQQEYYLDNSLKMNMPFYAYYDMFKKKQLVKWLKDHHDDI
LGGTGRMYTSDGSYIANSYLEVALESSRLGSGSYMLQMRFKDYSRSQEPIPSGRKNRLE
WIENNLENIR

[0179] **SEQID NO:65** / anti-CRISPR gene isolated from Bacteriophage 5093

ATGGAAATCAACAACGATATCAAAGAGTTAATTTTGGAATACGTGAAAAGATACTT
CAAGTTCGAGAACGACTTCTACAAATTGCCTGACATCAAATTCACTGATGCCAACTG
GCAGAAGTTCAAAAATGGCGAAACAGCCATTGAGAAGATGGGGGCAGCACGAGTA
AACGCAATGCTCGACTGCCTATTCGAAGATTTTGAGCTTGCCATGATTGGCAAGGCT
CAAACTAATTATTATATTGATAACTCGCTTAAATTAAACATGCCATTTTATGCTTACT
ATGATATGTTTAAGAAACAACAGCTCGTCAAATGGCTTGAAACTAGTCGTGAAGAC
ATCATCGGAGGGGCTGGCAGAATGTACACTTCAGACGGTAGTTACATTGCTAACGCT
TATTTAGAAGTAGCGTTAGAATCAAGCTCGCTTGGTGATAGTGAATACATGTTGCAA
ATGCGTTTTAAAAATTATTCAAAAAGTCAAGAACCTATTCCGTCTGGTCGTCAAAAT
AGACTGGAATGGATTGAAAACAATCTTAAAAACATTCGATAA

[0180] **SEQID NO:66** / anti-CRISPR protein encoded by SEQ ID NO:65

MEINNDIKELILEYVKRYFKFENDFYKLPDIKFTDANWQKFKNGETAIEKMGAARVNA
MLDCLFEDFELAMIGKAQTNYYIDNSLKLNMPFYAYYDMFKKQQLVKWLETSREDIIG

GAGRMYTSDGSYIANAYLEVALESSSLGDSEYMLQMRFKNYSKSQEPIPSGRQNRLEWI
ENNLKNIR

[0181] **SEQID NO:67** / anti-CRISPR gene isolated from Bacteriophage D4154

ATGCTAATAAATAACGACATCAAAGAGTTGATTTTGGAATACGTCAAACGCTATTTT
AAATATGAAAATGACTTCTACAGATTGCCGGGCATCAAGTTTACCGATGCAAATTGG
CAGAAGTTTAAAAATGGCGACACTTCCATCGAGAAGATGGGGGCAGCACGAGTAAA
CGCCATGCTCGACTGCCTATTCGAAGATTTTGAGCTTGCCATGATTGGTAAGGCTCA
AACCAATTATTATATCAATAATTCATTGAAAATGAATATGCCGTTTTACGCTTACTAT
GATATGTTCAAGAAGGAACAGCTTATGAAATGGCTTGAAACCAGCCGTGAAGACAT
CATAGGCGGAACTGGCAGGATGTACACTTCAGACGGTAGTTACATTGCTAACGCTTA
TTTGGAAATTGCATTAGAATCGAGCTCGCTTGGTAGTGGCTCTTACATGCTTCAAAT
GCGTTTTAAAGATTATTCAAAAGGTCAAGAGCCTATCCCGTCTGGTCGTCAAAACCG
ACTTGAGTGGATTGAAAACAATCTTGAAAACATTCGATAA

[0182] **SEQID NO:68** / anti-CRISPR protein encoded by SEQ ID NO:67

MLINNDIKELILEYVKRYFKYENDFYRLPGIKFTDANWQKFKNGDTSIEKMGAARVNA
MLDCLFEDFELAMIGKAQTNYYINNSLKMNMPFYAYYDMFKKEQLMKWLETSREDIIG
GTGRMYTSDGSYIANAYLEIALESSSLGSGSYMLQMRFKDYSKGQEPIPSGRQNRLEWIE
NNLENIR

[0183] **SEQID NO:69** / anti-CRISPR gene isolated from the genome of *Streptococcus thermophilus* DGCC11758

ATGCTAATAAATAACGACATCAAAGAGTTGATTTTGGAATACGTCAAACGCTATTTT
AAATTTGAAAATGACTTCTACAGATTGCCGGGCATCAAGTTTACCGATGCAAATTGG
CAGAAGTTTAAAAATGGCGACACTGCCATTGAGAAGATGGGGGCATCACGAGTAAA
CTCTATGCTTGACTGCCTGTTCGAAGATTTTGAGCTTGCTATGATTGGCAAGGCTCAA
GATGAATACTATTTGGATAATTCACTAAAGATGAACATGCCATTTTACGCTTATTAT
GATATGTTCAAGAAAAAACAGCTCGTCAAATGGCTTAAAGATCACCATGATGACAT
CCTAGGCGGAACTGGTAGGATGTATACTTCAAGCGGCAATTACATTGCTAACGCTTA
TTTAGAGGTAGCGTTAGAATCAAGCTCGCTTGGTAGTGGCTCTTACATGATTCAAAT
GCGTTTTAAAAATTATTCAAAAGGTCAAGAGCCTATCCCGTCTGGTCGTCAAAACCG
ACTTGAGTGGATTGAAAAAAACTTGGAGAACATTCGATAA

[0184] **SEQID NO:70** / anti-CRISPR protein encoded by SEQ ID NO:69

MLINNDIKELILEYVKRYFKFENDFYRLPGIKFTDANWQKFKNGDTAIEKMGASRVNSM
LDCLFEDFELAMIGKAQDEYYLDNSLKMNMPFYAYYDMFKKKQLVKWLKDHHDDILG
GTGRMYTSSGNYIANAYLEVALESSSLGSGSYMIQMRFKNYSKGQEPIPSGRQNRLEWI
EKNLENIR

[0185] **SEQID NO:71** / anti-CRISPR gene isolated from the genome of *Streptococcus thermophilus* DSM 20617

ATGGAAATCAACAACGATATTAAACAACTGATCTTGGAATACGCTAAACGTTATTTC
AAGTTTGAGAACGACTTTTATAAACTGCCAGGCATCAAGTTCACTGATGCAAATTGG
CAGAAGTTCAAAAATGGAGGCACTGCCATTGAGAAGATGGGGGCAGCACGAGTAA
ACGCCATGCTCGACTGCCTATTCGAAGATTTCGAGCTTGCAATGATTGGCAAGGCTC
AACAAGAATACTATTCGGATAATTCCTTGAAAGTAAATATGGCATTCTATGCTTATT
ACGATCAATTCAAAAAACAACAGCTTATGAAATGGCTTAAAGATAATCACGATGAC
ATCATAGGAGGGACTGGTAGAATGTACACGTCAAGCGGTAGTTACATTGCTAACGC
TTATTTAGAAATTGCGTTAGAATCTAGCCGTCTGGGTGGTGGTTCTTACATGATCCA
AATGAGGTTTAAAGACTATTCAAAAGGTCAAGAACCTATTCCGTCTGGTCGTCAGAA
TCGACTTGAATGGATTGAGAGCAACTTGGAAAACATTCGATAA

[0186] **SEQID NO:72** / anti-CRISPR protein encoded by SEQ ID NO:71

MEINNDIKQLILEYAKRYFKFENDFYKLPGIKFTDANWQKFKNGGTAIEKMGAARVNA
MLDCLFEDFELAMIGKAQQEYYSDNSLKVNMAFYAYYDQFKKQQLMKWLKDNHDDII
GGTGRMYTSSGSYIANAYLEIALESSRLGGGSYMIQMRFKDYSKGQEPIPSGRQNRLEWI
ESNLENIR

[0187] **SEQID NO:73** / anti-CRISPR gene isolated from Bacteriophage Sfi19

ATGGAAATCAACAACGACATTAAACAACTGATCTTGGAATACGTGGGACGCTATTTT
AAATTTGAAAATGACTTCTACAAATTGCCCGGCATCAAATTCACTGATGCCAATTGG
CAGAAGTTCAAAAATGGCGATACTTCCATCGAAAAGATGGGAGCAGCACGAGTAAA
CGCAATGCTTGACTGCCTGTTCGAAGATTTCGAACTTGCCATGATTGGCAAGGCTCA
AACTAATTATTATATTGATAATTCCCTTAAATTAAACATGCCATTTTACGCTTATTAT
GATATGTTCAAGAAGGAACAGCTTATGAAATGGCTTAAAGATCACCATGATGACAT
CATAGGCGGAACTGGTAGGATGTACATTTCAAGCGGTAGCTACATTGCTAACGCTTA
TTTGGAAATTGCACTAGAATCAAGTACGCTTGGTGGTGGTGAGTACATGTTGCAAAT

GCGCTTTAAAAATTATTCACGAAGCCAAGAACCTATTCCATCAGGTCGCAAAAATAG
ACTTGAATGGATTGAAAACAATCTTGAAAACATTCGATAA

[0188] **SEQID NO:74** / anti-CRISPR protein encoded by SEQ ID NO:73

MEINNDIKQLILEYVGRYFKFENDFYKLPGIKFTDANWQKFKNGDTSIEKMGAARVNA
MLDCLFEDFELAMIGKAQTNYYIDNSLKLNMPFYAYYDMFKKEQLMKWLKDHHDDII
GGTGRMYISSGSYIANAYLEIALESSTLGGGEYMLQMRFKNYSRSQEPIPSGRKNRLEWI
ENNLENIR

[0189] **SEQID NO:75** / anti-CRISPR gene isolated from Bacteriophage Sfi11

ATGGAAATCAACAACGACATTAAACAACTGATCTTGGAATACGTGGGACGCTATTTT
AAATTTGAAAATGACTTCTACAAATTGCCCGGCATCAAATTCACTGATGCCAATTGG
CAGAAGTTCAAAAATGGCGATACTTCCATCGAAAAGATGGGAGCAGCACGAGTAAA
CGCAATGCTTGACTGCCTGTTCGAAGATTTCGAACTTGCCATGATTGGCAAGGCTCA
AACTAATTATTATATTGATAATTCCCTTAAATTAAACATGCCATTTTACGCTTATTAT
GATATGTTCAAGAAGGAACAGCTTATGAAATGGCTTAAAGATCACCATGATGACAT
CATAGGCGGAACTGGTAGGATGTACACTTCAAGCGGTAGCTACATTGCTAACGCTTA
TTTGGAAATTGCACTAGAATCAAGTACGCTTGGTGGTGGTGAGTACATGTTGCAAAT
GCGCTTTAAAAATTATTCACGAAGCCAAGAACCTATTCCATCAGGTCGCAAAAATAG
ACTTGAATGGATTGAAAACAATCTTGAAAACATTCGATAA

[0190] **SEQID NO:76** / anti-CRISPR protein encoded by SEQ ID NO:75

MEINNDIKQLILEYVGRYFKFENDFYKLPGIKFTDANWQKFKNGDTSIEKMGAARVNA
MLDCLFEDFELAMIGKAQTNYYIDNSLKLNMPFYAYYDMFKKEQLMKWLKDHHDDII
GGTGRMYTSSGSYIANAYLEIALESSTLGGGEYMLQMRFKNYSRSQEPIPSGRKNRLEWI
ENNLENIR

[0191] **SEQID NO:77** / anti-CRISPR gene isolated from the genome of *Streptococcus thermophilus* M17PTZA496

ATGGAAATCAACAAAGACATCAAAGAGTTGATTTTGGAATACGTCAAACGCTATTTT
AAATTTGAAAATGATTTCTACAGATTGCCGGGCATCAAGTTTACCGATGCCAACTGG
CAAAAATTCAAGAATGGAGATACTTCCATCGAGAAGATGGGGGCAGCACGAGTAAA
CGCCATGCTTGACTGCCTGTTCGAAGATTTCGAACTTGCTATGATTGGCAAGGCTCA
AGATGAATACTATTTGGATAATTCACTTAAGTTTAATATGGCATTCCATACTTATTAC
GATCAATTTAAAAAACAACAGCTTATGAAATGGCTTGAAACTAGCCTCGAAGACAT

CATAGGCGGAACTGGTAGGATGTACACTTCAAGCGGTAGTTACATTGCTAACGCTTA
TTTGGAAATTGCACTAGAATCAAGCTCGCTTGGTGGTGGTGAGTACATGTTGCAAAT
GCGTTTTAAAAATTATTCACGAAGCCAAGAACCTATTCCGTCAGGTCGCAAAAACCG
ACTTGAGTGGATTGAAAACAATCTGGAAAATATCCGATAA

[0192] **SEQID NO:78** / anti-CRISPR protein encoded by SEQ ID NO:77

MEINKDIKELILEYVKRYFKFENDFYRLPGIKFTDANWQKFKNGDTSIEKMGAARVNAM
LDCLFEDFELAMIGKAQDEYYLDNSLKFNMAFHTYYDQFKKQQLMKWLETSLEDIIGG
TGRMYTSSGSYIANAYLEIALESSSLGGGEYMLQMRFKNYSRSQEPIPSGRKNRLEWIEN
NLENIR

[0193] **SEQID NO:79** / anti-CRISPR gene isolated from Bacteriophage D4769

ATGAAAATCAATAATGACATCAAAGAGCTAATTTTGGAATATGTAAGTCGCTATTTT
AAATTTGAAAACGACTTCTACAAATTACCTGGCATCAAATTCACTGATGCCAACTGG
CAAAAATTCAAGAATGGAGATACTTCCATCGAGAAGATGGGGGCAGCACGAGTAAA
CGCCATGCTTGACTGCCTGTTCGAAGATTTCGAACTTGCTATGATTGGCAAGGCTCA
AGATGAATACTATTTGGATAATTCACTTAAGTTTAATATGGCATTCCATACTTATTAC
GATCAATTTAAAAAACAACAGCTTATGAAATGGCTTGAAACTAGCCTCGAAGACAT
CATAGGCGGAACTGGTAGGATGTACACTTCAAGCGGTAGTTACATTGCTAACGCTTA
TTTGGAAATTGCACTAGAATCAAGCTCGCTTGGTGGTGGTGAGTACATGTTGCAAAT
GCGTTTTAAAAATTATTCACGAAGCCAAGAACCTATTCCGTCAGGTCGCAAAAACCG
ACTTGAGTGGATTGAAAACAATCTGGAAAATATCCGATAA

[0194] **SEQID NO:80** / anti-CRISPR protein encoded by SEQ ID NO:79

MKINNDIKELILEYVSRYFKFENDFYKLPGIKFTDANWQKFKNGDTSIEKMGAARVNAM
LDCLFEDFELAMIGKAQDEYYLDNSLKFNMAFHTYYDQFKKQQLMKWLETSLEDIIGG
TGRMYTSSGSYIANAYLEIALESSSLGGGEYMLQMRFKNYSRSQEPIPSGRKNRLEWIEN
NLENIR

[0195] **SEQID NO:81** / anti-CRISPR gene isolated from Bacteriophage D5691

ATGATTATAAATATTGATATCAAGGAATTGATTTTAGAGTATATGAGTAGATACTTC
AAATTTGAAAATGATTTCTACAAACTCCCCGGCATCAAATTCACTGATGCCAATTGG
CAAAAATTTAAGAATGGTGACACTTCCATCGAAAAGATGGGAGCGGCTCGAGTAAA
TGCCATGCTCGACTGTCTATTCGATGACTTTGAACTTGCTATGATTGGCAAGGCTCA
AATTAATTATTACATAGACAATTCCCTTAAATTGAACATGCCATTCTATGCTTATTAT

GACATGTTCAAAAAACAACAACTGATCAAATGGATTGAAACCAGCCGTGATGATGT
CATCGGAGGAACTGGCAGGATGTATACAGCAAGCGGAAGCTACATAGCTAACGCTT
ATCTAGAAATAGCACTAGAATCTAGCTCTCTGGGTGGTGGCTCTTATATGCTTCAAA
TGAGATTCAAAAACTACTCACGAAGCCAAGAGCCAATACCATCTGGTCGGAAAAAC
CGACTTGAGTGGATTGAGAGCAACTTGGAAAACATTAGATAA

[0196] **SEQID NO:82** / anti-CRISPR protein encoded by SEQ ID NO:81

MIINIDIKELILEYMSRYFKFENDFYKLPGIKFTDANWQKFKNGDTSIEKMGAARVNAML
DCLFDDFELAMIGKAQINYYIDNSLKLNMPFYAYYDMFKKQQLIKWIETSRDDVIGGTG
RMYTASGSYIANAYLEIALESSSLGGGSYMLQMRFKNYSRSQEPIPSGRKNRLEWIESNL
ENIR

[0197] **SEQID NO:83** / anti-CRISPR gene isolated from the genome of *Streptococcus* sp. HMSC10E12

ATGGAAATCAACAATGACATCAAAGAGTTAATCTTGGAATACGTGGGACGCTATTTC
AAGTTTGAAAATGATTTTTACAAATTGCCGGGCATCAAATTTACCGATGCAAATTGG
CAAAAATTCAAAAACGGTGATACATCCATCGAGAAAATGGGGGCGGCACGAGTAAA
CGCAATGCTCGACTGCCTATTCGATGATTTCGAGCTTGCTATGATTGGCAAGGCTCA
AACTGATTATTACATTGATAACTCACTTAAATTGAACATGCCATTTTATGCTTATTAT
GACATGTTCAAAAAACAACAGCTTCTAAAATGGATTGAGAATAGTCGTGAAGACAT
CATCGGAGGGGCTGGCAGAATGTACACAGCGGGCGGTAATTGGATTTCTAGCGCTT
ATTTAGAGATCGCATTAGAATCTAGTTCCATCGGTGGCGGTGGCTATATGCTTCAAA
TGCGGTTCAAAAACTACTCAAGAGACCCTAGACCGATTCCAGCAGGCCACCAAAAT
CGTCTCGAATGGATTGAAAACAACTTGGAGAATATCCGATAA

[0198] **SEQID NO:84** / anti-CRISPR protein encoded by SEQ ID NO:83

MEINNDIKELILEYVGRYFKFENDFYKLPGIKFTDANWQKFKNGDTSIEKMGAARVNAM
LDCLFDDFELAMIGKAQTDYYIDNSLKLNMPFYAYYDMFKKQQLLKWIENSREDIIGGA
GRMYTAGGNWISSAYLEIALESSSIGGGGYMLQMRFKNYSRDPRPIPAGHQNRLEWIEN
NLENIR

[0199] **SEQID NO:85** / anti-CRISPR gene isolated from the genome of *Streptococcus* sp. HSISS2

ATGGAAATCAACAATGACATCAAGGACCTAATTTTAGAATACGTAGGACGATATTTT
CGATTTGAAAACGACTTCTACAAACTTCCCAGAATCAAGTTTACCGATTCCAATTGG

CAAAAATTCAAGAACGGTGACACTTCCATCGAAAAAATGGGAGCTGGCAGAGTGAA
CGCAATGCTCGATTGTCTATTTGATGATTTTGAGCTTGCTATGATTGGTAAGGCTCAA
ACCGATTACTACATGGACAATTCTTTAAAGATGAATATGCCATTTTATGCCTATTATG
ACCAATTTAAGAAACAGCAACTATTGAAATGGATCGAGAATAGTAGAGAGGATATC
ATAGGCGGTGCTGGCAGAATGTACACAGCTAGTGGGAATTGGATTTCTAGTGCCTAT
TTAGAAATTGCATTGGAATCCAGCTCGTTAGGTGGTGGTGAGTACATGTTGCAAATG
CGTTTCAAAGACTACTCACGAAGCCAAGAGCCGATACCAGCAGGCCGCCAGAATCG
ACTTGAGTGGATTGAGAATAATTTGGAGAATATTCGATAA

[0200] **SEQID NO:86** / anti-CRISPR protein encoded by SEQ ID NO:85
MEINNDIKDLILEYVGRYFRFENDFYKLPRIKFTDSNWQKFKNGDTSIEKMGAGRVNAM
LDCLFDDFELAMIGKAQTDYYMDNSLKMNMPFYAYYDQFKKQQLLKWIENSREDIIGG
AGRMYTASGNWISSAYLEIALESSSLGGGEYMLQMRFKDYSRSQEPIPAGRQNRLEWIE
NNLENIR

[0201]

DETAILED DESCRIPTION

[0202] Compositions and methods are provided for novel anti-CRISPR ("ACR") polynucleotides and polypeptides as well as methods of use of such polynucleotides and polypeptides. The abbreviation "ACR" as used herein may be used as an alternative notation for "ACR polypeptide", "ACR protein", or "ACR polynucleotide", consistent with context. The disclosed methods include methods for inhibiting the activity of CRISPR-Cas complexes from modifying target DNA molecules. Accordingly, the disclosed compositions and methods find a wide range of uses in genome editing applications, particularly in plants.

[0203] The CRISPR-Cas system bases its utility as a genome-editing tool on its native function as an immune system in prokaryotes. The very first demonstration of its activity against bacterial viruses (phages) was also the first record of phages evading that immunity. This evasion can be due to point mutations, DNA modifications, or specific phage-encoded proteins that interfere with the CRISPR-Cas system, known as anti-CRISPRs (ACRs). The latter category is of considerable biotechnological interest, as these ACRs can serve as off-switches for CRISPR-based genome-editing. Every ACR characterized to date has originated from temperate phages, genomic islands, or prophages – and they have all been identified due to properties

shared with the first ACR discovered, such as an association with helix-turn-helix motifs. Here, with a phage-oriented approach, we provide entirely novel ACRs in a virulent phage of *Streptococcus thermophilus.* In challenging an *S. thermophilus* strain CRISPR-immunized against a set of related virulent phages, we found one phage that evaded the CRISPR-Cas system at greater than 40000 times the rate of the others. We then identified an ACR solely responsible for the abolished immunity. We extended our findings by demonstrating anti-CRISPR activity in another *S. thermophilus* strain, against unrelated phages, and in another bacterial genus immunized using the heterologous *Streptococcus pyogenes Cas9 (*SpCas9) system commonly used in genome-editing. This ACR has the largest effect on SpCas9 activity demonstrated to date. Our phage-oriented approach is likely to serve to uncover many more ACRs. We also identified a second ACR also having anti-CRISPR activity against the *S. thermophilus* strain.

[0204] Disclosed herein are methods of identifying an ACR, methods of using an ACR to modulate the activity of a Cas endonuclease, particularly in a cell, particularly in a plant cell, and exemplary but not limiting compositions of ACR polypeptides, and polynucleotides encoding the same.

Definitions

[0205] 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.

[0206] 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-naturally occurring, 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.

[0207] 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.

[0208] “Open reading frame” is abbreviated ORF.

[0209] The term "selectively hybridizes" or “selective hybridization” 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.

[0210] The term "stringent conditions" or “stringent hybridization conditions” includes reference to conditions under which a polynucleotide/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 polynucleotide/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 polynucleotide/probe is fewer than about 1000 nucleotides in length, fewer than 500 nucleotides, fewer than 100 nucleotides, fewer than 90 nucleotides, fewer than 80 nucleotides, fewer than 70 nucleotides, fewer than 60 nucleotides, fewer than 50 nucleotides, fewer than 40 nucleotides, fewer than 30 nucleotides, fewer than 20 nucleotides, 10 nucleotides, or even fewer than 10 nucleotides. 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 30°C for short polynucleotides/probes (e.g., 10 to 50 nucleotides) and at least 60°C for long polynucleotides/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 sulfate) 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.

[0211] 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 similarity to undergo homologous recombination with the corresponding genomic region. "Sufficient similarity" indicates that two polynucleotide sequences have sufficient structural equivalency to act as substrates for a homologous recombination reaction. The structural equivalency 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.

[0212] 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 a target site or, alternatively, also comprises a portion of a 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 similarity to undergo homologous recombination with the corresponding region of homology.

[0213] 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.

[0214] "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.

[0215] 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%, 95%, 96%, 97%, 98%, 99%, 100%, or any incremental or fractional percentage from 50% to 100%. These identities can be determined using any of the programs described herein.

[0216] 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.

[0217] 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%, 95%, 96%, 97%, 98%, 99%, 100%, or any incremental or fractional 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%.

[0218] 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, the association of an atom or a molecule to an existing nucleotide in a polynucleotide (for example but not limited to: a covalent addition of a methyl group, or an ionic interaction with a metal ion), the chemical alteration of at least one nucleotide, or any combination of the preceding. 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.

[0219] 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.

[0220] 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 embodiments, 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.

[0221] The term “fragment” refers to a contiguous set of polynucleotides or polypeptides. In one embodiment, 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 polynucleotides. In one embodiment, 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 polypeptides. A fragment may or may not exhibit the function of a sequence sharing some percent identity over the length of said fragment.

[0222] 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 promoter sequence.

[0223] "Variants" is intended to mean substantially similar sequences. For polynucleotides, a variant comprises a deletion and/or addition of one or more nucleotides at one or more sites within the native polynucleotide and/or a substitution of one or more nucleotides at one or more sites in the native polynucleotide. As used herein, a "native" or “wild type” polynucleotide or polypeptide comprises a naturally occurring nucleotide sequence or amino acid sequence, respectively. For polynucleotides, conservative variants include those sequences that, because of the degeneracy of the genetic code, encode the amino acid sequence of a polypeptide disclosed herein. Naturally occurring allelic variants such as these can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques as outlined below. Variant polynucleotides also include synthetically derived polynucleotides, such as those generated, for example, by using site-directed mutagenesis, and which may encode a polypeptide. Generally, variants of a particular polynucleotide disclosed herein will have at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to that particular polynucleotide (e.g., to the ACR sequences of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, and 99-374) as determined by sequence alignment programs and parameters described elsewhere herein or known in the art.

[0224] "Variant" protein is intended to mean a protein derived from the native protein by deletion or addition of one or more amino acids at one or more sites in the native protein and/or substitution of one or more amino acids at one or more sites in the native protein. In some embodiments, a variant proteins disclosed herein include those that are biologically active, that is they continue to possess biological activity of the native protein. Such variants are referred to as "functional variants", "biologically active variant" or "active variant" interchangeably herein, and may result from, for example, genetic polymorphism or human manipulation.

[0225] "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.

[0226] By the term "endogenous" it is meant a sequence or other molecule that naturally occurs in a cell or organism. In one aspect, an endogenous polynucleotide is normally found in the genome of the cell from which it is obtained; that is, not heterologous.

[0227] 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.

[0228] "Coding sequence" refers to a polynucleotide sequence that may be transcribed into an RNA molecule and optionally further translated into a polypeptide. "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.

[0229] 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 embodiments of the disclosure, the mutated gene comprises an alteration that results from a

guide polynucleotide/Cas endonuclease system as disclosed herein. A mutated plant is a plant comprising a mutated gene.

[0230] 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 endonuclease system as disclosed herein.

[0231] 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 protein; 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).

[0232] 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 protein (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.

[0233] By "domain" it is meant a contiguous stretch of nucleotides (that can be RNA, DNA, and/or RNA-DNA-combination sequence) or amino acids.

[0234] 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.

[0235] 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.

[0236] An "optimized" polynucleotide is a sequence that has been optimized for improved expression or function in a particular heterologous host cell.

[0237] A "plant-optimized nucleotide sequence" is a nucleotide sequence that has been optimized for expression or function 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 endonuclease 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.

[0238] 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.

[0239] 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.

[0240] "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).

[0241] "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.

[0242] "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 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.

[0243] 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.

[0244] 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.

[0245] "Introducing" is intended to mean presenting to the organism, such as a cell or organism, the polynucleotide or polypeptide or a polynucleotide-protein complex (e.g. an engineered CRISPR-Cas 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. The methods and compositions 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 polynucleotide or polypeptide to the cell.

[0246] 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 embodiments, 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*.

[0247] 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.

[0248] 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.

[0249] 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.

[0250] 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.

[0251] 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.

[0252] A "mature" protein refers to a post-translationally processed polypeptide (i.e., one from which any pre- or propeptides present in the primary translation product have been removed).

[0253] "Precursor" protein refers to the primary product of translation of mRNA (i.e., with pre- and propeptides still present). Pre- and propeptides may be but are not limited to intracellular localization signals.

[0254] "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.

[0255] 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. The "effector complex" of a CRISPR system includes Cas proteins involved in crRNA and target recognition and binding. Some of the component Cas proteins may additionally comprise domains involved in target polynucleotide cleavage.

[0256] The term "Cas protein" refers to a polypeptide encoded by a Cas (CRISPR-associated) gene. A Cas protein includes but is not limited to: the novel Cas9 orthologs 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 protein may be a "Cas endonuclease" , 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. A Cas endonuclease described herein comprises one or more nuclease domains. A Cas protein is further defined as a functional fragment or functional variant of a native Cas protein, 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% 85, and 99-374%, 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 protein, and retains at least partial activity.

[0257] A "functional fragment", "fragment that is functionally equivalent" and "functionally equivalent fragment" of a Cas endonuclease are used interchangeably herein, and

refer to a portion or subsequence of the Cas endonuclease 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 endonuclease can comprise a complete or partial (functional) peptide of any one of its domains such as for example, but not limiting to a complete or functional part of a HD domain, a complete or functional part of a helicase domain, a complete or functional part of an endonuclease domain, a complete or functional part of a PAM-interacting domain, a complete or functional part of a Wedge domain, a complete or functional part of an RuvC domain, a complete or functional part of a zinc-finger domain, or a complete or functional part of a Cas protein (such as but not limiting to a Cas9, Cpf1, Cas5, Cas5d, Cas7, Cas8b1, Cas1, Cas2, Cas4, or Cas9 ortholog).

[0258] The terms "functional variant", "variant that is functionally equivalent" and "functionally equivalent variant" of a Cas endonuclease or Cas endonuclease, including Cas9 ortholog described herein, are used interchangeably herein, and refer to a variant of the Cas endonuclease 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.

[0259] In some aspects, a functional fragment or functional variant retains about the same level and type (*e.g.*, target polynucleotide recognition, binding, and cleavage) of activity as the parental molecule from which it was derived. In some aspects, a functional fragment or functional variant displays improved activity of the same type (*e.g.*, increased specificity of target polynucleotide recognition) as the parental molecule from which it was derived. In some aspects, a functional fragment or functional variant displays reduced activity of the same type (*e.g.*, lower target polynucleotide binding affinity) as the parental molecule from which it was derived. In some aspects, a functional fragment or functional variant displays partial activity (*e.g.* polynucleotide recognition and binding, but not cleavage) as the parental molecule from which it was derived. In some aspects, a functional fragment or functional variant displays a different type of activity (*e.g.*, creation of a single-strand nick on a target polynucleotide vs. a double strand break) than the parental molecule from which it was derived. Any similarity or difference in type or level of activity may be chosen as a desired outcome, according to the needs of the practitioner.

[0260] 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 cascade (comprises at least a second protein domain that can form a cascade with other proteins). In one aspect, 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.

[0261] As used herein, the term "guide polynucleotide", relates to a polynucleotide sequence that can form a complex with a Cas endonuclease, including the Cas endonuclease described herein, and enables the Cas endonuclease 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).

[0262] The terms "functional fragment", "fragment that is functionally equivalent" and "functionally equivalent 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.

[0263] The terms "functional variant", "variant that is functionally equivalent" and "functionally equivalent 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.

[0264] 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 endonuclease, wherein said guide RNA/Cas endonuclease complex can direct the Cas endonuclease to a DNA target site, enabling the Cas endonuclease to recognize, optionally bind to, and optionally nick or cleave (introduce a single or double-strand break) the DNA target site.

[0265] 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 embodiments, 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.

[0266] 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 endonuclease 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.

[0267] As used herein, the terms "guide polynucleotide/Cas endonuclease complex", "guide polynucleotide/Cas endonuclease system", " guide polynucleotide/Cas complex", "guide polynucleotide/Cas system" and "guided Cas system" "polynucleotide-guided endonuclease", and "PGEN" are used interchangeably herein and refer to at least one guide polynucleotide and at least one Cas endonuclease, that are capable of forming a complex, wherein said guide polynucleotide/Cas endonuclease complex can direct the Cas endonuclease to a DNA target site, enabling the Cas endonuclease to recognize, bind to, and optionally nick or cleave (introduce a single or double-strand break) the DNA target site. A guide polynucleotide/Cas endonuclease complex herein can comprise Cas protein(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).

[0268] The terms "guide RNA/Cas endonuclease complex", "guide RNA/Cas endonuclease 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 endonuclease that are capable of forming a complex , wherein said guide RNA/Cas endonuclease complex can direct the Cas endonuclease to a DNA target site, enabling the Cas endonuclease to recognize, bind to, and optionally nick or cleave (introduce a single or double-strand break) the DNA target site. In some aspects, the components are provided as a ribonucleoprotein complex ("RNP") of a Cas endonuclease protein and a guide RNA.

[0269] The terms "target site", "target sequence", "target polynucleotide", "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 endonuclease 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.

[0270] 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 endonuclease system described herein. In some aspects, the Cas endonuclease may not successfully recognize a target DNA sequence if the target DNA sequence

is not adjacent to, or near, a PAM sequence. In some aspects, the PAM precedes the target sequence (*e.g.* Cas12a). In some aspects, the PAM follows the target sequence (*e.g. S. pyogenes* Cas9). The sequence and length of a PAM herein can differ depending on the Cas protein or Cas protein 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.

[0271] 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: modifications include deletion, substitution, and/or insertion of one or more nucleotides in the nucleic acid fragment, the association of an atom or a molecule to an existing nucleotide in a polynucleotide (for example but not limited to: a covalent addition of a methyl group, or an ionic interaction with a metal ion), the chemical alteration of at least one nucleotide, or any combination of the preceding.

[0272] 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: modifications include deletion, substitution, and/or insertion of one or more nucleotides in the nucleic acid fragment, the association of an atom or a molecule to an existing nucleotide in a polynucleotide (for example but not limited to: a covalent addition of a methyl group, or an ionic interaction with a metal ion), the chemical alteration of at least one nucleotide, or any combination of the preceding.

[0273] 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.

[0274] As used herein, "donor DNA" is a DNA construct that comprises a polynucleotide of interest to be inserted into the target site of a Cas endonuclease.

[0275] 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.

[0276] As used herein, the term "eukaryote" or "eukaryotic" refers to organisms or cells or tissues derived therefrom belonging to the phylogenetic domain Eukarya such as animals (e.g., mammals, insects, reptiles, and birds), ciliates, plants (e.g., monocots, dicots, and algae), fungi, yeasts, flagellates, microsporidia, and protists.

[0277] Eukaryotic cells include, but are not limited to, human, non-human, animal, mammalian, bacterial, fungal, insect, yeast, and plant cells as well as plants and seeds produced by the methods described herein.

[0278] The term “plant-optimized Cas endonuclease” herein refers to a Cas protein, including a multifunctional Cas protein, encoded by a nucleotide sequence that has been optimized for expression in a plant cell or plant.

[0279] A “plant-optimized nucleotide sequence encoding a Cas endonuclease”, “plant-optimized construct encoding a Cas endonuclease” and a “plant-optimized polynucleotide encoding a Cas endonuclease” are used interchangeably herein and refer to a nucleotide sequence encoding a Cas protein, 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 endonuclease includes a plant comprising the nucleotide sequence encoding for the Cas sequence and/or a plant comprising the Cas endonuclease protein. In one aspect, the plant-optimized Cas endonuclease nucleotide sequence is a maize-optimized, rice-optimized, wheat-optimized, soybean-optimized, cotton-optimized, or canola-optimized Cas endonuclease.

[0280] The term “plant” generically includes whole plants, plant organs, plant tissues, seeds, plant cells, seeds and progeny of the same. 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 embodiment, 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). 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.

[0281] "Progeny" comprises any subsequent generation of a plant.

[0282] 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 invention, provided that these parts comprise the introduced polynucleotides.

[0283] 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.

[0284] 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.

[0285] The term "isoline" is a comparative term, and references organisms that are genetically identical, but differ in treatment. In one example, two genetically identical maize plant embryos may be separated into two different groups, one receiving a treatment (such as the introduction of a CRISPR-Cas effector endonuclease) and one control that does not receive such treatment. In some aspects, "isoline" refers to two cells or organisms that are genetically identical except for the presence of a heterologous polynucleotide or polypeptide that has been introduced as part of an experiment. Any phenotypic differences between the two groups may

thus be attributed solely to the treatment or presence of the heterologous molecule, and not to any inherent property of the organism's endogenous genetic makeup.

[0286] "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.

[0287] A "polynucleotide of interest" includes any nucleotide sequence encoding a protein or polypeptide that improves desirability of crops. 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.

[0288] 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 embodiments, 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 host 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.

[0289] A protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, 5%, or 1% (by dry weight) of contaminating protein. When the polypeptides disclosed herein or biologically active portion thereof is recombinantly produced, optimally culture medium represents less than about 30%, 20%, 10%, 5%, or 1% (by dry weight) of chemical precursors or non-protein-of-interest chemicals.

[0290] The compositions and methods herein may provide for an improved "agronomic trait" or "trait of agronomic importance" or “trait of agronomic interest” to a plant, which may include, but not be limited to, the following: disease resistance, drought tolerance, heat tolerance, cold tolerance, salinity tolerance, metal tolerance, herbicide tolerance, improved water use efficiency, improved nitrogen utilization, improved nitrogen fixation, pest resistance, herbivore resistance, pathogen resistance, yield improvement, health enhancement, vigor improvement, growth improvement, photosynthetic capability improvement, nutrition enhancement, altered protein content, altered oil content, increased biomass, increased shoot length, increased root length, improved root architecture, modulation of a metabolite, modulation of the proteome, increased seed weight, altered seed carbohydrate composition, altered seed oil composition, altered seed protein composition, altered seed nutrient composition, as compared to an isoline plant not comprising a modification derived from the methods or compositions herein.

[0291] "Agronomic trait potential" is intended to mean a capability of a plant element for exhibiting a phenotype, preferably an improved agronomic trait, at some point during its life cycle, or conveying said phenotype to another plant element with which it is associated in the same plant.

[0292] 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 60%, between 60% and 70%, between 70% and 80%, at least 75%, at least 80%, between 80% and 90%, at least 90%, between 90% and 100%, at least 100%, between 100% and 200%, at least 200%, at least 300%, at least 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 60%, between 60% and 70%, between 70% and 80%, at least 75%, at least 80%, between 80% and 90%, at least 90%, between 90% and 100%, at least 100%, between 100% and 200%, at least 200%, at least 300%), at least 400% or more higher than the untreated control.

[0293] 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.

[0294] The meaning of abbreviations is as follows: "sec" means second(s), "min" means minute(s), "h" means hour(s), "d" means day(s), "µL" or "uL" or "ul" 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).

CRISPR-Cas Systems

[0295] In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex (comprising a Cas protein (e.g. a Cas9 protein), a tracr and a crRNA (having a repeat sequence and a spacer, or guide, sequence)) at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). In an engineered CRISPR-Cas9 complex, the natural spacer sequence has been replaced with a sequence designed to be complementary to a target sequence, for example, a target sequence in a eukaryotic cell. In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity. A target sequence can be any polynucleotide, such as DNA or RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell.

[0296] 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).

[0297] 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 protein components. The resulting ribonucleoprotein complex is called a Cascade, that 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).

[0298] The Cas endonuclease 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 1 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. Class 2 CRISPR-Cas systems comprise Types II, V, and VI. A characteristic feature of Class 2 systems is the presence of a single Cas protein instead of an effector module endonuclease complex. Types II and V Cas proteins comprise an RuvC-like endonuclease domain that adopts the RNase H fold.

[0299] Class 2 Type II CRISPR/Cas systems employ a crRNA and tracrRNA (trans-activating CRISPR RNA) to guide the Cas endonuclease 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 endonuclease to cleave the DNA target. For the *S. pyogenes* Cas9 endonuclease, the cleavage leaves a blunt end. 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.

[0300] Cas endonucleases 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 protein or gRNA. 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).

CRISPR-Cas System Compositions

[0301] Methods and compositions are provided for genome editing with a CRISPR Associated (Cas) endonuclease. Class I Cas endonucleases comprise multisubunit effector complexes (Types I, III, and IV), while Class 2 systems comprise single protein effectors (Types II, V, and 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). In Class 2 Type II systems, the Cas endonuclease acts in complex with a guide RNA (gRNA) that directs the Cas endonuclease to cleave the DNA target to enable target recognition, binding, and cleavage by the Cas endonuclease. The gRNA comprises a Cas endonuclease recognition (CER) domain that interacts with the Cas endonuclease, and a Variable Targeting (VT) domain that hybridizes to a nucleotide sequence in a target DNA . In some aspects, the gRNA comprises a CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA) to guide the Cas endonuclease 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, forming an RNA duplex. In some aspects, the gRNA is a "single guide RNA" (sgRNA) that comprises a synthetic fusion of crRNA and tracrRNA. In many systems, the Cas endonuclease-guide polynucleotide complex recognizes a short nucleotide sequence adjacent to the target sequence (protospacer), called a "protospacer adjacent motif" (PAM).

[0302] Examples of a Cas endonuclease include but are not limited to Cas9 and Cpf1. Cas9 (formerly referred to as Cas5, Csn1, or Csx12) is a Class 2 Type II Cas endonuclease (Makarova *et al.* 2015, *Nature Reviews Microbiology* Vol. 13:1-15). A Cas9-gRNA complex recognizes a 3' PAM sequence (NGG for the *S. pyogenes* Cas9) at the target site, permitting the

spacer of the guide RNA to invade the double-stranded DNA target, and, if sufficient homology between the spacer and protospacer exists, generate a double-strand break cleavage. Cas9 endonucleases comprise RuvC and HNH domains that together produce double strand breaks, and separately can produce single strand breaks. For the *S. pyogenes* Cas9 endonuclease, the double-strand break leaves a blunt end. Cpf1 is a Class 2 Type V Cas endonuclease, and comprises nuclease RuvC domain but lacks an HNH domain (Yamane et al., 2016, Cell 165:949-962). Cpf1 endonucleases create "sticky" overhang ends.

[0303] A large number of Cas9 orthologs are known in the art as well as their associated tracrRNA and crRNA components (see, e.g., "Supplementary Table S2. List of bacterial strains with identified Cas9 orthologs," Fonfara, Ines, et al., "Phylogeny of Cas9 Determines Functional Exchangeability of Dual-RNA and Cas9 among Orthologous Type II CRISPR/Cas Systems," Nucleic Acids Research 42.4 (2014): 2577-2590, including all Supplemental Data; Chylinski K., et al., "Classification and evolution of type II CRISPR-Cas systems," Nucleic Acids Research, 2014; 42(10):6091-6105, including all Supplemental Data; Kevin M Esvelt, K. M., et al., (2013) "Orthogonal Cas9 proteins for RNA-guided gene regulation and editing," Nature Methods 10, 1116-1121, a number of orthogonal Cas9 proteins identified including a Cas9 protein from *Neisseria meningitidis*). A representative list of Type-II CRISPR systems that find use with the compositions and methods disclosed herein, includes those described in Makarova et al. 2015, Nature Reviews Microbiology | AOP, published online 28 September 2015; doi:10.1038/nrmicro3569; and in Burstein, D. et al. New CRISPR–Cas systems from uncultivated microbes. Nature http://dx.doi.org/10.1038/nature21059 (2016) and in WO 2017 062 855. In some embodiments, the Cas endonuclease is identified from a Type II-A CRISPR complex, such as those derived from *Streptococcus thermophilus*, or *Streptococcus pyogenes*.

[0304] In some aspects, a "polynucleotide modification template" is provided 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, deletion, or chemical alteration. 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.

[0305] In some aspects, a polynucleotide of interest is inserted at a target site and provided as part of a "donor DNA" molecule. As used herein, "donor DNA" is a DNA construct that comprises a polynucleotide of interest to be inserted into the target site of a Cas endonuclease. 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. 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). The amount of homology or sequence identity shared by a target and a donor polynucleotide can vary and includes total lengths and/or regions.

[0306] To facilitate optimal expression and nuclear localization for eukaryotic cells, the gene comprising the Cas endonuclease 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.

[0307] In some aspects, the Cas endonuclease is provided as a polypeptide. In some aspects, the Cas endonuclease is provided as a polynucleotide encoding a polypeptide. In some aspects, the guide RNA is provided as a DNA molecule encoding one or more RNA molecules. In some aspects, the guide RNA is provide as RNA or chemically-modified RNA. In some aspects, the Cas endonuclease protein and guide RNA are provided as a ribonucleoprotein complex (RNP).

[0308] Also of use with the ACR compositions and methods provided herein are Cas endonuclease variants that have a reduced activity towards off-target sequences. Such Cas endonuclease variants include those disclosed, for example, in WO2016 205 613. Such combinations may provide for an even greater reduction in off-target activity.

CRISPR-Cas Mediated Genome Editing

[0309] As used herein, CRISPR-Cas-mediated genome editing composition (or, an engineered CRISPR-Cas complex) refers to the elements of a CRISPR system needed to carry out CRISPR-Cas-mediated genome editing in a host cell, such as a eukaryotic cell. Engineered

CRISPR-Cas complex compositions typically include one or more nucleic acids comprising a crRNA, a tracrRNA (or chimeric thereof also referred to a guide RNA or single guide RNA) and a Cas enzyme, for example, Cas9. The crRNA and tracrRNAs of engineered -Cas complex compositions can also be provided to the system indirectly by nucleic acids encoding the crRNA, tracrRNA and/or guide RNA. The CRISPR/Cas-mediated genome editing composition can optionally include a donor polynucleotide that can be recombined into the target cell's genome at or adjacent to the target site (e.g., the site of single or double stand break induced by the Cas9). Examples of engineered CRISPR-Cas complexes include those disclosed in U.S. Publication No. 2015/0045546 and International Application publication number WO 2013/176772.

[0310] Some uses for Cas9-gRNA systems at a genomic target site include but are not limited to insertions, deletions, substitutions, or modifications of one or more nucleotides at the target site; 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.

[0311] The process for editing a genomic sequence at a Cas9-gRNA double-strand-break site with a modification template generally comprises: providing a host cell with a Cas9-gRNA complex that recognizes a target sequence in the genome of the host cell and is able to induce a double-strand-break 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 double-strand break. Genome editing using double-strand-break-inducing agents, such as Cas9-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 WO2016025131 published on 18 February 2016.

[0312] To facilitate optimal expression and nuclear localization for eukaryotic cells, the gene comprising the Cas endonuclease 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. In some aspects, the Cas endonuclease is provided as a polypeptide. In some aspects, the Cas endonuclease is provided as a polynucleotide encoding a polypeptide. In some aspects, the guide RNA is provided as a DNA molecule encoding one or more RNA molecules. In some aspects, the guide RNA is provide as RNA or chemically-modified RNA. In some aspects, the Cas endonuclease protein and guide RNA are provided as a ribonucleoprotein complex (RNP).

Anti-CRISPR (ACR) Proteins

[0313] Compositions disclosed herein include isolated polynucleotides and polypeptides encoding anti-CRISPR ("ACR") proteins. In some embodiments, the disclosed ACR polypeptides are capable of reducing or inhibiting the ability of Cas endonuclease, for example but not limited to a Cas9 protein, to recognize, bind, and optionally modify, nick, or cleave a target polynucleotide.

[0314] In one embodiment polynucleotides and polynucleotide encoding polypeptides are provided which reduce and/or inhibit Cas9 activity against a target DNA molecule. In certain embodiments, polypeptides that reduce and/or inhibit the activity of Type II-A Cas9 proteins are provided.

[0315] isolated or identified from a bacteriophage or bacterium

[0316] In one embodiment, isolated or recombinant polynucleotides are provided which comprise a nucleotide sequence set forth in SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, and 99-374), or functional fragments or variants thereof. Also provided are recombinant polynucleotides that encode the polypeptides having a sequence set forth in SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, and 375-650, or functional fragments or variants thereof. Further provided are isolated or recombinant polypeptides which comprise an amino acid sequence set forth in SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, and 375-650), or functional fragments or variants thereof.

[0317] The ACR polynucleotides and polypeptides disclosed herein include both the naturally occurring sequences as well as nucleic acid variants. Likewise, the polypeptides and proteins encompass both naturally occurring polypeptides as well as variations and modified

forms thereof. Such polynucleotide and polypeptide variants may continue to possess the desired activity, in which case the mutations that will be made in the DNA encoding the variant will not place the sequence out of reading frame.

[0318] Functional variants of a protein disclosed herein will have at least about 40%, 45%, 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 more sequence identity to the amino acid sequence for the native protein (e.g. the polypeptides provided in SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, and 375-650). Functional variants of a protein disclosed herein may also have at least about 40%, 45%, 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 more sequence identity to the amino acid sequence for the native protein (e.g. the polypeptides provided in SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, and 26) and have a coiled-coil motif. A functional variant of a protein disclosed herein may differ from that protein by as few as 1-15 amino acid residues, as few as 1-10, such as 6-10, as few as 5, as few as 4, 3, 2, or even 1 amino acid residue.

[0319] In some embodiments, the ACR polypeptides include those that contain a coiled coil motif. In some embodiments, coiled coil motifs of the ACR proteins include those polypeptide sequences that contain a repeated pattern of amino acids, **hxxhcxc**, of hydrophobic (h) and charged (c) amino acids, also sometimes referred to as a heptad repeat. In some embodiments, the coiled coil motif includes the polypeptide sequences KQRREYAQEMDRLEKAFENLD and/or ENKLDKIIEKIDKL and those that contain 70%, 75%, 80%, 85%, 90%, 90%, 95%, 96%, 97%, 98%, and 99% sequence identity to SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, and 26. and retain the coiled coil structure.

[0320] The proteins disclosed herein may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants and fragments of polypeptides can be prepared by mutations in the DNA. Methods for mutagenesis and polynucleotide alterations are well known in the art. See, for example, Kunkel (1985) *Proc. Natl. Acad. Sci. USA* 82:488-492; Kunkel *et al.* (1987) *Methods in Enzymol.* 154:367-382; 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 as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found 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 optimal.

[0321] An "active" polypeptide, or fragments thereof, retains a biological activity of the native or naturally-occurring counterpart of the active polypeptide. Biological activity refers to a function mediated by the native or naturally-occurring counterpart of the active polypeptide. For example, binding or protein-protein interaction constitutes a biological activity.

[0322] In some embodiments, certain deletions, insertions, and substitutions of the protein sequences encompassed herein are not expected to produce radical changes in the characteristics of the protein. However, when it is difficult to predict the exact effect of the substitution, deletion, or insertion in advance of doing so, one skilled in the art will appreciate that the effect may be evaluated by screening assays, such as those described herein.

[0323] Variant functional polynucleotides and proteins also encompass sequences and proteins derived from a mutagenic and recombinogenic procedure such as DNA shuffling. With such a procedure, one or more different sequences can be manipulated to create a new polypeptide possessing desired properties. In this manner, libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined *in vitro* or *in vivo*. For example, using this approach, sequence motifs encoding a domain of interest may be shuffled between the polynucleotides disclosed herein and other known polynucleotides to obtain a new gene coding for a protein with an improved property of interest, such as an increased K_m in the case of an enzyme. Strategies for such DNA shuffling are known in the art. See, for example, Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91:10747-10751; Stemmer (1994) *Nature* 370:389-391; Crameri *et al.* (1997) *Nature Biotech.* 15:436-438; Moore *et al.* (1997) *J. Mol. Biol.* 272:336-347; Zhang *et al. (1997) Proc. Natl. Acad. Sci. USA* 94:4504-4509; Crameri *et al.* (1998) *Nature* 391:288-291; and U.S. Patent Nos. 5,605,793 and 5,837,458.

[0324] Fragments and variants of the disclosed polynucleotides and proteins encoded thereby are also provided. By "fragment" is intended a portion of the polynucleotide or a portion of the amino acid sequence and hence protein encoded thereby. Fragments of a polynucleotide

may encode protein fragments that retain the biological activity of the native protein, or fragments of a polynucleotide, may retain the biological activity of the full size polynucleotide; these fragments are referred to herein as "functional fragments". The terms "functional fragment ", "active fragment", "fragment that is functionally equivalent" and "functionally equivalent fragment" are used interchangeably herein.

[0325] A functional fragment of a polynucleotide that encodes a biologically active portion of an ACR polypeptide will encode at least 15, 25, 30, 50, 100, or 125 contiguous amino acids, or up to the total number of amino acids present in a full-length ACR polypeptide (e.g. the polypeptides provided in SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, and 375-650). Such functional fragments of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, and 26 may optionally include a coiled-coil motif.

[0326] Functional fragments of ACR proteins of the present disclosure include fragments comprising 50-130, 60-120, 70-110, 80-100 amino acids of an ACR protein and retain activity. Functional fragments of ACR proteins of the present disclosure can also include fragments comprising 50-130, 60-120, 70-110, 80-100 amino acids of an ACR protein and have a coiled-coil motif.

[0327] A biologically active portion of a polypeptide can be prepared by isolating a portion of one of the polynucleotides disclosed herein, expressing the encoded portion of the protein (e.g., by recombinant expression *in vitro*), and assessing the activity of the encoded portion of the polypeptide. Polynucleotides that are functional fragments of a polynucleotide encoding an ACR protein comprise at least 50, 75, 100, 150, 200, 250, 300, 350, or 400 nucleotides, or up to the number of nucleotides present in a full-length polynucleotide disclosed herein.

[0328]

[0329]

[0330]

Recombinant Constructs

[0331] The ACR compositions provided herein, as well as any of the CRISPR-Cas compositions, may be provided as part of a recombinant construct. The recombinant construct may be part of an expression cassette for use in transforming a heterologous host cell with said

compositions. 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).

[0332] A recombinant DNA construct may comprise regulatory sequences and coding sequences that are derived from different sources. 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.

[0333] In one aspect, the recombinant DNA construct includes heterologous 5' and 3' regulatory sequences operably linked to an ACR polynucleotide as disclosed herein. These regulatory sequences include but are not limited to a transcriptional and translational initiation region (i.e., a promoter), a nuclear localization signal, and a transcriptional and translational termination region (i.e., termination region) functional in the host cell (such as eukaryotic cell).

[0334] In one aspect, the recombinant DNA construct comprises a DNA encoding an ACR protein described herein, wherein the ACR protein is operably linked to or comprises a heterologous regulatory element such as a nuclear localization sequence (NLS).

[0335] In some embodiments, the ACR vectors can be combined with expression cassettes for the expression of one more components of an engineered CRISPR-Cas complex. In one example, one or more constructs are provided that comprise an expression cassette having a promoter functional in a eukaryotic cell operably linked to a polynucleotide encoding an ACR protein as disclosed herein, a second cassette having a promoter functional in a eukaryotic cell

operably linked to a single-guide sequence and a third cassette comprising a promoter functional in a eukaryotic cell operably linked to a Cas9 protein, where the guide and the Cas9 are capable of forming a complex that can modify a target DNA molecule. The cassettes may be provided on a single recombinant construct or on multiple recombinant constructs, which can be used for introduction into host cells either simultaneously or sequentially.

Expression Cassettes

[0336] The ACR polynucleotides disclosed herein can be provided in an expression cassette (also referred to as DNA construct) for expression of the ACR polypeptides in a host cell. The cassette can include 5' and 3' regulatory sequences operably linked to a polynucleotide as disclosed herein. "Operably linked" is intended to mean a functional linkage between two or more elements. For example, an operable linkage between a polynucleotide of interest and a regulatory sequence (e.g., a promoter) is a functional link that allows for expression of the polynucleotide of interest. Operably linked elements may be contiguous or non-contiguous. When used to refer to the joining of two protein coding regions, by operably linked is intended that the coding regions are in the same reading frame.

[0337] Where appropriate, the ACR polynucleotides may be optimized for increased expression in the transformed or targeted host cell. For example, the polynucleotides can be synthesized or altered to use mammalian-preferred or 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. 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.

[0338] The expression cassettes disclosed herein may include in the 5'-3' direction of transcription, a transcriptional and translational initiation region (i.e., a promoter), an ACR polynucleotide, and a transcriptional and translational termination region (i.e., termination region) functional in the host cell (e.g., a eukaryotic cell). Expression cassettes are also provided with a plurality of restriction sites and/or recombination sites for insertion of the polynucleotide to be under the transcriptional regulation of the regulatory regions described elsewhere herein. The regulatory regions (i.e., promoters, transcriptional regulatory regions, and translational termination regions) and/or the polynucleotide of interest may be native/analogous to the host cell or to each other. Alternatively, the regulatory regions and/or the polynucleotide of interest

may be heterologous to the host cell or to each other. As used herein, "heterologous" in reference to a polynucleotide or polypeptide sequence is a sequence that originates from a 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. As used herein, unless otherwise specified, a chimeric polynucleotide comprises a coding sequence operably linked to a transcription initiation region that is heterologous to the coding sequence.

[0339] In some embodiments, a nucleotide sequence encoding an ACR protein is operably linked to a control element, e.g., a transcriptional control element, such as a promoter. The transcriptional control element may be functional in either a eukaryotic cell, e.g., a plant or mammalian cell; or a prokaryotic cell (e.g., bacterial or archaeal cell). In some embodiments, an ACR nucleotide sequence encoding an ACR protein is operably linked to multiple control elements that allow expression of the nucleotide sequence encoding an ACR protein in both prokaryotic and eukaryotic cells.

Expression Elements

[0340] The recombinant construct, or expression cassette, may further comprise a non-coding regulatory element for use in expressing the ACR and/or CRISPR components in a heterologous cell, particularly a plant cell.

[0341] In one embodiment, expression cassettes are provided that comprise a promoter functional in a eukaryotic cell operably linked to a polynucleotide encoding an ACR protein, variant or fragment thereof as disclosed herein.

[0342] The expression cassettes may comprise a promoter operably linked to an ACR polynucleotide, along with a corresponding termination region. The termination region may be native to the transcriptional initiation region, may be native to the operably linked polynucleotide of interest or to the promoter sequences, may be native to the host cell, or may be derived from another source (i.e., foreign or heterologous). Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also Guerineau *et al.* (1991) *Mol. Gen. Genet.* 262:141-144; Proudfoot

(1991) *Cell* 64:671-674; Sanfacon *et al.* (1991) *Genes Dev.* 5:141-149; Mogen *et al.* (1990) *Plant Cell* 2:1261-1272; Munroe *et al.* (1990) *Gene* 91:151-158; Ballas *et al.* (1989) *Nucleic Acids Res.* 17:7891-7903; and Joshi *et al.* (1987) *Nucleic Acids Res.* 15:9627-9639.

[0343] Non-limiting examples of suitable eukaryotic promoters (promoters functional in a eukaryotic cell) include those from cytomegalovirus (CMV) immediate early, herpes simplex virus (HSV) thymidine kinase, early and late SV40, long terminal repeats (LTRs) from retrovirus, and mouse metallothionein-I. The expression vector may also contain a ribosome binding site for translation initiation and a transcription terminator. The expression vector may also contain one or more nuclear localization sequences (NLS sequences) to direct the ACR protein to the nucleus in a eukaryotic cell. The expression vector may also include appropriate sequences for amplifying expression. The expression vector may also include nucleotide sequences encoding protein tags (e.g., 6x His tag, hemagglutinin tag, green fluorescent protein, etc.) that are fused to the ACR protein, thus resulting in a chimeric polypeptide.

[0344] In embodiments where plant cells are employed, plant promoters will find use in the constructs. 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".

[0345] 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.

[0346] 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.

[0347] Tissue-preferred promoters can be utilized to target enhanced expression within a particular plant tissue. Tissue-preferred promoters include, for example, WO2013/103367 published on 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.

[0348] 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); (WO00/11177; 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, WO00/12733, where seed-preferred promoters from *END1* and *END2* genes are disclosed.

[0349] 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.

[0350] 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, WO93/01294), 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).

[0351] 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.*

[0352] 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.

[0353] Another example of an inducible promoter useful in plant cells, is the ZmCAS1 promoter, described in US patent application, US 2013-0312137A1, published on November 21, 2013, incorporated by reference herein.

[0354] 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.

[0355] The expression cassettes may additionally contain 5' leader sequences. Such leader sequences can act to enhance translation. Translation leaders are known in the art and include: picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5' noncoding region) (Elroy-Stein *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6126-6130); potyvirus leaders, for example, TEV leader (Tobacco Etch Virus) (Gallie *et al.* (1995) *Gene* 165(2):233-238), MDMV leader (Maize Dwarf Mosaic Virus) (Johnson *et al.* (1986) *Virology* 154:9-20), and human immunoglobulin heavy-chain binding protein (BiP) (Macejak *et al.* (1991) *Nature* 353:90-94); untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4) (Jobling *et al.* (1987) *Nature* 325:622-625); tobacco mosaic virus leader (TMV) (Gallie *et al.* (1989) in *Molecular Biology of RNA*, ed. Cech (Liss, New York), pp. 237-256); and maize chlorotic mottle virus leader (MCMV) (Lommel *et al.* (1991) *Virology* 81:382-385). See also, Della-Cioppa *et al.* (1987) *Plant Physiol.* 84:965-968. Other methods known to enhance translation can also be utilized, for example, introns, and the like.

Optimized and Modified Sequences

[0356] Where appropriate, the ACR polynucleotides may be optimized for increased expression in the transformed or targeted host cell. For example, the polynucleotides can be synthesized or altered to use mammalian-preferred or 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. 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.

[0357] Additional sequence modifications are known to enhance gene expression in a cellular host. These include elimination of sequences encoding spurious polyadenylation signals, exon-intron splice site signals, 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 cellular host, as calculated by reference to known genes expressed in the host cell. When possible, the sequence is modified to avoid predicted hairpin secondary mRNA structures.

[0358] In preparing the expression cassette, the various DNA fragments may be manipulated so as to provide for the DNA sequences in the proper orientation and, as appropriate, in the proper reading frame. Toward this end, adapters or linkers may be employed to join the DNA fragments or other manipulations may be involved to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like. For this purpose, *in vitro* mutagenesis, primer repair, restriction, annealing, resubstitutions, e.g., transitions and transversions, may be involved.

Cells

[0359] The polynucleotides, polypeptides or expression cassettes disclosed herein can be introduced into a host cell using any method available.

[0360] Methods for introducing polynucleotides or polypeptides into a cell or organism, include, but are 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. Stable transformation is intended to mean that the nucleotide construct introduced into host cell 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 cell 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.

[0361] An ACR protein can be introduced into a cell by directly introducing the ACR protein itself or an mRNA encoding the ACR protein. The ACR protein can also be introduced into a cell indirectly by introducing a recombinant DNA molecule that encodes the ACR protein. The ACR protein can be introduced into a cell transiently or can be incorporated into the genome of the host cell. Uptake of the ACR protein into the cell can be facilitated with a Cell Penetrating Peptide (CPP). Any promoter capable of expressing the ACR protein in a cell can be used and includes a heat shock /heat inducible promoter operably linked to a nucleotide sequence encoding the ACR protein.

[0362] Direct delivery of any one of the ACR polynucleotides or polypeptides, or CRISPR-Cas complex components can be accompanied by direct delivery (co-delivery) of other mRNAs that can promote the enrichment and/or visualization of cells receiving the components. For example, direct co-delivery of the ACR compositions or CRISPR-Cas complex components 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 PCT/US16/57272 filed October 17, 2016 and PCT/US16/57279, filed October 17, 2016.

[0363] Alternatively, polynucleotides may be introduced into 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.

[0364] 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.

[0365] 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, such as cell-penetrating peptides and nanocarriers. See also US20110035836 Nanocarrier-based plant transfection and transduction, and EP 2821486 A1 Method of introducing nucleic acid into plant cells.

[0366] Other methods of introducing polynucleotides into a prokaryotic and eukaryotic cell or organism or animal or plant part can be used, including transformation methods, and the methods for introducing polynucleotides into tissues, for example in plants from seedlings or mature seeds.

Animals

[0367] The presently disclosed polynucleotides and polypeptides can be introduced into a cell, such as a prokaryotic and eukaryotic cells, such as animal cells, in particular mammalian cells.

[0368] Numerous mammalian cell lines have been utilized for expression of gene products including HEK 293 (Human embryonic kidney) and CHO (Chinese Hamster Ovary). These cell lines can be transfected by standard methods (e.g., using calcium phosphate or polyethyleneimine (PEI), or electroporation). Other typical mammalian cell lines include, but are not limited to: HeLa, U2OS, 549, HT1080, CAD, P19, NIH 3T3, L929, N2a, Human embryonic kidney 293 cells, MCF-7, Y79, SO-Rb50, Hep G2, DUKX-X11, J558L, and Baby hamster kidney (BHK) cells.

[0369] The terms "therapeutic composition," "pharmaceutical composition," "therapeutic preparation," and "pharmaceutical preparation" are used interchangeably herein and encompass compositions of the present invention suitable for application or administration to a subject, typically a human. In general such compositions are safe, sterile, and preferably free of contaminants that are capable of eliciting undesirable responses in the subject (i.e., the compound(s) comprising the composition are pharmaceutically acceptable). Compositions can be formulated for application or administration to a subject in need thereof by a number of different routes of administration including oral (i.e., administered by mouth or alimentary canal) or parenteral (e.g., buccal, rectal, transdermal, transmucosal, subcutaneous, intravenous, intraperitoneal, intradermal, intratracheal, intrathecal, pulmonary, and the like).

[0370] The term "subject" as used herein refers to any member of the subphylum chordata, including, without limitation, humans and other primates, including non-human primates such as rhesus macaque, chimpanzees and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats and guinea pigs; birds, including domestic, wild and game birds such as chickens, turkeys and other gallinaceous birds, ducks, geese; and the like. The term does not denote a particular age. Thus, adult, young, and newborn individuals are intended to be covered.

Plants

[0371] The presently disclosed polynucleotides and polypeptides can be introduced into a cell, such as a prokaryotic and eukaryotic cells.

[0372] Numerous plant cells also find use with the compositions and methods provided herein. Plants are further provided comprising an expression cassette comprising a polynucleotide disclosed herein operably linked to a promoter that is active in the plant.

[0373] As used herein, the term plant includes plant cells, plant protoplasts, plant cell tissue cultures from which a plant 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, grain and the like. As used herein, by "grain" is intended 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 genomic modifications of the regenerated plant such as those resulting from transformation or genome editing.

[0374] Any plant or plant part can be used, including monocot and dicot plants or plant parts.

[0375] 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, *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.

[0376] The term “dicotyledonous “or “dicot” refers to the subclass of angiosperm plants also knows as “dicotyledoneae” and includes reference to whole plants, plant organs (e.g., leaves, stems, roots, etc.), seeds, plant cells, and progeny of the same. Examples of dicot plants that can be used include, but are not limited to, soybean (*Glycine max*), *Brassica* species (Canola) (Brassica *napus, B. campestris, Brassica rapa, Brassica. juncea), alfalfa (Medicago sativa),*), 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*.

[0377] Plant 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.

[0378] Vegetables 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.

[0379] Conifers include, for example, 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*).

[0380] The term "plant" includes whole plants, plant organs, plant tissues, seeds, plant cells, seeds and progeny of the same. Plant cells include, without limitation, cells from seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen and microspores. Plant parts include differentiated and undifferentiated tissues including, but not limited to roots, stems, shoots, leaves, pollens, seeds, tumor tissue and various forms of cells and culture (e.g., single cells, protoplasts, embryos, and callus tissue). The plant tissue may be in plant or in a plant organ, tissue or cell culture. The term "plant organ" refers to plant tissue or a group of tissues that constitute a morphologically and functionally distinct part of a plant. 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. "Progeny" comprises any subsequent generation of a plant.

[0381] 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 invention, provided that these parts comprise the introduced polynucleotides.

[0382] A transgenic plant includes, for example, a plant which comprises within its genome a heterologous polynucleotide introduced by a transformation step. The heterologous polynucleotide can be stably integrated within the genome such that the polynucleotide is passed on to successive generations. The heterologous polynucleotide may be integrated into the genome alone or as part of a recombinant DNA construct. A transgenic plant can also comprise more than one heterologous polynucleotide within its genome. Each heterologous polynucleotide may confer a different trait to the transgenic plant. Transgenic can include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic. The alterations of the genome (chromosomal or extra-chromosomal) by conventional plant breeding methods, by the genome editing procedure described herein that does not result in an insertion of a foreign polynucleotide, or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation are not intended to be regarded as transgenic.

[0383] In certain embodiments of the disclosure, a fertile plant is a plant that produces viable male and female gametes and is self-fertile.

ACR Protein Activity

[0384] The ACR polypeptides, variants and fragments described herein can be expressed and/or purified and their biological activity can be confirmed by any method, including those

methods disclosed herein. For example, the biological activity of ACR polypeptides, variants, and fragments thereof can assayed by co-expressing an ACR polynucleotide expressing an ACR polypeptide, variant or fragment thereof in a bacterial cell containing and expressing a CRISPR-Cas9 of a *Streptococcus thermophilus* or *Streptococcus pyogenes* CRISPR system targeting a target sequence of a virulent phage strain and assaying for a reduction in viral titre between bacteria expressing the ACR polypeptide, variant or fragment compared to bacteria lacking the ACR polypeptide, variant or fragment.

Analysis of Cas inhibitory activity of ACR proteins

[0385] In one aspect, the biological activity of the ACR protein of the present disclosure (including the polypeptides encoded by the polynucleotides of the present disclosure), and functional fragment and variants thereof, is an ability to inhibit the cleavage activity of a Cas protein, for example, a Type II Cas9 protein. Methods to determine inhibitory activity by an ACR protein are disclosed herein.

[0386] The disclosure thus provides methods for identifying anti-CRISPR proteins, where the method comprises obtaining a bacterial host cell comprising a recombinant construct capable of expressing a Type II-A CRISPR system having a targeting sequence (also referred to as spacer sequence) capable of targeting a genomic target sequence in a virulent phage, then introducing a construct comprising a promoter functional in the bacterial host cell operably linked to a polynucleotide encoding a polypeptide to be assayed for anti-CRISPR activity, challenging the bacterial host with the virulent phage, and identifying one or more bacterial colonies having a phage titre substantially similar to a bacterial cell lacking the recombinant construct encoding the Type II-A CRISPR system having the targeting sequence capable of targeting a genomic target sequence in the virulent phage challenged with the virulent phage.

[0387] In some embodiments, the anti-CRISPR activity assayed is the ability of a polypeptide to substantially restore the phage titre levels in a bacterial culture having a Type IIA CRISPR system challenged with a virulent phage. In some embodiments, the anti-CRISPR activity assayed results in a bacterial culture having a given Type II CRISPR system a substantially similar susceptibility to a given phage in the presence of the ACR protein as that of the same bacterial strain lacking the Type II CRISPR system being challenged with the same phage.

[0388] Modification of CRISPR system activity and/or genome modification activity by CRISPR systems, such as but not limiting to Type II-A CRISPR-Cas9 complexes, can also be measured as disclosed in described in Rauch et al., 2017, cell 168:150-158.

Methods of Use for ACR proteins

[0389] The compositions and methods provided herein find use in a wide variety of host cells, for example but not limited to those embodiments described herein. As used herein, a "host cell," refers to an in vivo or in vitro eukaryotic cell, a prokaryotic cell (e.g., bacterial or archaeal cell), or a cell from a multicellular organism (e.g., a cell line) cultured as a unicellular entity, which eukaryotic or prokaryotic cells can be, or have been, used as recipients for a nucleic acid, and include the progeny of the original cell which has been transformed by the nucleic acid. A "recombinant host cell" (also referred to as a "genetically modified host cell") is a host cell into which has been introduced a heterologous nucleic acid, e.g., an expression vector. For example, a subject bacterial host cell is a genetically modified bacterial host cell by virtue of introduction into a suitable bacterial host cell of an exogenous nucleic acid (e.g., a plasmid or recombinant expression vector) and a subject eukaryotic host cell is a genetically modified eukaryotic host cell (e.g., a mammalian germ cell or plant cell), by virtue of introduction into a suitable eukaryotic host cell of an exogenous nucleic acid.

[0390] In some embodiments, 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 yeast 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. For example, where the cell is a human cell, the human cell can be either in tissue culture or in vivo.

[0391] The methods provided herein can be used with any CRISPR-Cas system. In one embodiment, the methods and compositions provided herein can be used in combination with CRISPR-Cas systems (e.g. engineered CRISPR-Cas complexes derived from bacterial CRISPR systems) belonging to the Type II CRISPR-Cas systems. Such systems include engineered Type II-A CRISPR-Cas9 complexes.

[0392] In one embodiment, methods are provided for the immunization of eukaryotic cells against CRISPR-Cas9-mediated DNA modification, for example, to reduce or prevent the

cleavage of DNA in a eukaryotic cell by a Cas9 protein complex. Such methods include introducing an ACR polypeptide (or a polynucleotide encoding an ACR polypeptide) into a cell containing a CRISPR-Cas9 complex capable of directing the cleavage of a target DNA in the cell. Such methods can be used in prokaryotic cells. The ACR polypeptide can be introduced simultaneously with the engineered CRISPR-Cas9 complex, or components of the CRISPR-Cas9 complex or sequentially to the CRISPR-Cas9 complex or components thereof. Where the introduction is sequentially, the ACR polypeptide can be introduced prior to the CRISPR-Cas9 complex or after the CRISPR-Cas9 complex. In other embodiments, the ACR polypeptide can be introduced via an expression cassette that provides for an inducible expression of the ACR or a temporal expression of the ACR polypeptide.

[0393] Where activity of an engineered CRISPR-Cas complex is reduced, the reduction in activity can be compared to the activity of the engineered CRISPR-Cas complex in the absence of the anti-CRISPR protein. The reduction in activity can be any measurably amount of reduction when compared to the activity of the engineered CRISPR-Cas complex in the absence of the ACR protein, and includes a reduction of about 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or more in activity. The activity measured can be assayed using a viral titre assay in a bacterial host as described herein, or can be measured as the cleavage activity of the engineered CRISPR-Cas complex itself.

[0394] In some embodiments, the methods include methods for reducing the activity of CRISPR-Cas complexes in a host cell. In some embodiments, such methods include introducing into a host cell a recombinant construct comprising a promoter operably linked to a polynucleotide encoding an ACR protein, variant, or fragment thereof, where the host cell also comprises a CRISPR-Cas complex capable of modifying a target DNA molecule.

[0395] Also provided, are methods for providing an inducible expression of an ACR polypeptide in a cell. Such methods include introducing into a cell an expression cassette comprising an ACR polynucleotide encoding an ACR polypeptide under the operable linkage of an inducible promoter. Examples of inducible promoters for use in methods where an inducible expression of the ACR polypeptide is desired include, but are not limited toT7 RNA polymerase promoter, T3 RNA polymerase promoter, Isopropyl-beta-D-thiogalactopyranoside (IPTG)-regulated promoter, lactose induced promoter, heat shock promoter, Tetracycline-regulated promoter (e.g., Tet-ON, Tet-OFF, etc.), Steroid-regulated promoter, Metal-regulated promoter,

estrogen receptor-regulated promoter, etc. Inducible promoters can therefore be regulated by molecules including, but not limited to, doxycycline; RNA polymerase, e.g., T7 RNA polymerase; an estrogen receptor; an estrogen receptor fusion; and the like.

[0396] Also provided are methods for controlling the cleavage (e.g. single or double-stranded cleavage) of a target DNA by a Type-II CRISPR complex in a eukaryotic cell. Such methods involve the expression or introduction of an ACR polypeptide disclosed herein into a cell containing or expressing a Type-II CRISPR complex capable of cleaving a target DNA molecule. In some embodiments, the methods involve the inducible expression of an ACR polypeptide disclosed herein to allow for the control of the timing of the expression of the ACR polypeptide.

[0397] Also provided are methods for reducing off-target DNA cleavage by a Type-II CRISPR complex in a eukaryotic cell. Such methods involve the expression or introduction of an ACR polypeptide disclosed herein into a cell containing or expressing a Type-II CRISPR complex capable of cleaving a target DNA molecule.

[0398] Such off-target DNA cleavage may be reduced by 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or more compared to the cleavage by the Type-II CRISPR complex in a cell in the absence of the ACR polypeptide.

[0399] Also provided are methods for immunizing a population of cells in which genome editing by a CRISPR-Cas9 complex is desired, but complete penetration is not desired. The methods general comprise expressing or introducing an ACR polypeptide disclosed herein into a population of cells containing a CRISPR-Cas9 complex that is capable of cleaving a target DNA molecule. Such methods may also further involve the identification or selection of a cell having a modification of the target DNA molecule.

[0400] Also provided are methods for protecting cells, e.g. eukaryotic cells, from DNA damage from the activity of an engineered CRISPR-Cas9 complex. Such methods include introducing into a cell a recombinant construct for the expression of an ACR polypeptide as provided herein.

[0401] Also provided are methods for modulating the activity of a Cas endonuclease via the usage of an ACR, by controlling the expression or activity of the Cas endonuclease or the ACR protein during one more cell cycles. In some aspects, the cell cycle is selected from a meiotic phase. In some aspects, the cell cycle is selected from the mitotic phase.

[0402] Also provided are methods for increasing the frequency of homologous recombination during genome editing, and/or reducing the frequency of non-homologous end joining during genome editing.

[0403] The ACR polynucleotides, polypeptides, and methods disclosed herein find use in combination with a wide variety of CRISPR complexes, for example, to inhibit the activity of CRISPR complexes against target DNA molecules. The CRISPR complexes of particular interest include those from Type II CRISPR systems, including those derived from Type II-A CRISPR systems. In some embodiments, the Type II-A CRISPR complexes are those derived from *Streptococcus thermophilus*, *Streptococcus pyogenes*, and *S. aureus*. In other embodiments, the Type II-A CRISPR complexes are those derived from *Streptococcus thermophilus*. In other embodiments, the Type II-A CRISPR complexes are those derived from *Streptococcus thermophilus,* CRISPR1 locus.

[0404] In *Streptococcus thermophilus*, although CRISPR1 and CRISPR3 belong to class 2 type II-A systems, they are different in terms of sequence including Cas9 sequence. For the distinction of CRISPR1 and CRISPR3, reference is made herein to the publication of Chylinski et al. 2014, where the CRISPR1-Cas system is represented by the Cas9 sequence of LMD-9 116628213, and the CRISPR3-Cas system is represented by the Cas9 sequence of LMD-9 116627542.

[0405] In some embodiments, an engineered CRISPR-Cas endonuclease (e.g. an engineered Type II-A CRISPR-Cas9) can (or is capable of) recognize, bind to a DNA target sequence and introduce a single strand (nick) or double-strand break. Once a single or double-strand break is induced in the DNA, the cell's DNA repair mechanism is activated to repair the break. Error-prone DNA repair mechanisms can produce mutations at double-strand break sites. The most common repair mechanism to bring the broken ends together is the nonhomologous end-joining (NHEJ) 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 (such as chromosomal translocations) are possible (Siebert and Puchta, 2002, Plant Cell 14:1121-31; Pacher et al., 2007, Genetics 175:21-9).

[0406] While the invention has been particularly shown and described with reference to a preferred embodiment and various alternate embodiments, it will be understood by persons skilled in the relevant art that various changes in form and details can be made therein without

departing from the spirit and scope of the invention. For instance, while the particular examples below may illustrate the methods and embodiments described herein using a specific plant, the principles in these examples may be applied to any plant. Therefore, it will be appreciated that the scope of this invention is encompassed by the embodiments of the inventions recited herein and in the specification rather than the specific examples that are exemplified below. All cited patents and publications referred to in this application are herein incorporated by reference in their entirety, for all purposes, to the same extent as if each were individually and specifically incorporated by reference.

EXAMPLES

[0407] The following are examples of specific embodiments of some aspects of the invention. The examples are offered for illustrative purposes only, and are not intended to limit the scope of the invention 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.

[0408] The meaning of abbreviations is as follows: "sec" means second(s), "min" means minute(s), "h" means hour(s), "d" means day(s), "µL" or "uL" or "ul" 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).

[0409] Anti-CRISPR (ACR) proteins may be identified, characterized, and utilized according to a number of techniques, some of which are described herein.

Example 1

[0410] Genomic sequences are obtained for a phage that displays virulence against a bacterium comprising a CRISPR-Cas system, wherein the CRISPR-Cas system comprises a targeting sequence that is substantially complementary to a sequence in the phage genome. The sequences are analyzed, and compared to the polynucleotide sequence of at least one known anti-CRISPR protein. In some aspects, at least one polynucleotide of the phage genome shares at least 70% sequence identity with at least 100 bases of a sequence selected from the group consisting of: SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71,73, 75, 77, 79, 81, 83, 85, and 99-374.

Example 2

[0411] A first bacterial host cell comprising a recombinant construct having a Type II-A CRISPR system having a targeting sequence capable of targeting a genomic target sequence in a first virulent phage is obtained. The first bacterial host is challenged with the virulent phage. A second bacterial host cell, preferably of the same strain and genetic composition (isoline) as the first bacterial host cell, comprising a recombinant construct having a Type II-A CRISPR system having a targeting sequence capable of targeting a genomic target sequence in a second virulent

phage is obtained. The second bacterial host is challenged with the second virulent phage. One or more bacterial colonies of the first bacterial host cell is/are identified having a phage titre substantially similar to an otherwise isoline bacterial cell lacking the recombinant construct encoding the CRISPR system having the targeting sequence capable of targeting a genomic target sequence in the first virulent phage challenged with the first virulent phage. One or more bacterial colonies of the second bacterial host cell are identified having a phage titre substantially different than a bacterial cell lacking the recombinant construct encoding the CRISPR system having the targeting sequence capable of targeting a genomic target sequence in the second virulent phage challenged with the second virulent phage. The genomes of the first and second virulent phages are sequenced. One ore more gene(s) is/are present in the first virulent phage but not the second virulent phage. A third bacterial host cell is obtained, preferably of the same strain and genetic composition (isoline) as the first bacterial host cell, comprising a recombinant construct having a CRISPR system having a targeting sequence capable of targeting a genomic target sequence in the first virulent phage. A construct is introduced into the third bacterial host cell, wherein the construct comprises a promoter functional in the third bacterial host cell operably linked to a polynucleotide identical to the gene identified as present in the first virulent phage but not the second virulent phage. The third bacterial host is challenged with the first virulent phage. One or more bacterial colonies of the third bacterial host cell is/are identified, having a phage titre substantially similar to a bacterial cell lacking the recombinant construct encoding the CRISPR system having the targeting sequence capable of targeting a genomic target sequence in the first virulent phage challenged with the first virulent phage.

Example 3

[0412] A bacterial host cell comprising a recombinant construct having a CRISPR system having a targeting sequence capable of targeting a genomic target sequence in a virulent phage was obtained. A construct comprising a promoter functional in the bacterial host cell operably linked to a polynucleotide encoding a polypeptide to be assayed for anti-CRISPR activity was introduced into the bacterial host cell. The bacterial host was challenged with the virulent phage. One or more bacterial colonies were identified, that displayed a phage titre substantially similar to a bacterial cell lacking the recombinant construct encoding the CRISPR system having the targeting sequence capable of targeting a genomic target sequence in the virulent phage challenged with the virulent phage.

Example 4

[0413] As *Streptococcus thermophilus* is a model for the study of CRISPR adaptation, a detailed step-by-step protocol for many of the methods used here is available elsewhere.

Strain Culturing

[0414] *S. thermophilus* cultures were grown in M17 medium (Oxoid, Ontario, Canada) supplemented with 0.5% w/v lactose (LM17). Chloramphenicol, when necessary, was added at 5 ug/ml. When used to generate an overnight culture for use the following day, cultures were grown at 37 °C without shaking. In all other cases, they were grown at 42 °C without shaking. If phages were to be added, the media was further supplemented with 10 mM $CaCl_2$.

[0415] *Lactococcus lactis* cultures were grown in M17 medium (Oxoid, Ontario, Canada) supplemented with 0.5% w/v glucose monohydrate (GM17). Chloramphenicol or erythromycin, when necessary, were added at 5 ug/ml. Cultures were grown at 30 °C without shaking, except when the activity of an *Sp*Cas9-containing construct was assayed, in which case incubations took place at 33 °C. If phages were to be added, the media was further supplemented with 10 mM $CaCl_2$.

[0416] *Escherichia coli* cultures were grown in LB medium. Chloramphenicol, when necessary, was provided at 20 ug/ml. Cultures were grown at 37 °C with shaking.

Phage amplification

[0417] A scraping from a phage lysate preserved at -80 °C with 15% glycerol was co-inoculated with its host strain, in media supplemented with 10 mM $CaCl_2$, and grown until complete lysis was observed. This first amplification lysate was then filtered through a 0.45 um PES filter, and 100 ul used to inoculate its host strain grown to an OD_{600} of 0.1 in media supplemented with 10 mM $CaCl_2$. This second amplification lysate was also filtered through a 0.45 um PES filter, then stored at 4°C.

Phage Titering

[0418] As depicted in FIG. 1, phages were serially diluted in phage buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 8 mM $MgSO_4$). Three ml of molten 0.75 % agar medium at 55 °C, supplemented with 10 mM $CaCl_2$, was inoculated with 300 ul of an overnight culture of the host strain, then rapidly poured over a pre-set plate of the same medium with 1 % agar. The plate was allowed to set and dry. Phage dilutions, 3 ul from each, were spotted onto the dry overlay and

allowed to dry for 20 min. The plates were then incubated overnight, and plaques counted at the lowest dilution at which they were visible.

[0419] As depicted in FIG. 2, phages were serially diluted in phage buffer. Three ml of molten 0.75 % agar medium at 55 °C, supplemented with 10 mM $CaCl_2$, is co-inoculated with 300 ul of an OD_{600} 0.6 culture of the host strain and 100 ul of diluted phage. The plates were then incubated overnight, and plaques counted from plates with between 30-300 plaques.

Immunizing Assays for BIM (Bacteriophage Insensitive Mutants)

[0420] Phages were diluted in phage buffer in order to obtain a final multiplicity of infection (MOI) of 0.1 plaque forming units per colony forming units (pfu/cfu). Three ml of molten 0.75% agar medium at 55 °C, supplemented with 10 mM $CaCl_2$ was co-inoculated with 300 ul of a culture at an OD_{600} of 0.6 (~1.2 x 10^8 cfu/ml) of the host strain and 100 ul of the appropriate phage dilution. The plates were then incubated overnight, and surviving colonies counted.

Characterization of Surviving Colonies

[0421] Random surviving bacterial colonies were screened by PCR for acquisition of new spacers at the CRISPR1 & CRISPR3 loci (*S. thermophilus* strain DGCC7710) or CRISPR1 locus (*S. thermophilus* strain DGCC7854). An increase in the size of the PCR product relative to the wild type was indicative of CRISPR immunization. The resulting PCR products were sequenced to confirm the identity of the newly acquired spacer. For assays in Figure 2, presence of the insert in pNZAcr was confirmed by sequencing in cells that had acquired spacers.

Plasmid Programming

[0422] A plasmid was designed to contain a protospacer (CRISPR-acquirable sequence) targeting the five phages used in the challenges. Two oligos consisting of a conserved protospacer in the gene encoding the tape measure protein, as well as overhangs suited for cloning, were annealed together by mixing them in equal parts, heating them to 98°C, then cooling them slowly to 50°C. This annealed construct was then ligated directly into an EcoRI/XhoI double-digested pNZ123, transformed into commercial NEB5α, and selected for with chloramphenicol. The constructed plasmid was then isolated using Qiaprep Spin Miniprep kit (Qiagen, Ontario, Canada) according to the manufacturer's recommendations. *S. thermophilus* DGCC7854 was transformed with this plasmid, pNZ5phage, then grown in the absence of selection for 7 generations and subjected to an immunizing assay (see above) with

virulent phage D5842. The surviving colonies had naturally acquired the desired spacer from the plasmid, immunizing them to the phages. The spacer sequence was confirmed as described in "characterization of surviving colonies" above.

Phage Genome Sequencing & Annotation

[0423] DNA from the phage D4276 was purified using a PureLink Viral RNA/DNA kit (Invitrogen, MA, USA). The purified DNA was sequenced on a MiSeq system using a MiSeq reagent kit v2 after preparation using the Nextera XT DNA library preparation kit (Illumina, British Columbia, Canada). The resulting reads assembled using Ray version 2.2.0 (32). The genome was annotated using NCBI ORF finder and GeneMark.hmm prokaryotic, and those annotations then manually curated based on comparisons to related phages.

Phage Gene Cloning and pNZAcr Construction

[0424] Primers were designed to systematically clone all of phage D4276 into pNZ123 oriented so as to drive transcription from the promoter upstream of the chloramphenicol resistance gene, *cat*. Initially, inserts were designed to contain several genes, but if cloning failed the inserts were redesigned as smaller, single-gene constructs. The gene of greatest interest, D4276_028, exemplifies this cloning technique. Primers were designed to amplify the gene and append 30 nt extensions overlapping the pNZ123 MCS (SEQ ID NO:87 and SEQ ID NO:88). The amplified gene was then cloned by Gibson reaction into XhoI digested pNZ123. The resulting plasmid, pNZAcr, was transformed into commercial NEB5a, isolated using a Qiaprep Spin Miniprep kit, and then transformed into the relevant *S. thermophilus* and *L. lactis* strains. The sequence of the insert was confirmed by sequencing using primers (SEQ ID NO:89 and SEQ ID NO:90).

Plasmid Loss Assays

[0425] Cultures carrying pNZAcr were serially grown in the absence of selection, inoculating fresh 10 ml of LM17 broth media with 100 ul of a culture grown to saturation. This was repeated 5 times. Dilutions of the resulting culture were spread upon plates in order to obtain isolated colonies, and 120 such colonies were then patch-plated on LM17 with and without chloramphenicol. Colonies, which grew on LM17 (all 120) but failed to grow on LM17 Cm (two), were screened by PCR to confirm plasmid loss using pNZinsF and pNZinsR, and their CRISPR1 locus was amplified to confirm the presence of the immunizing spacer. Colonies were

then used to titer the phages D4276 and D5842, and confirm that they had regained resistance to the phages from losing the plasmid.

pL2Cas9-44 construction

[0426] pL2Cas9 (Lemay et al. 2017) is a derivative of the lactococcal vector pTRKL2 (O' Sullivan et al. 1993) with the *Sp*Cas9 module of pCas9 (Jiang et al. 2013). A pair of oligos comprising a spacer sequence targeting *orf44* of phage p2 and overhangs for ligation into pL2cas9 were designed (SEQ ID NO:90 and SEQ ID NO:91). They were annealed together by mixing them in equal parts, heating them to 98 °C, then cooling them slowly to 50 °C. This annealed construct was then ligated directly into digested pL2Cas9 and transformed directly into *L. lactis*. The resulting transformants were screened by PCR amplification and sequencing to confirm the presence of the desired spacer, using primers (SEQ ID NO:93 and SEQ ID NO:94).

Efficiency of centers of infection (ECOI)

[0427] Cultures of all four strains depicted in Figure 3B were grown at 33 °C to an OD_{600} of 0.8 (~$1.9*10^8$ cfu/ml), 2 ml spun down and resuspended 1 ml in fresh GM17 media with 10 mM $CaCl_2$. Then, phage p2 was added to an MOI of 0.2, mixed by inversion, and given 5 min to allow adsorption to the cells at 33 °C. The phage-cell mix was then spun down and resuspended in fresh media thrice in order to wash away unbound phages, then serially diluted. 100 ul of the resulting dilutions were then added to 300 ul of indicator strain (MG1363 pNZ123 pL2Cas9), embedded in a soft agar overlay (see phage titering), and incubated at 33 °C overnight.

Results

[0428] *Streptococcus thermophilus* has become a model for acquisition of new CRISPR immunities, shares its genus with the source of *Sp*Cas9, and its active CRISPR-Cas systems are also of type II-A. A set of five virulent phages infecting *S. thermophilus* strain DGCC7854 proved ideal for identifying phages that were less likely to lead to the acquisition of new spacers (phage-derived sequences in the CRISPR array, conferring immunity); while two of the phages readily gave rise to CRISPR-immune colonies, three did not (Figure 1, Top). The dearth of spacer acquisition from these three phages did not necessarily confirm the presence of anti-CRISPRs. Those phages could simply be more sensitive to non-CRISPR forms of resistance, be quicker to take over the host cell, or produce fewer immunogenic defective particles. It was necessary to establish whether the CRISPR-Cas system was impeded during the adaptation ('memorization' of new targets) or interference (cleavage of that target) process. Using plasmid-

programing, a strain targeting a protospacer conserved in all five phage genomes was generated. Then, in plaquing each phage upon this strain, it was observed that four of the five phages suffered a drastic reduction (~ 6 Log) in titer, consistent with CRISPR interference, but one phage, D4276, did not (Figure 1, bottom). These phages were categorized according to these CRISPR-interacting phenotypes; permissive (white) phages D5842 and D5843, impeded adaptation (fractal pattern) phages D1024 and D5891, and restrictive (black) phage D4276 – a candidate to harbor an anti-CRISPR.

[0429] Genes from the restrictive phage D4276 were cloned into a vector where they could be expressed in the immunized strain (Figure 2A). A phage gene encoding an anti-CRISPR protein should inactivate the pre-existing immunity and thereby restore the titer of a sensitive (permissive) phage plated upon the strain. A new anti-CRISPR gene (*acr* gene, described herein as SEQ ID NO:9 encoding an anti-CRISPR protein as described in SEQ ID NO:10) was obtained, which completely restored the immunized strain's sensitivity to the permissive phage D5842 (~6 Log increase), as well as increased sensitivity to the restrictive phage D4276 back to wild-type levels (Figure 2B). We attribute this increase in titer for even the anti-CRISPR-containing phage D4276 to high anti-CRISPR production before phage exposure, which would otherwise be a time-sensitive process whereby production must outrace CRISPR activity. In order to ensure the gain-of-sensitivity phenotypes were due *only* to this anti-CRISPR, we allowed loss of the anti-CRISPR -bearing plasmid and confirmed a return-of-resistance phenotype (data not shown).

[0430] As all five phages infecting *S. thermophilus* DGCC7854 are related *cos*-type phages, we could not rule out that the anti-CRISPR might be dependent upon interaction with partner proteins present in these phages. Furthermore, strain DGCC7854 contains only a single active CRISPR-Cas system (CRISPR1), as opposed to the two systems (CRISPR1 & CRISPR3) commonly active in *S. thermophilus* strains. We ported our anti-CRISPR vector over to the well-characterized model strain, *S. thermophilus* DGCC7710, which is sensitive to an unrelated virulent *pac*-type phage, 2972 – and for which we have strains immunized at either the CRISPR1 or CRISPR3 locus (Figure 2C). In this system, the anti-CRISPR activity was maintained, completely restoring phage sensitivity to a CRISPR1-immunized strain, and partially restoring sensitivity for a CRISPR3-immunized strain. When we attempted immunizing assays on DGCC7710 bearing the *acr* gene, the number of surviving colonies fell sharply (Figure 2D).

Moreover, the nature of those survivors changed drastically. The number of CRISPR3-immune colonies dropped, a number of previously undetectable non-CRISPR mutants were observed, and whereas CRISPR1 immunizations normally compose upwards of 90% of surviving colonies, only a single colony with a CRISPR1 spacer acquisition was recovered (Figure 2D). Of note, the CRISPR1 spacer in question targeted the plasmid (still present and carrying an intact *acr* gene), rather than the phage genome. This indicates that the ACR protein likely impedes Cas9-mediated cleavage, but not Cas9's role in spacer acquisition.

[0431] The ACR protein (SEQ ID NO:10) is 140 amino acids long and is predicted to contain a distinctive coiled-coil motif, which might act in a nucleic acid binding role, similar to HTH and AP2 motifs associated with other anti-CRISPR proteins. We have found several new anti-CRISPR genes both in phage genome (SEQ ID NO: 1, 3, 5, 11, 13, 15 and 17 encoding respectively anti-CRISPR protein as defined in SEQ ID NO: 2, 4, 6, 12 14, 16 and 18) as well as in the genome of *Streptococcus* strains (SEQ ID NO:5, 7, 19, 21, 23 and 25 encoding respectively anti-CRISPR protein as defined in SEQ ID NO: 6, 8, 20, 22, 24 and 26).

[0432] Finally, despite the fact that the genome-editing tool *Sp*Cas9 (Cas9 from *S. pyogenes*) is more closely related to the Cas9 of the CRISPR3 system of *S. thermophilus*, we were keen to determine whether the ACR protein would have any effectiveness against *Sp*Cas9. We initially attempted to assay the effectiveness of ACR (pNZAcr) on *Sp*Cas9 (pCas9) in *Escherichia coli*, but despite the ability to clone each separately, the two systems were not able to co-exist. Some aspects of the ACR-Cas9 interaction may be pernicious to *E. coli*. Instead, we used a pCas9 derivative adjusted for use in *Lactococcus lactis*, with demonstrated efficacy in the genome-editing of virulent phages (Figure 3A). The pL2Cas9-44 construct, targeting *orf44* of the virulent phage p2, resulted in a 4 Log decrease in measurable phage titer (Figure 3B), which was completely restored by the presence of the *acr* gene. This is the strongest reported anti-CRISPR activity against *Sp*Cas9 to date.

[0433] The 4 Log reduction associated with pL2Cas9-44 was also accompanied by a 'tiny plaque' phenotype that proved difficult to quantify, as they were only observable on some technical replicates. The maximum number of tiny plaques observed is displayed in pale orange (Figure 3B). We characterized the phenotype and genotype of these smaller plaques, establishing that they were CRISPR-bypassing mutants genetically indistinguishable from those in the larger plaques, but arising only after several rounds of replication on the 'leaky' targeting strain.

Notably, however, the expression of the ACR completely rescued both the titer and the tiny-plaque phenotype.

[0434] The ACR protein (SEQ ID NO:10) is the first anti-CRISPR protein with demonstrated activity from a virulent phage, is structurally distinct from previously characterized anti-CRISPRs, and displays the strongest *in vivo* activity against SpCas9 to date.

Example 5

[0435] Strain culturing, phage amplification, phage titering, immunizing assays, characterization of surviving colonies and transformation were done the same way as in Example 1.

Phage Genome Sequencing and Annotation

[0436] DNA from the phage D1811 was purified using a PureLink Viral RNA/DNA kit (Invitrogen, MA, USA). The purified DNA was sequenced on a MiSeq system using a MiSeq reagent kit v2 after preparation using the Nextera XT DNA library preparation kit (Illumina, British Columbia, Canada). The resulting reads assembled using Ray version 2.2.0 (32). The genome was annotated using NCBI ORF finder and GeneMark.hmm prokaryotic, and those annotations then manually curated based on comparisons to related phages.

Phage Gene Cloning and pNZAcr Construction

[0437] Primers were designed to amplify a gene of interest, D1811_026, and append 30 nt extensions overlapping the pNZ123 MCS (SEQ ID NO:95 and SEQ ID NO:96). The amplified gene was then cloned by Gibson reaction into XhoI digested pNZ123. The resulting plasmid, pNZAcr-1811, was transformed into commercial NEB5a, isolated using a Qiaprep Spin Miniprep kit, and then transformed into the relevant *S. thermophilus*. The sequence of the insert was confirmed by sequencing using primers (SEQ ID NO:97 and SEQ ID NO:98).

Plasmid Loss Assays

[0438] Cultures carrying pNZAcr were serially grown in the absence of selection, inoculating fresh 10 ml of LM17 broth media with 100 ul of a culture grown to saturation. This was repeated 14 times. Dilutions of the resulting culture were spread upon plates in order to obtain isolated colonies, and 160 such colonies were then patch-plated on LM17 with and without chloramphenicol. Colonies, which grew on LM17 (all 120) but failed to grow on LM17 Cm (two), were screened by PCR to confirm plasmid loss using pNZinsF and pNZinsR, and their CRISPR1 locus was amplified to confirm the presence of the immunizing spacer. Colonies were

then used to titer the phages D1811 and D5842, and confirm that they had regained resistance to the phages from losing the plasmid.

Results

[0439] In plaquing an additional phage (D1811) upon the DGCC7854 strain, it was observed that this phage suffers a much smaller reduction in titer than 4 other related phages (Figure 1, bottom). This phage was categorized as restrictive (black) phage.

[0440] We found a second new anti-CRISPR gene (*acr2* gene, defined herein as SEQ ID NO:27 encoding an anti-CRISPR protein as defined in SEQ ID NO:28), which completely restored the immunized strain's sensitivity to the permissive phage D5842 (~5 Log increase), as well as increased sensitivity to the restrictive phage D1811 back to wild-type levels (Figure 4A). We attribute this increase in titer for even the anti-CRISPR-containing phage D1811 to high anti-CRISPR production before phage exposure, which would otherwise be a time-sensitive process whereby production must outrace CRISPR activity. In order to ensure the gain-of-sensitivity phenotypes were due *only* to this anti-CRISPR, we allowed loss of the anti-CRISPR -bearing plasmid and confirmed a return-of-resistance phenotype (data not shown).

[0441] Since D1811 is related to the five other phages disclosed in example 1, we could not rule out that the anti-CRISPR might be dependent upon interaction with partner proteins present in these phages. We ported our anti-CRISPR vector over to the well-characterized model strain, *S. thermophilus* DGCC7710, which is sensitive to an unrelated virulent *pac*-type phage, 2972 – and for which we have strains immunized at either the CRISPR1 or CRISPR3 locus (Figure 4B). In this system, the anti-CRISPR activity was maintained, completely restoring phage sensitivity to a CRISPR1-immunized strain.

[0442] The Acr2 protein (SEQ ID NO:28) is 183 amino acids long. We have found several new anti-CRISPR genes both in phage genome (SEQ ID NO: 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79 and 81 encoding respectively anti-CRISPR protein as defined in SEQ ID NO:, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 80 and 82,) as well as in the genome of *Streptococcus* strains (SEQ ID NO: 83 85, and 99-374 encoding respectively anti-CRISPR protein as defined in SEQ ID NO: 84 86, and 375-650).

Example 6

[0443] In this example, methods to reduce "off-target" chromosomal DNA cleavage by the RNA guided endonuclease are described. In some aspects, any Cas endonuclease may be used to generate double-strand breaks. In some aspects, a Type II Cas endonuclease may be used to generate double-strand breaks. In some aspects, a Cas9 endonuclease from any organism may be used to generate double-strand breaks. In some aspects, the Cas9 endonuclease is from *S. pyogenes* or *S. thermophilus*.

[0444] In one example, a Cas endonuclease, for example but not limited to *S. pyogenes* Cas9 (SpCas9), can be directed by guide RNAs (gRNAs) to cleave DNA targets and introduce double-strand breaks (DSBs) at high efficiencies in multiple organisms including plants (Hsu, P.D. *et al.* (2014) *Cell.* 157:1262-1278). The cellular repair of the DSB(s) is then used to introduce genetic modifications. This may include small insertion or deletion (indel) mutations, large deletions (if more than one DNA site is targeted), purposeful edits (for example alteration of a codon in a gene), and insertion of DNA within or near the DNA target sequence. Depending on the nature of the experimental system used, SpCas9 may generate DSBs and chromosomal alterations in other locations in the genome besides those intended (Fu, Y. *et al.* (2013) *Nat. Biotechnol.* 31:822-826). To reduce these potential "off-target" effects, an anti-CRISPR (ACR) can be utilized. The method relies on the recombinant expression of an ACR that inhibits SpCas9 or other Cas9s or any other Cas endonuclease from binding, nicking, or cleaving DNA. In this case, the timing of ACR expression is relevant. If it is expressed before the Cas endonuclease has cleaved the intended on-target sequence, then the intended target site may have no or reduced DSB activity. Alternatively, if it is expressed too late, continued activity of the Cas endonuclease protein may not be affected by ACR expression, resulting in less specific activity. To ensure proper timing, an ACR recombinant gene expression cassette can be designed to be non-functional, then following Cas endonuclease expression and RNA guided cleavage, or after a sufficient time has passed for site-specific cleavage, converted into a functional expression cassette. To restore functionality by cleavage, a method like that described in PCT application publication number WO2017070032 may be utilized. In an embodiment, the translational open-reading frame (ORF) of the ACR protein of interest can be designed to be out of frame (for example but not limited to the deletion of a single base) (Figure 9) where the ORF results in RNA transcript that contains a premature stop codon or encodes a non-sense protein. The

resulting non-functional ORF can be then converted into a functional ORF by targeting the out of frame sequence for Cas endonuclease cleavage and cellular DSB repair (Figure 9). Additionally, the gene encoding the out of frame ACR protein may be combined with other genes (for example but not limited to selective agents and marker genes (Miki, B. (2004) *J Biotechnol.* 107:193-232)) in a polycistron (separated by sequences encoding 'self-cleaving' 2A peptides (Szymczak, A.L. *et al.* (2004) *Nature Biotech.* 22:589-594) (Figure 10). Then, following restoration of the ACR ORF, the other genes in the multicistronic expression cassette can also be converted into a functional state (Figure 10). Thus, permitting simultaneous positive selection for functional ACR expression and Cas9 genome editing. Taken together, this approach results in an auto-regulatory feedback loop (ARFL) which through Cas endonuclease activity results in rapid inactivation of Cas endonuclease by ACR reducing the potential for off-target cleavage.

Example 7

[0445] In this example, methods to enhance the chromosomal DNA repair of Cas-generated double strand breaks (DSBs) with the homologous recombination (HR) DNA repair pathway using an anti-CRISPR protein are described. In some aspects, any Cas endonuclease may be used to generate double-strand breaks. In some aspects, a Type II Cas endonuclease may be used to generate double-strand breaks. In some aspects, a Cas9 endonuclease from any organism may be used to generate double-strand breaks. In some aspects, the Cas9 endonuclease is from *S. pyogenes* or *S. thermophilus*.

[0446] Cellular repair of Cas endonuclease induced DSBs utilizes the non-homologous end-joining (NHEJ) and the HR DNA repair pathways (Hsu, P.D. *et al.* (2014) *Cell.* 157:1262-1278). NHEJ repair may result in the imprecise insertion or deletion (indel) of DNA base pairs (bps) at a chromosomal DNA target site and is useful for disrupting (knocking-out) gene expression. In contrast, HR-mediated repair offers a highly precise method to introduce desired changes into DNA using an exogenously supplied DNA repair template (Capecchi, M.R. (1989) *Science.* 244:1288-1292). The NHEJ pathway is typically the most prevalent DSB repair outcome making the recovery of HR-mediated alterations infrequent (Capecchi, M.R. (1989) *Science.* 244:1288-1292). To increase the frequency of HR repair, an anti-CRISPR (ACR) may be used to time Cas endonuclease cleavage activity with the part of the cell cycle where HR repair occurs, S (Synthesis) and G2 (Gap 2) phases (Heyer, W.D. *et al.* (2010) *Annu. Rev. Genet.* 44:113-139). To accomplish this, the ubiquitin-mediated proteolysis pathway may be leveraged.

By fusing part or all of Cdt, a protein that is targeted for degradation by the SCF^{Skp2} ubiquitination complex in the S and G2 cellular phases (Nishitani, H. *et al.* (2000) *Nature.* 404:625-628), to ACR, its expression can be limited to the G1 (Gap 1) phase. Thus, inactivating Cas endonuclease during G1 when HR repair is inactive and permitting Cas endonuclease re-activation during S and G2 when HR repair machinery is expressed and active (Figure 11). Taken together, ACR may be used as a tool to modulate Cas endonuclease activity in a cell cycle-dependent manner resulting in enhanced HR repair. In some aspects, an ACR may be used to modulate Cas endonuclease activity during meiosis. In some aspects, an ACR may be used to modulate Cas endonuclease activity during mitosis. In some aspects, an ACR may be used to modulate Cas endonuclease activity during S phase or G2 phase.

Example 8

[0447] Methods for controlling the expression of a Cas endonuclease in a plant via spatial regulated expression, temporal regulated expression, or inducible expression of the ACR are contemplated. In some aspects, the ACR, the Cas endonuclease, or both is/are pre-integrated into the genome of at least one plant cell in the plant. In some aspects, ACR, the Cas endonuclease, or both is/are introduced as polynucleotides into at least one cell of the plant, or into a cell from which a whole plant or plant tissue may be derived. In some aspects, ACR, the Cas endonuclease, or both is/are introduced as polypeptides into at least one cell of the plant, or into a cell from which a whole plant or plant tissue may be derived.

[0448] In this example, methods to regulate the binding, nicking, and cleavage activity of a RNA guided CRISPR endonucleases in a tissue specific manner using an anti-CRISPR protein are described. In some aspects, any Cas endonuclease may be used. In some aspects, any Type II Cas endonuclease may be used. In some aspects, a Cas9 endonuclease from any organism may be used.

[0449] MicroRNAs (miRNAs) are small non-coding RNAs that provide a pleotropic cellular mechanism for modulating gene expression and key determinant for cellular differentiation (Baskerville, S. et al. (2005) RNA. 11: 241-247, Lagos-Quintana, M. et al. (2002) Curr. Biol. 12: 735-739, Chen, C.Z. et al. (2004) Science. 303: 83-86 and Lu, J. et al. (2005) Nature. 435: 834-838). They act to regulate gene expression post-transcriptionally by targeting transcribed RNAs for degradation (Lagos-Quintana, M. et al. (2001) Science. 294: 853-858). Additionally, they have been repurposed to regulate transgene expression in a tissue specific

manner by placing their binding site(s) in the 3 prime untranslated region (UTR) of foreign genes (Brown, B. et al. (2006) Nat. Med. 12:585-591). To regulate the binding, nicking, and cleavage activity of a Cas9 protein in a tissue specific manner, a recombinant anti-CRISPR (ACR) encoding gene can be made into a substrate for cellular miRNA(s) regulation. By placing one or more miRNA binding sites in the 3' UTR of the ACR gene, its expression and consequential activity of Cas9 may be modulated as a function of tissue type, developmental stage, or growth condition. Cas9 will be active in the presence of miRNA translational repression and inactive in the absence of miRNA translation repression. As an alternative, miRNA binding sites can be also placed into the 3' UTR of the Cas9 gene to directly regulated its expression in a tissue specific manner. Moreover, miRNAs with different tissue specificities can be incorporated into the 3' UTRs of both the recombinant ACR and Cas9 expression constructs to provide additional layers of regulation.

We claim:

1. A method for modulating the activity of a Cas endonuclease with a target polynucleotide in a cell, comprising providing an anti-CRISPR (ACR) polypeptide to the cell , wherein the anti-CRISPR polypeptide modulates the activity of the Cas endonuclease in the cell.
2. The method of Claim 1, wherein the Cas endonuclease activity selected from the group consisting of: target polynucleotide binding, target polynucleotide nicking, target polynucleotide double-strand-break creation, and target polynucleotide modification.
3. The method of Claim 2, wherein said target polynucleotide modification is selected from the group consisting of: insertion of at least one nucleotide, deletion of at least one nucleotide, substitution of at least one nucleotide, and chemical alteration of at least one nucleotide.
4. The method of Claim 1, wherein the Cas endonuclease lacks the ability to nick or cleave a polynucleotide.
5. The method of Claim 4, wherein the Cas endonuclease is operably linked to a deaminase.
6. The method of Claim 5, further comprising editing at least one base of the target polynucleotide.
7. The method of Claim 1, wherein said activity is decreased as compared to a control cell comprising said Cas endonuclease and guide RNA but not comprising said anti-CRISPR polypeptide.
8. The method of Claim 1, wherein the cell is selected from the group consisting of: plant cell, animal cell, mammalian cell, microbial cell, fungal cell.
9. The method of Claim 1, wherein the Cas endonuclease is a Type II-A Cas endonuclease.
10. The method of Claim 1, wherein the Cas endonuclease is Cas9 or Cpf1.
11. A method of increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a cell, comprising introducing an anti-CRISPR (ACR) polypeptide to the cell, wherein the ACR polypeptide interacts with the Cas endonuclease.
12. The method of Claim 11, wherein the specificity of the Cas endonuclease is improved by at least 5% compared to a control cell lacking said ACR polypeptide.

13. The method of Claim 11, wherein the ACR polypeptide interacts with the Cas endonuclease during a specific cell cycle stage.

14. The method of Claim 9, wherein the cell is selected from the group consisting of: plant cell, animal cell, mammalian cell, microbial cell, fungal cell.

15. The method of Claim 13, wherein the cell cycle stage is during meiosis.

16. The method of Claim 13, wherein the cell cycle stage is during mitosis.

17. The method of Claim 11, wherein the ACR polypeptide modulates the activity of the Cas endonuclease during a specific stage of an organism's development.

18. The method of Claim 17, wherein said stage is selected from the group consisting of: growth, reproductive, vegetative, and senescence.

19. The method of Claim 11, wherein the Cas endonuclease and a guide polynucleotide are provided as polynucleotides.

20. The method of Claim 11, wherein the Cas endonuclease is provided as a polypeptide and the guide polynucleotide is provided as an RNA molecule.

21. The method of Claim 11, wherein the ACR is provided as a polynucleotide encoding a polypeptide.

22. The method of Claim 11, wherein the ACR is provided as a polypeptide.

23. The method of Claim 11, wherein the ACR is provided concurrently with either the Cas endonuclease or the guide polynucleotide.

24. The method of Claim 11, wherein the ACR is provided sequentially in relation to the introduction of the Cas endonuclease or the guide polynucleotide.

25. The method of Claim 11, wherein a polynucleotide encoding the ACR and/or the polynucleotide encoding the Cas endonuclease is/are pre-integrated into the genome of the cell or organism.

26. The method of Claim 11, wherein the relative ratio of the concentration of Cas endonuclease to ACR is between 100:1 and 1:100.

27. The method of Claim 11, wherein the expression or activity of the ACR is inducible.

28. The method of Claim 27, wherein induction is in response to a condition selected from the group consisting of: temperature, presence or absence of an exogenously-applied molecule, activation or inhibition of an endogenous gene, light, dark, cell cycle, organism phase, tissue or cell type, and environmental stress.

29. The method of Claim 11, wherein the ACR protein comprises a coiled-coil motif.

30. The method of Claim 11, wherein the ACR protein comprises a heptad repeat pattern of amino acids in the pattern of "hxxhcxc", wherein h = a hydrophobic amino acid, c = a charged amino acid, and x = any amino acid.

31. The method of Claim 11, wherein the ACR has an amino acid sequence at least 70% identical to a sequence selected from the group consisting of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, and 375-650.

32. The method of Claim 11, wherein the Cas endonuclease is a Type II-A Cas endonuclease.

33. The method of Claim 11, wherein the Cas endonuclease is Cas9 or Cpf1.

34. A method of increasing site-specific homologous recombination frequency of a donor polynucleotide in a cell, comprising introducing to the cell an anti-CRISPR (ACR) polypeptide to increase the homologous recombination of the donor polynucleotide by a polynucleotide-guided Cas endonuclease.

35. The method of Claim 34, wherein the cell is selected from the group consisting of: plant cell, animal cell, mammalian cell, insect cell, human cell, microbial cell, fungal cell.

36. The method of Claim 34, wherein the ACR polypeptide interacts with the Cas endonuclease during a specific cell cycle stage.

37. The method of Claim 36, wherein the cell cycle is during meiosis.

38. The method of Claim 36, wherein the cell cycle is during mitosis.

39. The method of Claim 34, wherein said method is performed during a specific stage of an organism's development.

40. The method of Claim 39, wherein said stage is selected from the group consisting of: growth, reproductive, vegetative, and senescence.

41. The method of Claim 34, wherein the Cas endonuclease and the guide RNA are provided as polynucleotides.

42. The method of Claim 34, wherein the Cas endonuclease is provided as a protein and the guide polynucleotide is provided as an RNA molecule.

43. The method of Claim 34, wherein the ACR is provided as a polynucleotide encoding a polypeptide.

44. The method of Claim 34, wherein the ACR is provided as a polypeptide.

45. The method of Claim 34, wherein the ACR is provided concurrently with either the Cas endonuclease or the guide polynucleotide.

46. The method of Claim 34, wherein the ACR is provided sequentially in relation to the introduction of the Cas endonuclease or the guide polynucleotide.

47. The method of Claim 34, wherein a polynucleotide encoding the ACR and/or the polynucleotide encoding the Cas endonuclease is/are pre-integrated into the genome of the cell or organism.

48. The method of Claim 34, wherein the expression or activity of the ACR is inducible.

49. The method of Claim 48, wherein induction is in response to a condition selected from the group consisting of: temperature, presence or absence of an exogenously-applied molecule, activation or inhibition of an endogenous gene, light, dark, cell cycle, organism phase, tissue or cell type, and environmental stress.

50. The method of Claim 34, wherein the ACR protein comprises a coiled-coil motif.

51. The method of Claim 34, wherein the ACR protein comprises a heptad repeat pattern of amino acids in the pattern of "hxxhcxc", wherein h = a hydrophobic amino acid, c = a charged amino acid, and x = any amino acid.

52. The method of Claim 34, wherein the Cas endonuclease is a Type II-A Cas endonuclease.

53. The method of Claim 34, wherein the Cas endonuclease is Cas9.

54. A cell comprising a Cas endonuclease and an anti-CRISPR (ACR) protein.

55. The cell of Claim 54, wherein said ACR protein is provided by a phage, virus, or recombinant construct.

56. The cell of Claim 54, further comprising a guide polynucleotide.

57. The cell of Claim 54, further comprising a heterologous polynucleotide.

58. The cell of Claim 54, wherein the cell is selected from the group consisting of: plant cell, animal cell, mammalian cell, insect cell, human cell, microbial cell, fungal cell.

59. The plant cell of Claim 58, selected from the group consisting of: maize, rice, sorghum, rye, barley, wheat, millet, oats, sugarcane, turfgrass, switchgrass, soybean, canola, alfalfa, sunflower, cotton, tobacco, peanut, potato, tobacco, *Arabidopsis*, vegetable, and safflower.

60. A cell comprising a recombinant construct comprising a polynucleotide sequence encoding an anti-CRISPR (ACR) protein, operably linked to a heterologous regulatory expression element.

61. The cell of Claim 60, wherein the heterologous regulatory expression element is inducible in response to a condition selected from the group consisting of: temperature, presence or absence of an exogenously-applied molecule, activation or inhibition of an endogenous gene, light, cell cycle, organism phase, tissue or cell type, and environmental stress.

62. The cell of Claim 60, wherein the ACR protein comprises a coiled-coil motif.

63. The cell of Claim 60, wherein the ACR protein comprises a heptad repeat pattern of amino acids in the pattern of "hxxhcxc", wherein h = a hydrophobic amino acid, c = a charged amino acid, and x = any amino acid.

64. The method of Claim 60, wherein the cell is selected from the group consisting of: plant cell, animal cell, mammalian cell, insect cell, human cell, microbial cell, fungal cell.

65. The plant cell of Claim 63, selected from the group consisting of: maize, rice, sorghum, rye, barley, wheat, millet, oats, sugarcane, turfgrass, switchgrass, soybean, canola, alfalfa, sunflower, cotton, tobacco, peanut, potato, tobacco, *Arabidopsis*, vegetable, and safflower.

66. A method for gene activation or gene repression in a cell, comprising introducing an anti-CRISPR (ACR) polypeptide and a Cas endonuclease into the cell comprising the gene to be activated or repressed.

FIG. 1

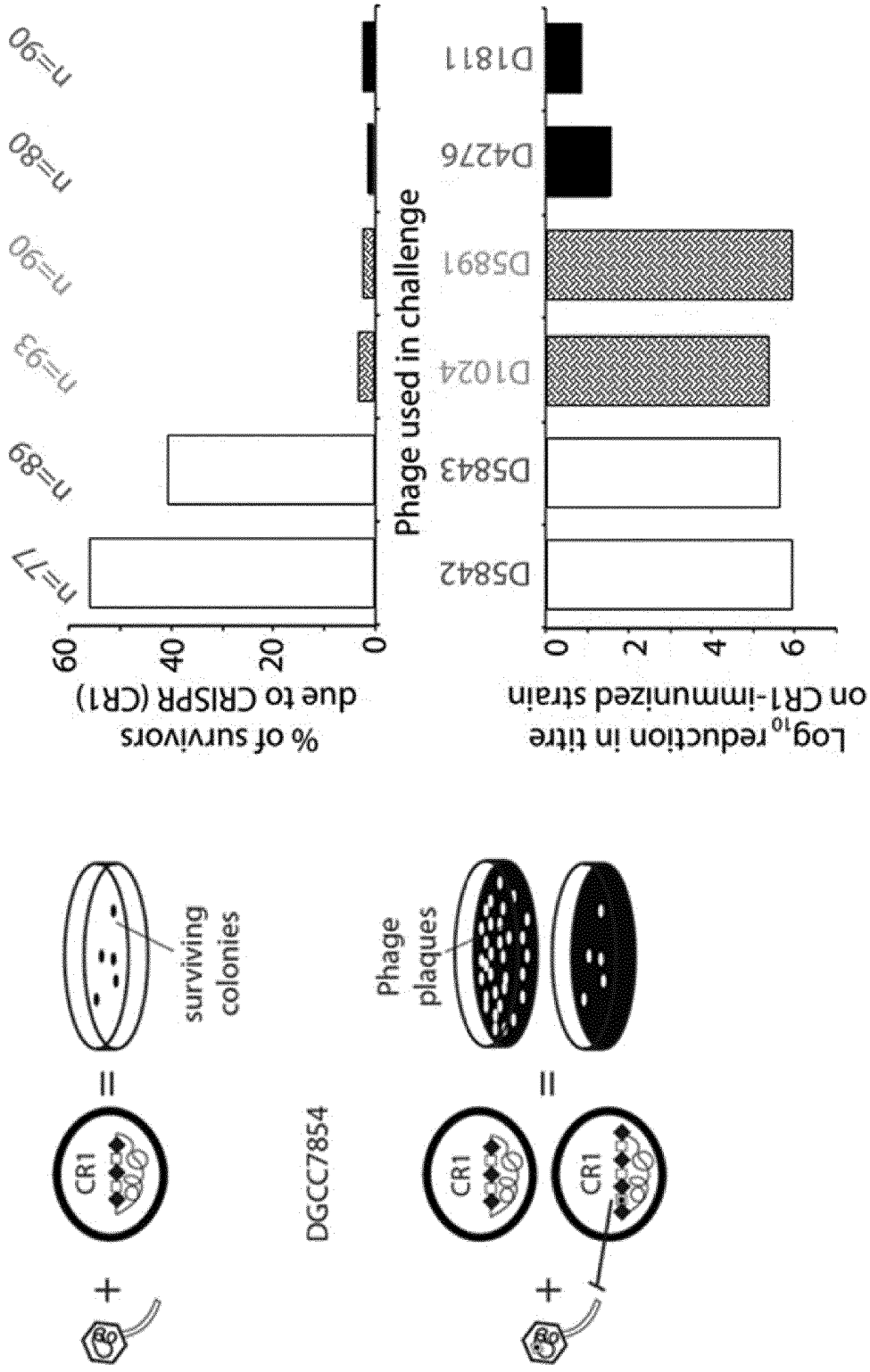

Log10 reduction in titre on CR1-immunized strain
D5842
D5843
D1024
D5891
D4276
D1811
Phage used in challenge
0
2
4
6
% of survivors due to CRISPR (CR1)
0
20
40
60
n=77
n=89
n=93
n=90
n=80
n=90
DGCC7854
CR1
CR1
Phage plaques
=
+
CR1
surviving colonies
=
+

FIG. 2A

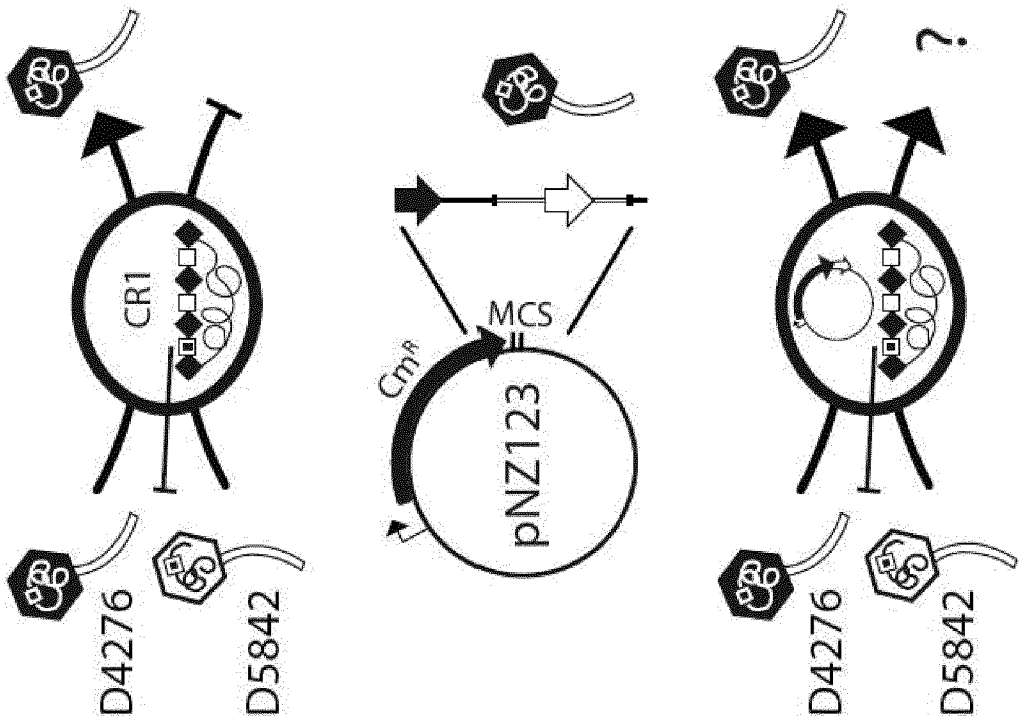


FIG. 2B

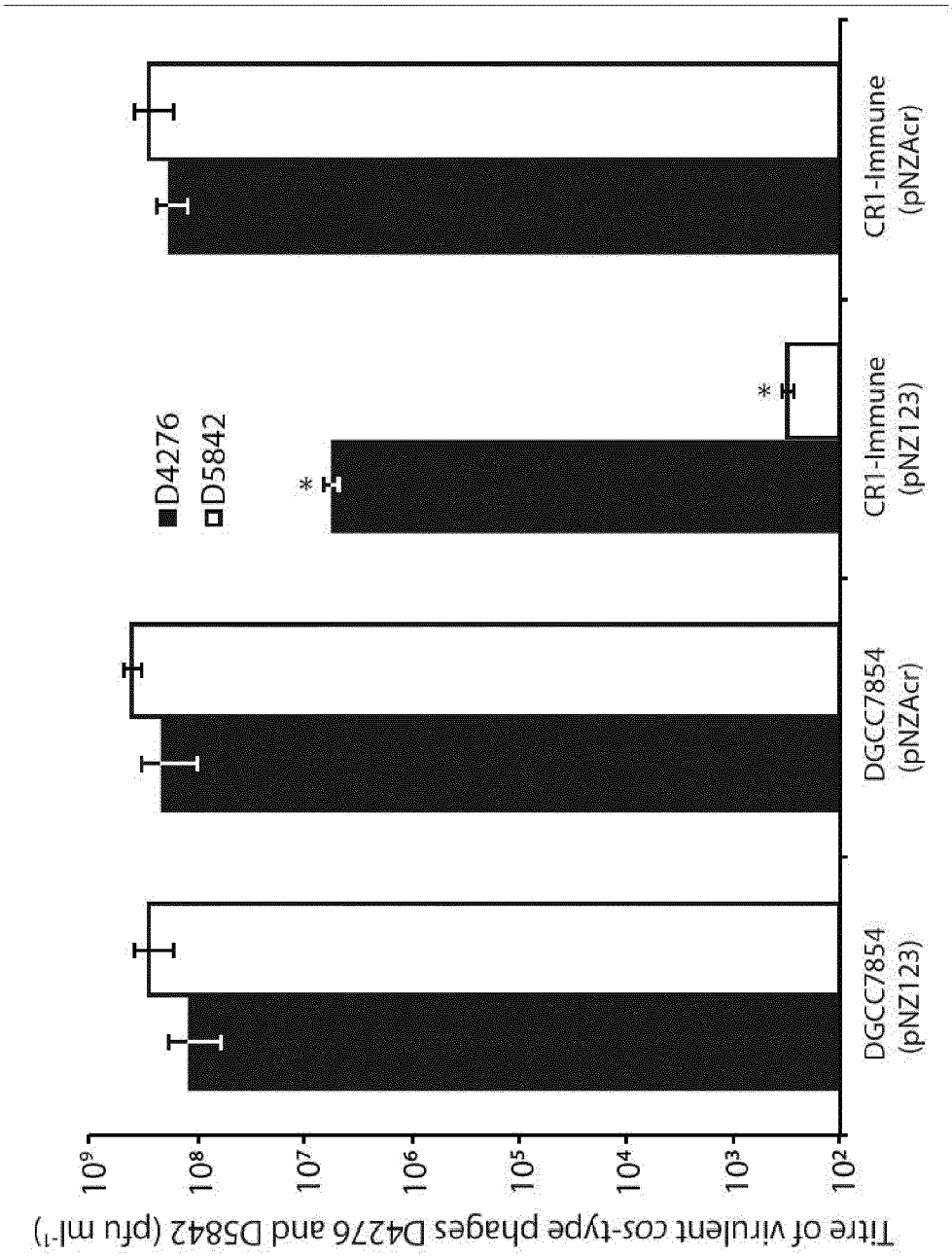

Titre of virulent cos-type phages D4276 and D5842 (pfu ml-1)
10⁹
10⁸
10⁷
10⁶
10⁵
10⁴
10³
10²
D4276
D5842
*
*
DGCC7854 (pNZ123)
DGCC7854 (pNZAcr)
CR1-Immune (pNZ123)
CR1-Immune (pNZAcr)

FIG. 2C

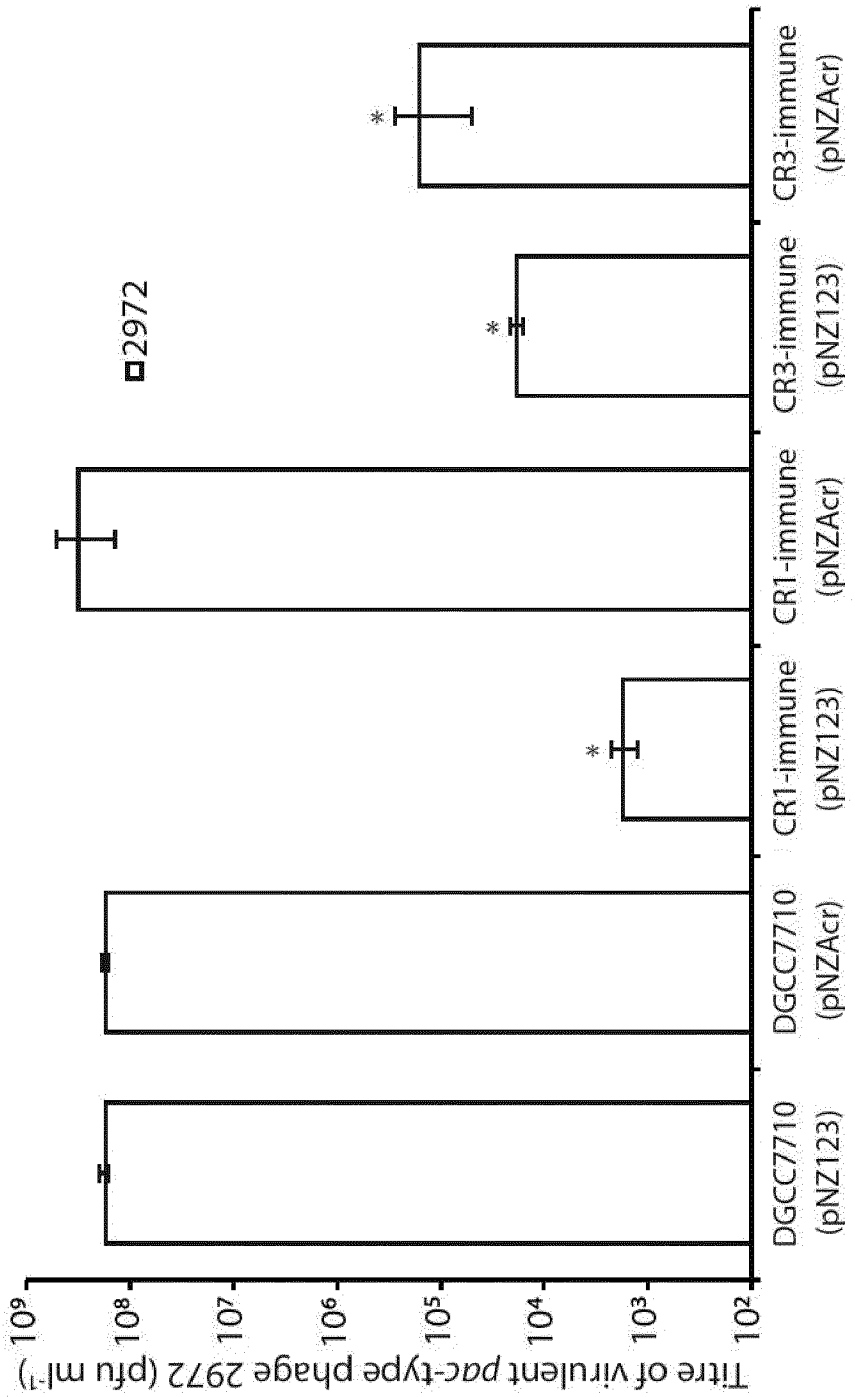
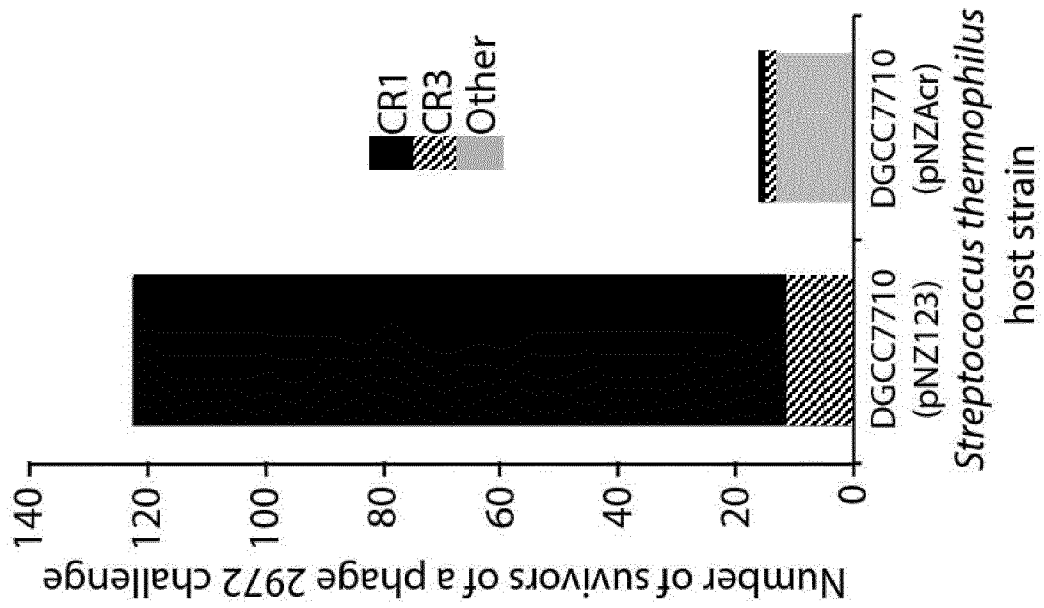

Titre of virulent pac-type phage 2972 (pfu ml-1)
10⁹
10⁸
10⁷
10⁶
10⁵
10⁴
10³
10²
2972
*
*
*
DGCC7710 (pNZ123)
DGCC7710 (pNZAcr)
CR1-immune (pNZ123)
CR1-immune (pNZAcr)
CR3-immune (pNZ123)
CR3-immune (pNZAcr)

FIG. 2D


Number of suviyors of a phage 2972 challenge
140
120
100
80
60
40
20
0
CR1
CR3
Other
DGCC7710 (pNZ123)
DGCC7710 (pNZAcr)
Streptococcus thermophilus host strain

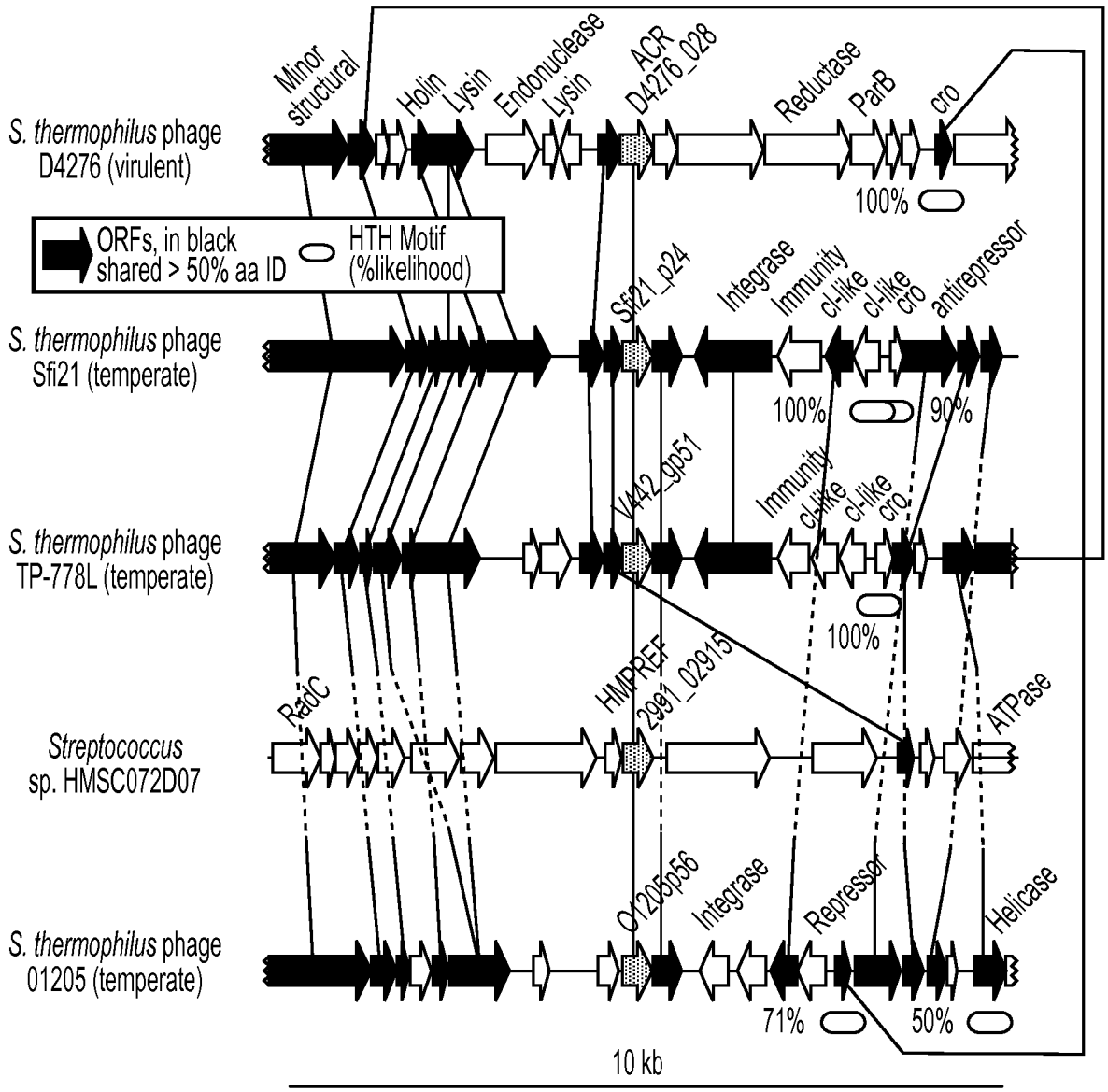

S. thermophilus phage D4276 (virulent)
Minor structural
Holin
Lysin
Endonuclease
Lysin
ACR D4276_028
Reductase
ParB
cro
100%
ORFs, in black shared > 50% aa ID
HTH Motif (%likelihood)
S. thermophilus phage Sfi21 (temperate)
Sfi21_p24
Integrase
Immunity
cI-like
cI-like
cro
antirepressor
100%
90%
S. thermophilus phage TP-778L (temperate)
V442_gp51
Immunity
cI-like
cI-like
cro
100%
Streptococcus sp. HMSC072D07
RadC
HMPREF 2091_02915
ATPase
S. thermophilus phage 01205 (temperate)
01205p56
Integrase
Repressor
Helicase
71%
50%
10 kb

FIG. 3A

FIG. 3B

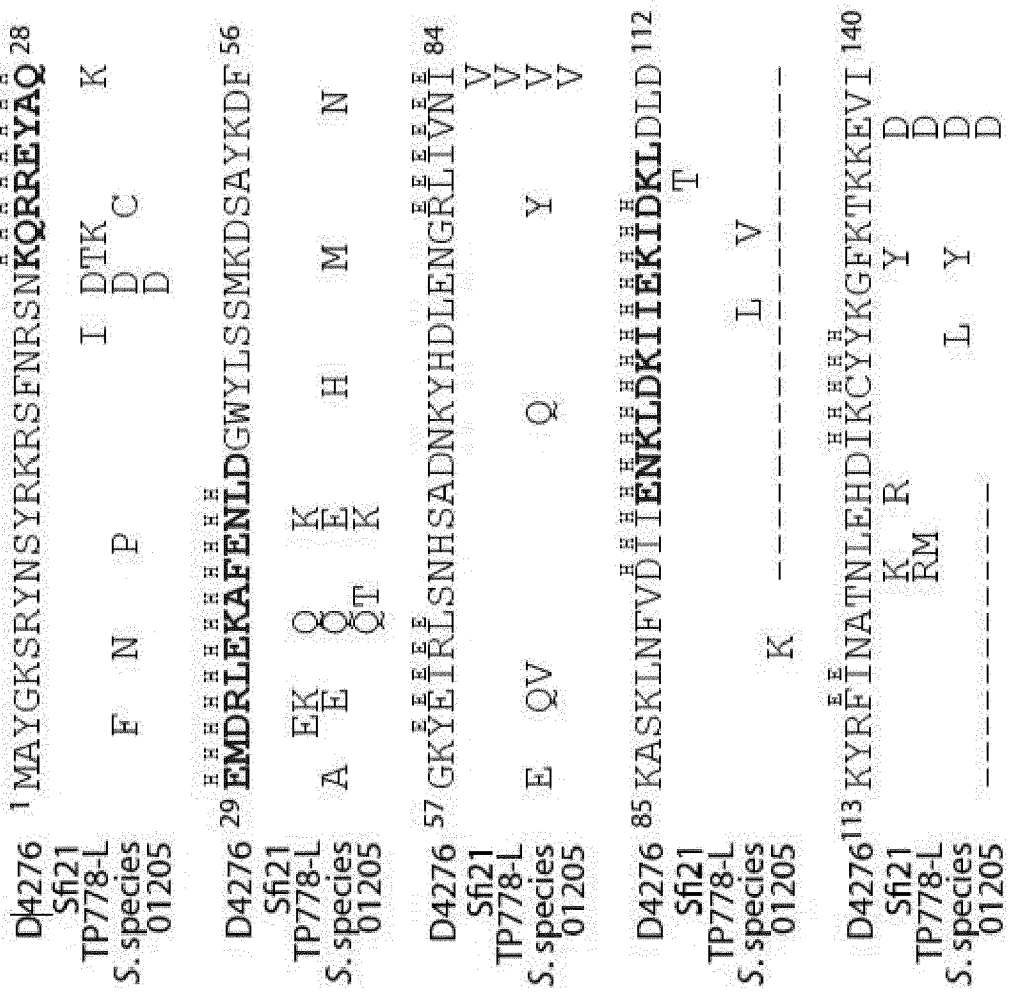

D4276 1 MAYGKSRYNSYRKRSFNRSNKQRREYAQ 28
Sfi21 I DTK K
TP778-L F N P D C
S. species 01205 D
D4276 29 EMDRLEKAFENLDGWYLSSMKDSAYKDE 56
Sfi21 EK Q K
TP778-L A E Q E H M N
S. species 01205 QT K
D4276 57 GKYEIRLSNHSADNKYHDLENGRLIVNI 84
Sfi21 V
TP778-L E QV Q Y V
S. species 01205 V
D4276 85 KASKLNFVDIIENKLDKIIEKIDKLDLD 112
Sfi21 T
TP778-L L V
S. species 01205 K
D4276 113 KYREINATNLEHDIKCYYKGFKTKKEVI 140
Sfi21 K R Y D
TP778-L RM D
S. species 01205 L Y D

FIG. 4A

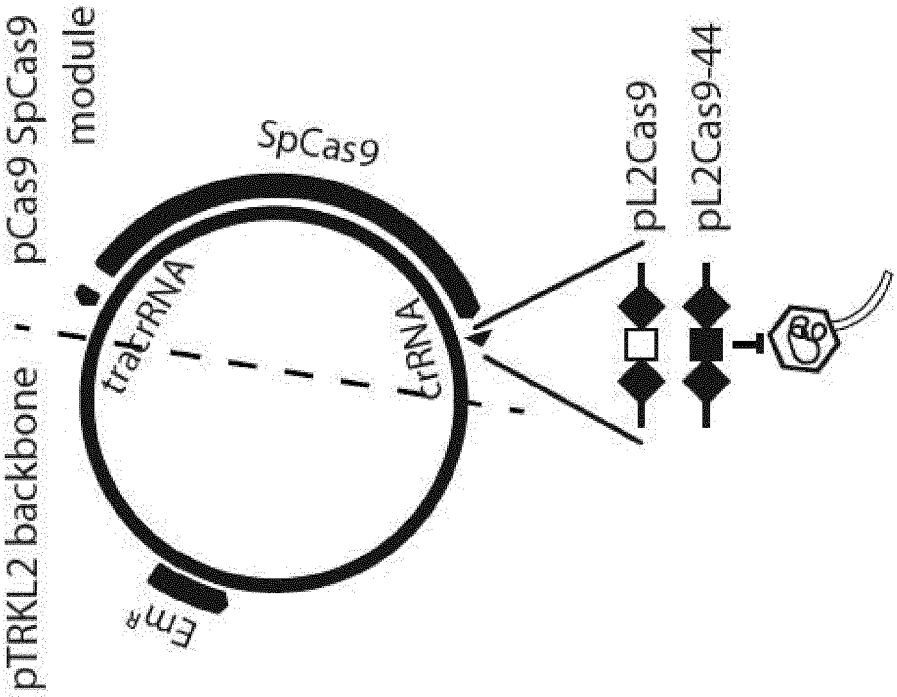

pTRKL2 backbone
pCas9 SpCas9
module
SpCas9
traᵗRNA
ᵗRNA
Emᴿ
pL2Cas9
pL2Cas9-44

FIG. 4B

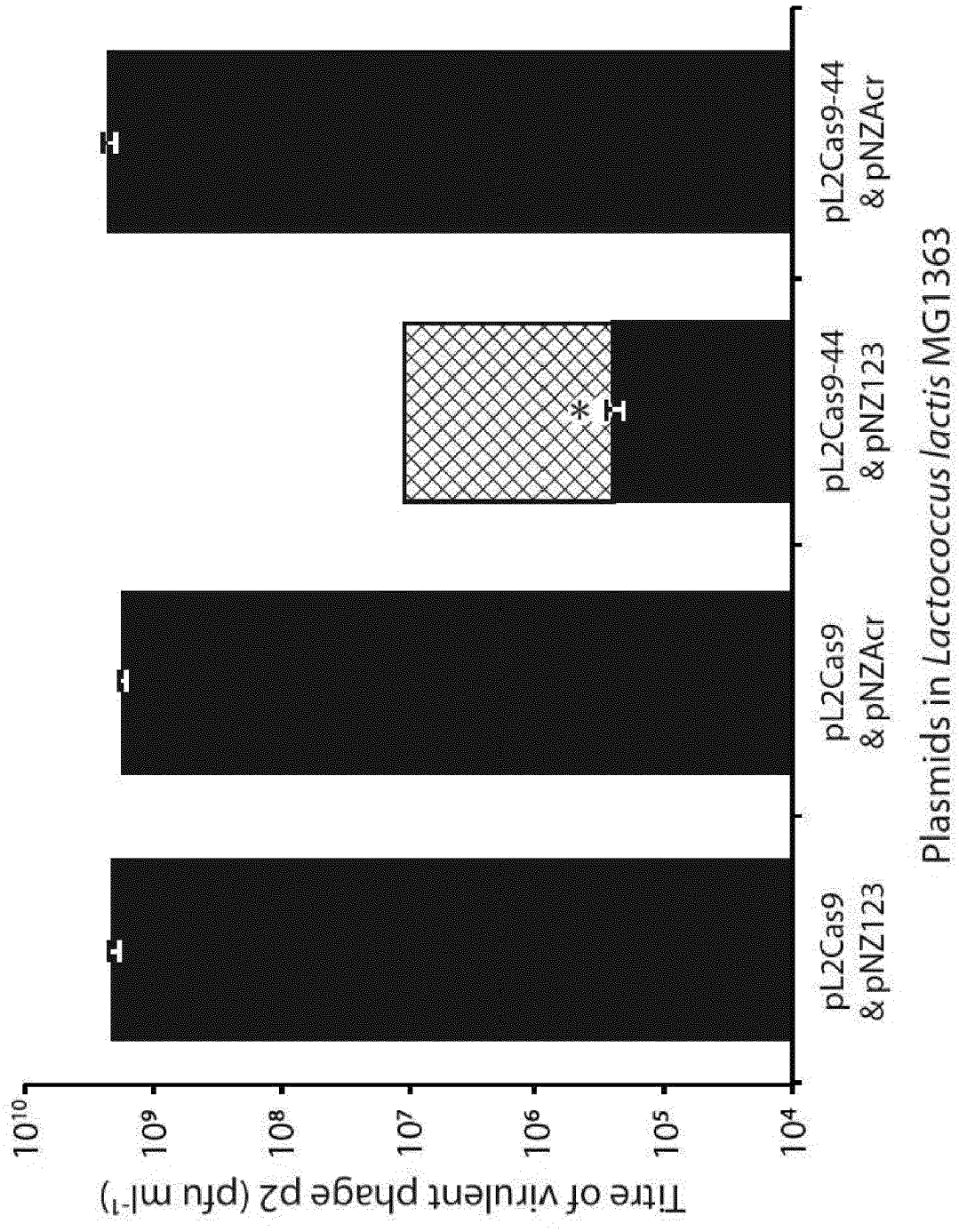

Titre of virulent phage p2 (pfu ml-1)
10^10
10^9
10^8
10^7
10^6
10^5
10^4
*
pL2Cas9 & pNZ123
pL2Cas9 & pNZAcr
pL2Cas9-44 & pNZ123
pL2Cas9-44 & pNZAcr
Plasmids in Lactococcus lactis MG1363

FIG. 5A

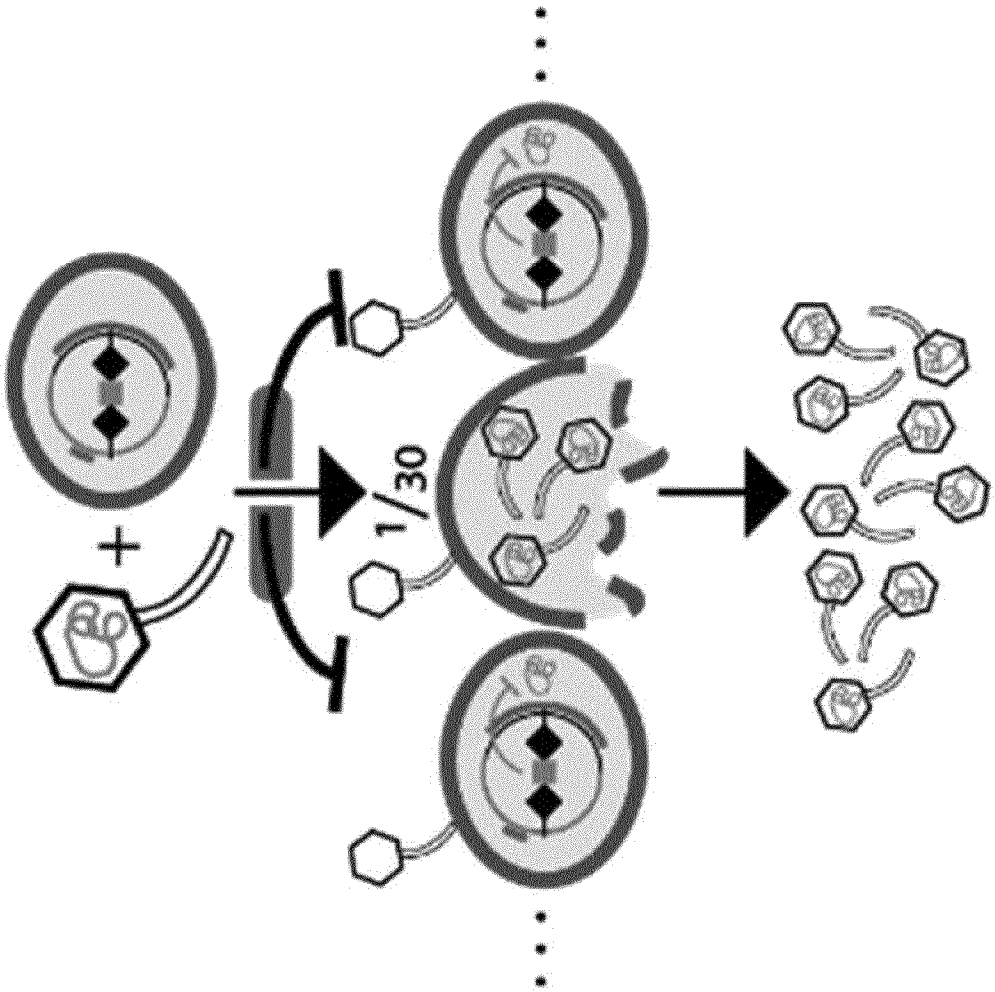


FIG. 5B

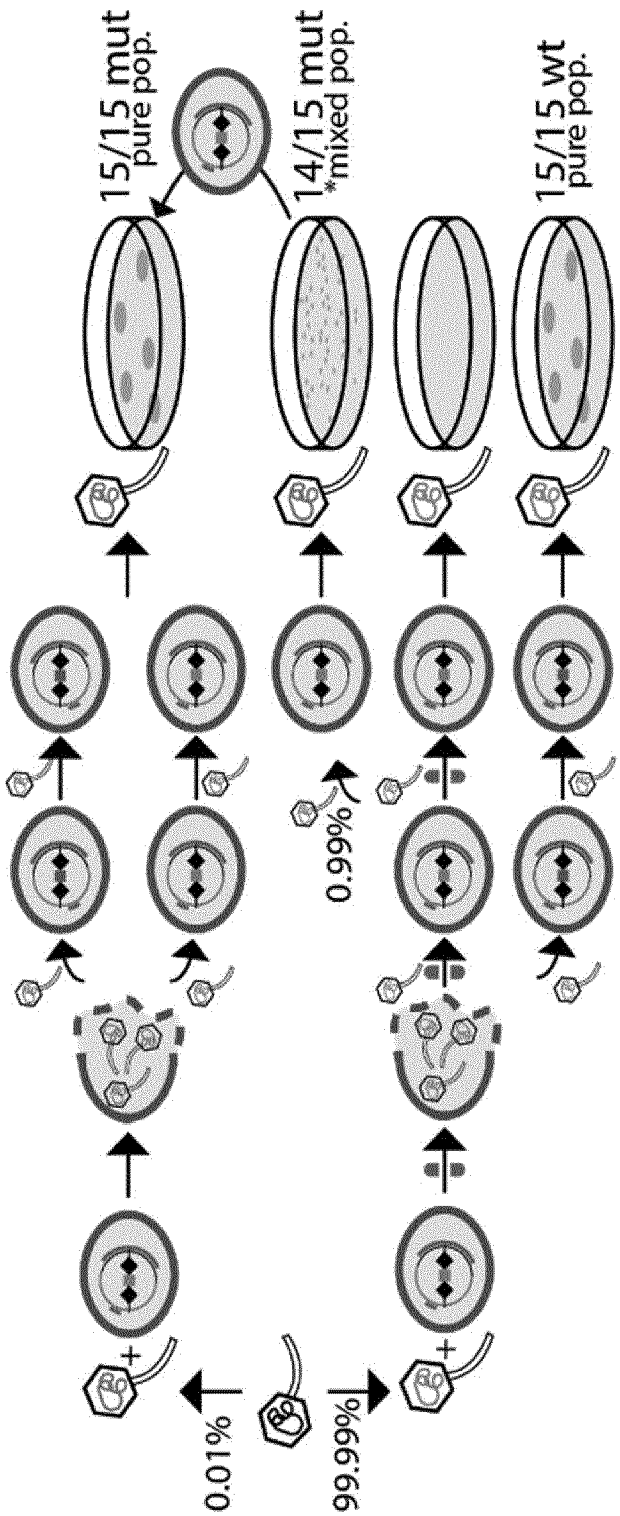

15/15 mut
pure pop.
14/15 mut
*mixed pop.
15/15 wt
pure pop.
0.99%
0.01%
99.99%

FIG. 6A

>Bacteriophage O1205 Nucleic acid (SEQ ID NO: 1)

ATGGCATACGGAAAAAGCAGATACAACTCATATAGGAAACGCAGTTTCAATAGAAGCGATAAGCAACGTAGAGAATACGCACACAAGAAATGGATAGATTAGAACAAACATTTGAA

AAACTTGATGGTTGGTATCTATCTAGCATGAAAGATAGTGCGTATATAAAGATTTCGGAAAATACGAAATTCGCTTATCAAATCATTCAGCAGACAACAAATATCATGACCTAGAAAA

TGGTCGTTTAATTGTTAATGTCAAAGCAAGTAAATTGAAATTCGTTGATATCAAATGTTACTATAAGGGATTTAAGACAAAGAAGGATGTAATCTAA

FIG. 6B

> Bacteriophage O1205 Amino acid (SEQ ID NO: 2)

MAYGIKSRYNSYRKRSFNRSDKQRREYAQEMDRLEQTFEKLDGWYLSSMKDSAYKDFGKYEIRLSNHSADNKYHDLENGRLIVNVKASKLKFVDIKCYYKGFKTKKDVI

FIG. 6C

> Bacteriophage Sfi21 Nucleic acid sequence (SEQ ID NO: 3)

ATGGCATACGGAAAAAGTAGATATAACTCATATAGAAAGCGCAGTTTTAACAGAAGTAATAAGCAACGTAGAGAATACGCACACAAGAAATGGATAGATTAGAGAAAGCTTTCGAA

AATCTTGACGGATGGTATCTATCTAGCATGAAAGATAGTGCGTACAAAGATTTCGGAAAATACGAAATTCGCTTATCAAATCATTCAGCAGACAACAAATATCATGACCTAGAAAA

TGGTCGTTTAATTGTTAATGTCAAAGCAAGTAAATTGAACTTCGTTGATATCATCGAGAACAAACTTGATAAAATCATTGAGAAGATTGATACTCTTGATTTAGATAAGTACAGATT

CATTAATGCTACTAAATTGGAACGTGATATCAAATGCTACTATAAAGGCTATAAGACAAAGAAGGATGTAATCTAA

FIG. 6D

> Bacteriophage Sfi21 Amino acid sequence (SEQ ID NO: 4)

MAYGIKSRYNSYRKRSFNRSNKQRREYAQEMDRLEKAFENLDGWYLSSMKDSAYKDFGKYEIRLSNHSADNKYHDLENGRLIVNVKASKLNFVDIIENKLDKIIEKIDTLDLDKYRFINATKLE

RDIKCYYKGYKTKKDVI

FIG. 6E

> Bacteriophage TP-778L Nucleic acid sequence (SEQ ID NO: 5)

ATGGCATACGGAAAAAGCAGATACAACTCATATAGAAAACGTAGTTTCAACATAAGTGACACAAAGCGTAGGGAATATGCAAAAGAAATGGAGAAATTAGAACAAGCATTTGAA

AAGCTAGATGGTTGGTATCTATCTAGCATGAAGGATAGTGCATACAAGGATTTTGGAAAATACGAAATCCGCTTATCAAATCATTCAGCAGACAATAAATATCATGACCTAGAAAA

TGGTCGTTTAATTGTTAATGTTAAAGCAAGTAAATTGAACTTCGTTGATATCATCGAAAACAAACTTGATAAAATCATCGAGAAGATTGATAAGCTTGATTTAGATAAGTACAGATT

TATTAACGCTACTAGAATGGAGCATGACATTAAATGCTACTATAAAGGATTTAAGACAAAGAAAGATGTAATCTAA

FIG. 6F

> Bacteriophage TP-778L Amino acid sequence (SEQ ID NO: 6)

MAYGKSRYNSYRKRSFNISDTKRREYAKEMEKLEQAFEKLDGWYLSSMKDSAYKDFGKYEIRLSNHSADNKYHDLENGRLIVNVKASKLNFVDIIENKLDKIIEKIDKLDLDIKYRFINATRMEH DIKCYYKGFKTKKDVI

FIG. 6G

> Streptococcus sp. HMSC072D07 Nucleic acid sequence (SEQ ID NO: 7)

ATGGCATTTGGCAAGAACAGATACAATCCATACAGGAAACGTAGTTTTAATCGTAGTGATAAACAATGTCGAGAGTATGCTCAGGCAATGGACGAACTAGAACAAGCCTTTGAGG AACTTGATGGATGGCACTTATCTAGTATGATGGATAGTGCTTATAAGAATTTTGAAAAGTACCAGGTTCGCCTATCAAATCATTCAGCAGACAACCAATATCATGACTTAGAAAATG GTTACTTGATTGTCAATGTTAAAGCAAGTAAATTGAACTTTGTCGATATTATCGAAAATAAATTGGATAAGATTTTAGAGAAAGTAGACAAGCTTGATCTTGATAAGTATAGGTTTA TCAATGCGACCAATCTGGAACATGATATTAAATGTTATCTCAAAGGCTATAAGACGAAAAAAGACGTGATTTAA

FIG. 6H

> Streptococcus sp. HMSC072D07 Amino acid sequence (SEQ ID NO: 8)

MAFGKNRYNPYRKRSFNRSDKQCREYAQAMDELEQAFEELDGWHLSSMMDSAYKNFEKYQVRLSNHSADNQYHDLENGYLIVNVKASKLNFVDIIENKLDKILEKVDKLDLDIKYRFINAT NLEHDIKCYLKGYKTKKDVI

FIG. 6I

> Bacteriophage D4276 Nucleic acid sequence (SEQ ID NO: 9)

ATGGCATACGGAAAAAGTAGATATAACTCATATAGAAAGCGCAGTTTTAACAGAAGTAATAAGCAACGTAGAGAATACGCACACAAGAAATGGATAGATTAGAGAAAGCTTTCGAA AATCTTGACGGATGGTATCTATCTAGCATGAAAGACAGTGCTTACAAGGATTTTGGGAAATACGAAATTCGCTTATCAAATCATTCGGCAGACAACAAATATCACGACTTAGAAAA CGGTCGTTTAATTGTTAATATTAAAGCTAGTAAATTGAATTCGTTGATATCATCGAGAATAAGCTTGATAAAATAATCGAGAAGATTGATAAGCTTGATTTAGATAAGTACCGATT CATCAATGCGACCAACCTAGAGCATGATATCAAATGCTATTACAAGGGGTTTAAAACGAAAAAGGAGGTAATCTAA

FIG. 6J

> Bacteriophage D4276 Amino acid sequence (SEQ ID NO: 10)

MAYGKSRYNSYRKRSFNRSNKQRREYAQEMDRLEKAFENLDGWYLSSMKDSAYKDFGKYEIRLSNHSADNKYHDLENGRLIVNIKASKLNFVDIIENKLDKLIEKIDKLDLDIKYRFINATNLE HDIKCYYKGFKTKKEVI

FIG. 6K

> Bacteriophage D1126 Nucleic acid sequence (SEQ ID NO: 11)

ATGGCATACGGAAAAAGCAGATACAATTCATATAGGAAGCGAAACTTCTCTATAAGCGACAATCAGCGTAGGGAATATGCTAAAAAAATGAAGGAGTTAGAACAAGCGTTTG
AAAACCTTGACGGATGGTATCTATCTAGCATGAAAGATAGTGCGTACAAAGATTTCGGAAAATACGAAATTCGCTTATCAAATCATTCAGCAGACAATAGATATCATGACCTAG
AAAATGGTCGCTTAATCGTTAATGTTAAAGCTAGTAAATTGAACTTCGTTGATATCATCGAGAATAAACTTGGTAAAATCATTGAGAAGATTGATACTCTTGATTTAGATAAGTA
CAGATTCATTAATGCTACTAAATTGGAACGTGATATCAAATGCTACTATAAGACAAAGAAGGATGTAATCTAA

FIG. 6L

> Bacteriophage D1126 Amino acid sequence (SEQ ID NO: 12)

MAYGKSRYNSYRKRNFSISDNQRREYAKKMKELEQAFENLDGWYLSSMKDSAYKDFGKYEIRLSNHSADNRYHDLENGRLIVNVKASKLNFVDIIENKLGKIIEKIDTLDLDKYRFINATKL
ERDIKCYYKGYKTKKDVI

FIG. 6M

> Bacteriophage D4250 Nucleic acid sequence (SEQ ID NO: 13)

ATGGCATACGGAAAAAGTAGATATAACTCATATAGAAAACGCAGTTTCAACAGAAGCGATAAACAGCGTAGAGAATACGCACAAGCAATGGAAGAATTAGAGCAAGCATTTG
AAAACTTTGATGATTGGTATCTATCAAGCATGAAAGACAGTGCTTACAAGGATTTTGGGAAATACGAAATTCGCTTATCAAATCATTCGGCAGACAACAAATATCACGACTTAG
AAAACGGTCGTTTAATTGTTAATATTAAAGCTAGTAAATTGAATTTCGTTGATATCATCGAGAATAAGCTTGATAAAATAATCGAGAAGATTGATAAGCTTGATTTAGATAAGTA
CCGATTCATCAATGCGACCAACCTAGAGCATGATATCAAATGCTATTACAAGGGGTTTAAAACGAAAAAGGAGGTAATCTAA

FIG. 6N

> Bacteriophage D4250 Amino acid sequence (SEQ ID NO: 14)

MAYGKSRYNSYRKRSFNRSDKQRREYAQAMEELEQAFENFDDWYLSSMKDSAYKDFGKYEIRLSNHSADNKYHDLENGRLIVNIKASKLNFVDIIENKLDKIIEKIDKLDLDKYRFINATN
LEHDIKCYYKGFKTKKEVI

FIG. 6O

> Bacteriophage D4252 Nucleic acid sequence (SEQ ID NO: 15)
ATGGCATACGGAAAAAGCAGATACAACTCATATAGAAAGCGCAGTTTTAACAGAAGTGATAAGCAACGTAGAGAATACGCTAAAAAAATGAAGGAGTTAGAACAAGCGTTTGAAAACCT
TGATGGTTGGTATCTATCGAGCATGAATGACAGTGCTTATAAAAATTTTGGCAAATATGAAGTTCGATTGTCAAATCATTCGGCAGATAATAAATATCACGACATAGAAAACGGTCGTTTAA
TTGTTAATGTTAAAGCTAGTAAATTGAATTTCGTTGATATCATCGAGAACAAGCTTGATAAAATAATCGAGAAGAATTGATAAGCTTGATTTAGATAAGTACCGATTCATCAACGCTACCAAT
CTAGAGCATAATATTAAATGCTATTACAAGGGATTTAAGACAAAGAAGGGATGTAATATAA

FIG. 6P

> Bacteriophage D4252 Amino acid sequence (SEQ ID NO: 16)
MAYGKSRYNSYRKRSFNRSDKQRREYAKKMKELEQAFENLDGWYLSSMNDSAYKNFGKYEVRLSNHSADNKYHDIENGRLIVNVKASKLNFVDIIENKLDKIIEKIDKLDLDKYRFINATNLEHNIKCY
YKGFKTKKDVI

FIG. 6Q

> Bacteriophage D4598 Nucleic acid sequence (SEQ ID NO: 17)
ATGGCATACGGAAAAAGTAGATATAACTCATATAGAAAACGCAGTTTCAACAGAAGCGATAAACAGCGTGGAGAATACGCACACAAGCAATGGAAGAATTAGAGCAAGCATTTGAAAACTT
TGATGATTGGTATCTATCAAGCATGAAAGACAGTGCTTACAAGGATTTTGGGAAATACGAAATTCGCTTATCAAATCATTCGGCAGACAATAAATATCATGACCTAGAAAATGGTCGCTTA
ATCGTTAATGTTAAAGCTAGTAAATTGAACTTCGTCGATATCATCGAGAATAAAATCGATAAAATCATTGAGAAGATTGATAAGCTTGATTTAGATAAGTACCGATTCATCAACGCTACCAA
CCTAGAGCATGATATCAAATGTTATTACAAGGGATTTAAGACAAAAAAGGATGTAATCTAA

FIG. 6R

> Bacteriophage D4598 Amino acid sequence (SEQ ID NO: 18)
MAYGKSRYNSYRKRSFNRSDKQRGEYAQAMEELEQAFENFDDWYLSSMKDSAYKDFGKYEIRLSNHSADNKYHDLENGRLIVNVKASKLNFVDIIENKIDKLDLDKYRFINATNLEHDIKCY
YKGFKTKKDVI

FIG. 6S

>Streptococcus_mutans_NLML9 (SEQ ID NO: 19)
ATGGCATTTGGAAAAAGAAGATATAACTCGTATCGTAAACGCAGTTTTAATAGAAGTGATAAGCAACGTCGAGAATATGCACAAGCAATGGAAGAACTTGAACAAACATTTGAAAATCTTGAAG
GTTGGAATTTATCAAGCATGAAAGATAGTGCTTATAAAGATTATGATAAATATGAAGTTCGACTTTCAAATCATTCAGCTGATAATCAATATCATAACTTACAAGATGGTAAATTAATCATCAATAT
CAAAGCTAGTAAAATGAATTTTGTTTGGATTATAGAAAATAAACTTGATGCAATTCTTGAAAAAGTAAATAAGTTAGACCTTAGCAAATACAGATTTATTAATGCTACAAGTTTAGATCATGATAT
CAAATGTTATTACAAAAATTATAAAACAAAGAAAGATGTAATTTAA

FIG. 6T

>Streptococcus_mutans_NLML9 (SEQ ID NO: 20)
MAFGKRRYNSYRKRSFNRSDKQRREYAQAMEELEQTFENLEGWNLSSMKDSAYKDYDKYEVRLSNHSADNQYHNLQDGKLIINIKASKMNFVWIIENKLDAILEKVNKLDLSKYRFINATSLDHDIKCYYK
NYKTKKDVI

FIG. 6U

>Streptococcus_mutans_AD01 (SEQ ID NO: 21)
ATGGCATTTGGAACAAGAAGATATAATTCATATCGTAAACGCAGTTTTAATAGAAGTGATAAGCAACGTCGAGAATATGCACAAGCAATGGAAGAACTTGAACAAACATTTGAAAATCTTGAAG
ATTGGAATTTGTCGAGCATGAAAGATAGTGCTTATAAAGATTATGATAAATATGAAGTTCGACTTTCAAATCATTCAGCTGATAATCAATATCATAACTTACAAGATGGTAAATTAATCATCAATAT
CAAAGCTAGTAAAATGAATTTTGTTTGGATTATAGAAAATAAACTTGATGCAATTCTTGAAAAAGTAAATAAGTTAGACCTTAGCAAATACAGATTTATTAATGCTACAAATTTAGAACATGATAT
CAAATGTTATTACAAAAAATTATAAAAACAAAGAAAGATGTAATTTAA

FIG. 6V

>Streptococcus_mutans_AD01 (SEQ ID NO: 22)
MAFGTRRYNSYRKRSFNRSDKQRREYAQAMEELEQTFENLEDWNLSSMKDSAYKDYDKYEVRLSNHSADNQYHNLQDGKLIINIKASKMNFVWIIENKLDAILEKVNKLDLSRYRFINATNLEHDIKCYYK
NYKTKKDVI

FIG. 6W

>Streptococcus_mutans_N66 (SEQ ID NO: 23)
ATGGCATTTGGAACAAGAAGATATAATTCATATCGTAAACGCAGTTTTAATAGAAGTGATAAGCAACGTCGAGAATATGCACAAGCAATGGAAGAACTTGAACAAACATTTGAAAATC
TTGAAGATTGGAATTTGTCGAGCATGAAAGATAGTGCTTATAAAGATTATGATAAATATGAAGTTAGACTTTCAAATCATTCAGCTGATAATCAATATCATAACTTACAAGATGGTAAAT
TAATCATCAATATCAAAGCTAGTAAAATGAATTTTGTTTGGATTATAGAAAATAAACTTGATGTAATTCTTGAAAAAGTAAATAAGTTAGACCTTAGCAAATACAGATTTATTAATGCTAC
AAGTTTAGATCATGATATCAAATGTTATTACAAAAATTATAAAACAAAGAAAGATGTAATCTAA

FIG. 6X

>Streptococcus_mutans_N66 (SEQ ID NO: 24)
MAFGTRRYNSYRKRSFNRSDKQRREYAQAMEELEQTFENLEDWNLSSMKDSAYKDYDKYEVRLSNHSADNQYHNLQDGKLIINIKASKMNFVWIIENKLDVILEKVNKLDLSKYRFINATSLDH
DIKCYYKNYKTKKDVI

FIG. 6Y

>Streptococcus_mutans_24 (SEQ ID NO: 25)
ATGGCATTTGGAACAAGAAGATATAATTCATATCGTAAACGCAATTTTAATAGAAGTGATAAACAACGTCGAGAATATGCACAAGCAATGGAAGAACTTGAAACAAACATTTGAAAATCT
TGAAGATTGGAATTTGTCGAGCATGAAAGATAGTGCTTATAAAGATTATGATAAATTTGAAGTTCGACTTTCAAATCATTCAGCTGATAATCAATATCATAACTTACAAGATGGTAAATT
AATCATCAATATCAAAGCTAGTAAAATGAATTTTGTTTGGATTATAGAAAATAAACTTGATGCAATTCTTGAAAAAGGTAAATAAGTTAGACCTTAGCAAATACAGATTTATTAATGCTAC
AAGTTTAGATCATGATATCAAATGTTATTACAAAAATTATAAAACAAAAAAAGATGTAATTTAA

FIG. 6Z

>Streptococcus_mutans_24 (SEQ ID NO: 26)
MAFGTRRYNSYRKRNFNRSDKQRREYAQAMEELEQTFENLEDWNLSSMKDSAYKDYDKFEVRLSNHSADNQYHNLQDGKLIINIKASKMNFVWIIENKLDAILEKVNKLDLSKYRFINATSLDH
DIKCYYKNYKTKKDVI

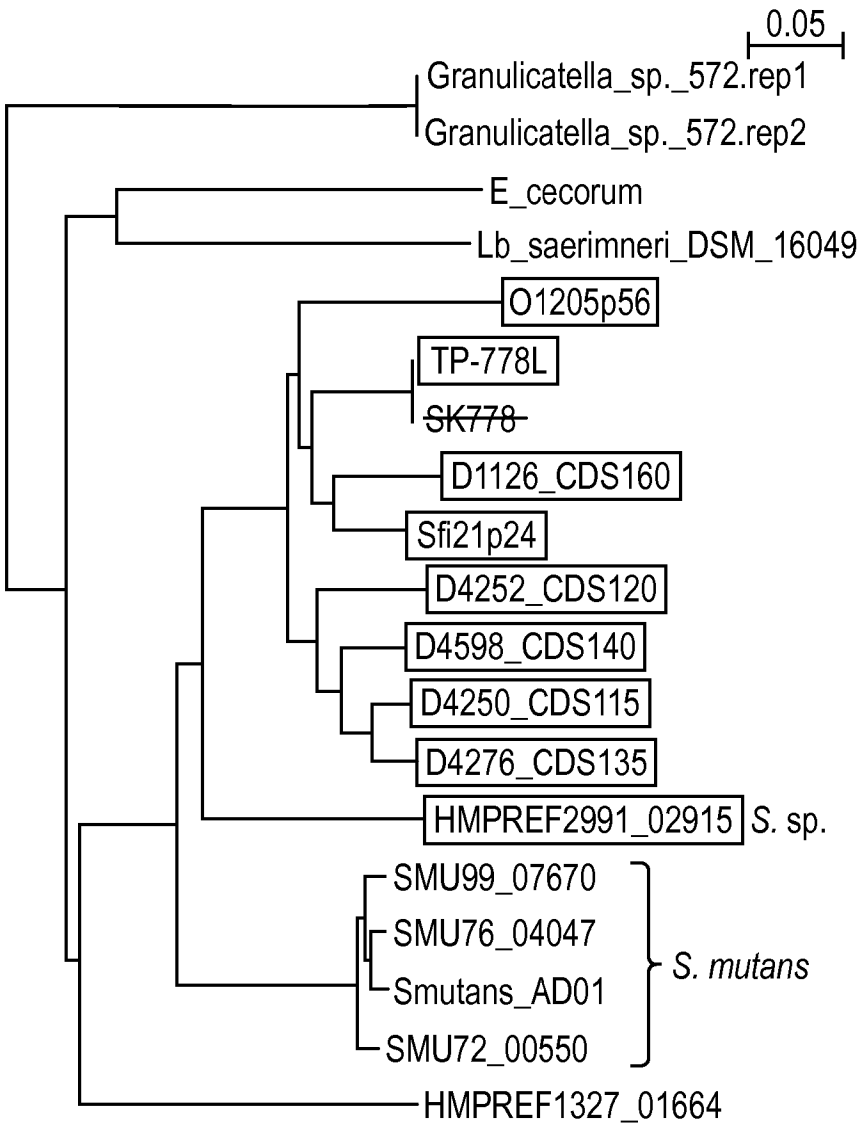

0.05
Granulicatella_sp._572.rep1
Granulicatella_sp._572.rep2
E_cecorum
Lb_saerimneri_DSM_16049
O1205p56
TP-778L
SK778
D1126_CDS160
Sfi21p24
D4252_CDS120
D4598_CDS140
D4250_CDS115
D4276_CDS135
HMPREF2991_02915
S. sp.
SMU99_07670
SMU76_04047
Smutans_AD01
SMU72_00550
S. mutans
HMPREF1327_01664

FIG. 7

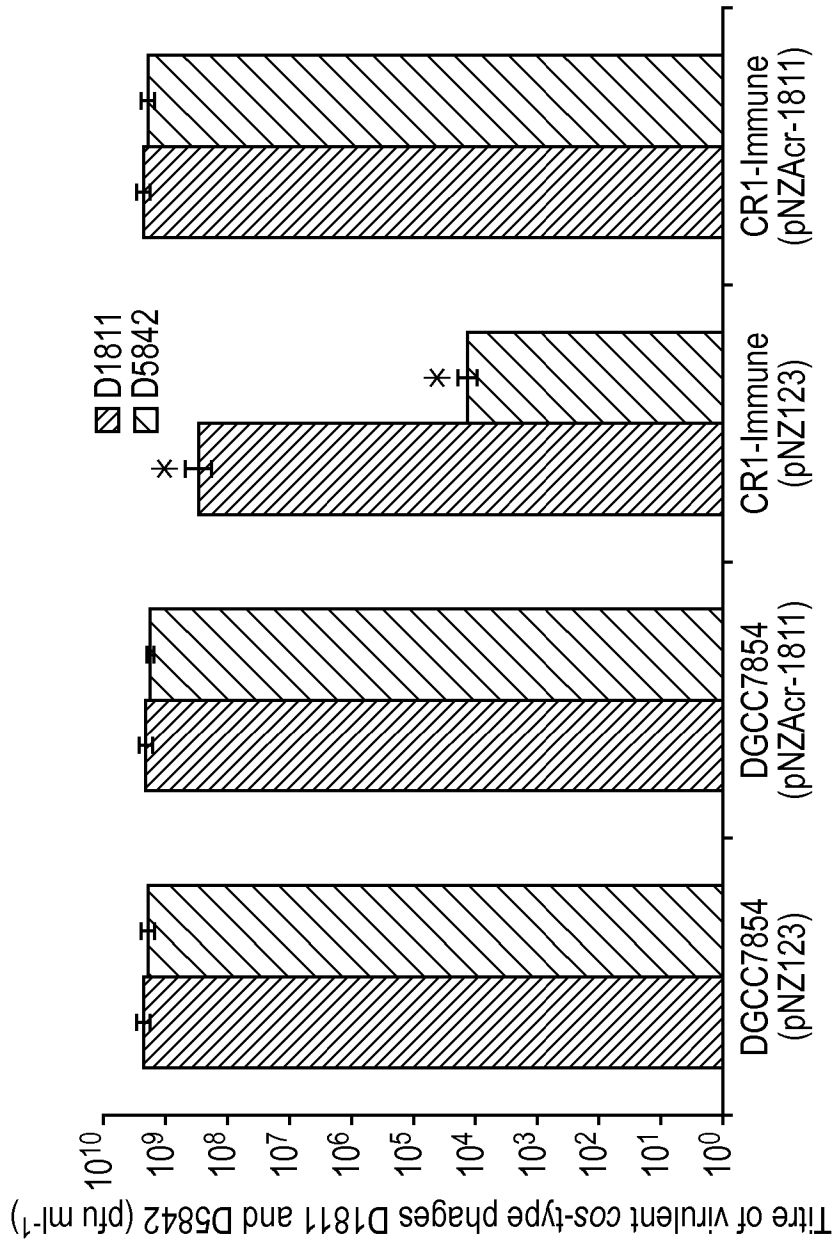

Titre of virulent cos-type phages D1811 and D5842 (pfu ml-1)
10^10
10^9
10^8
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
D1811
D5842
DGCC7854 (pNZ123)
DGCC7854 (pNZAcr-1811)
CR1-Immune (pNZ123)
CR1-Immune (pNZAcr-1811)

FIG. 8A

FIG. 8B

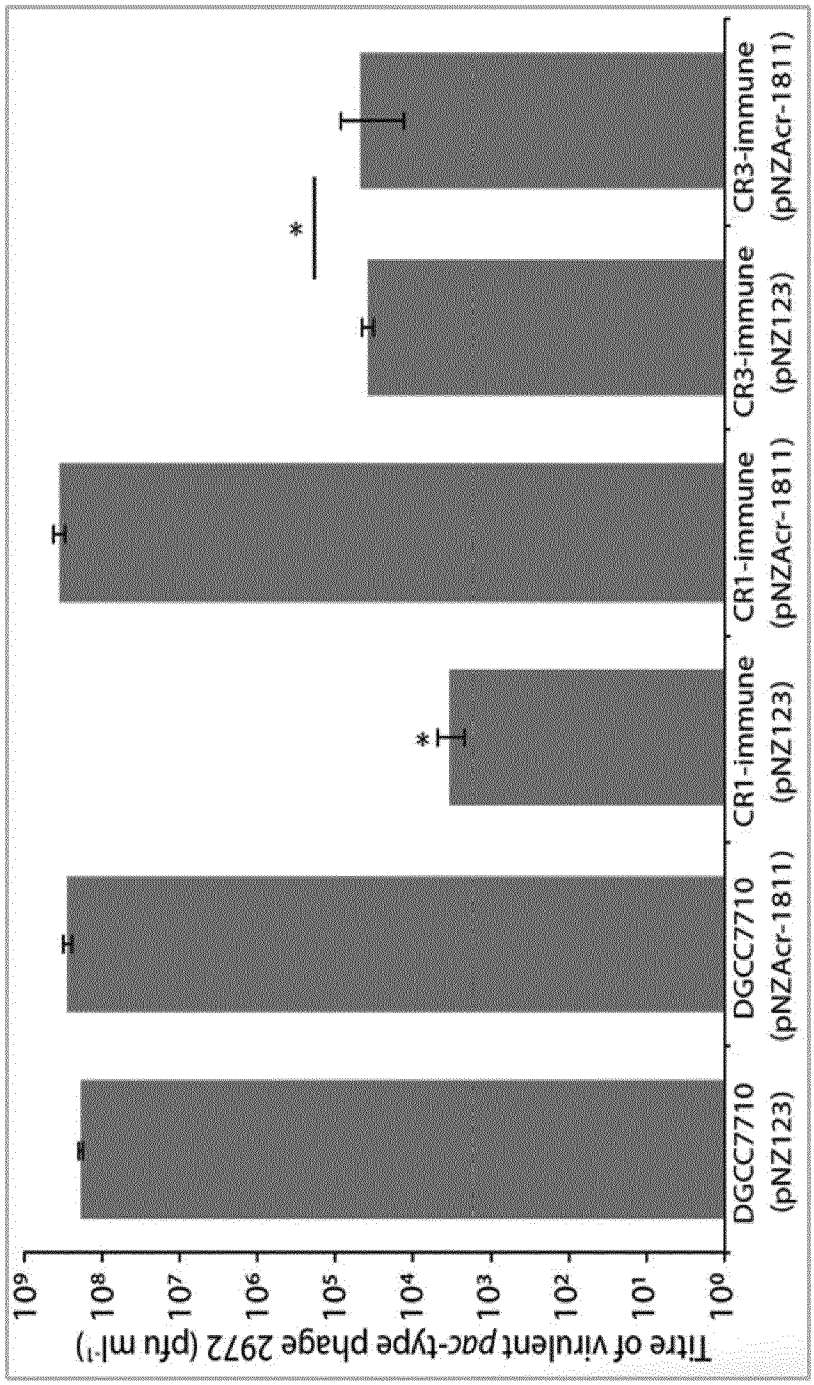

Titre of virulent pac-type phage 2972 (pfu ml-1)
10⁹
10⁸
10⁷
10⁶
10⁵
10⁴
10³
10²
10¹
10⁰
*
*
DGCC7710 (pNZ123)
DGCC7710 (pNZAcr-1811)
CR1-immune (pNZ123)
CR1-immune (pNZAcr-1811)
CR3-immune (pNZ123)
CR3-immune (pNZAcr-1811)

FIG. 9

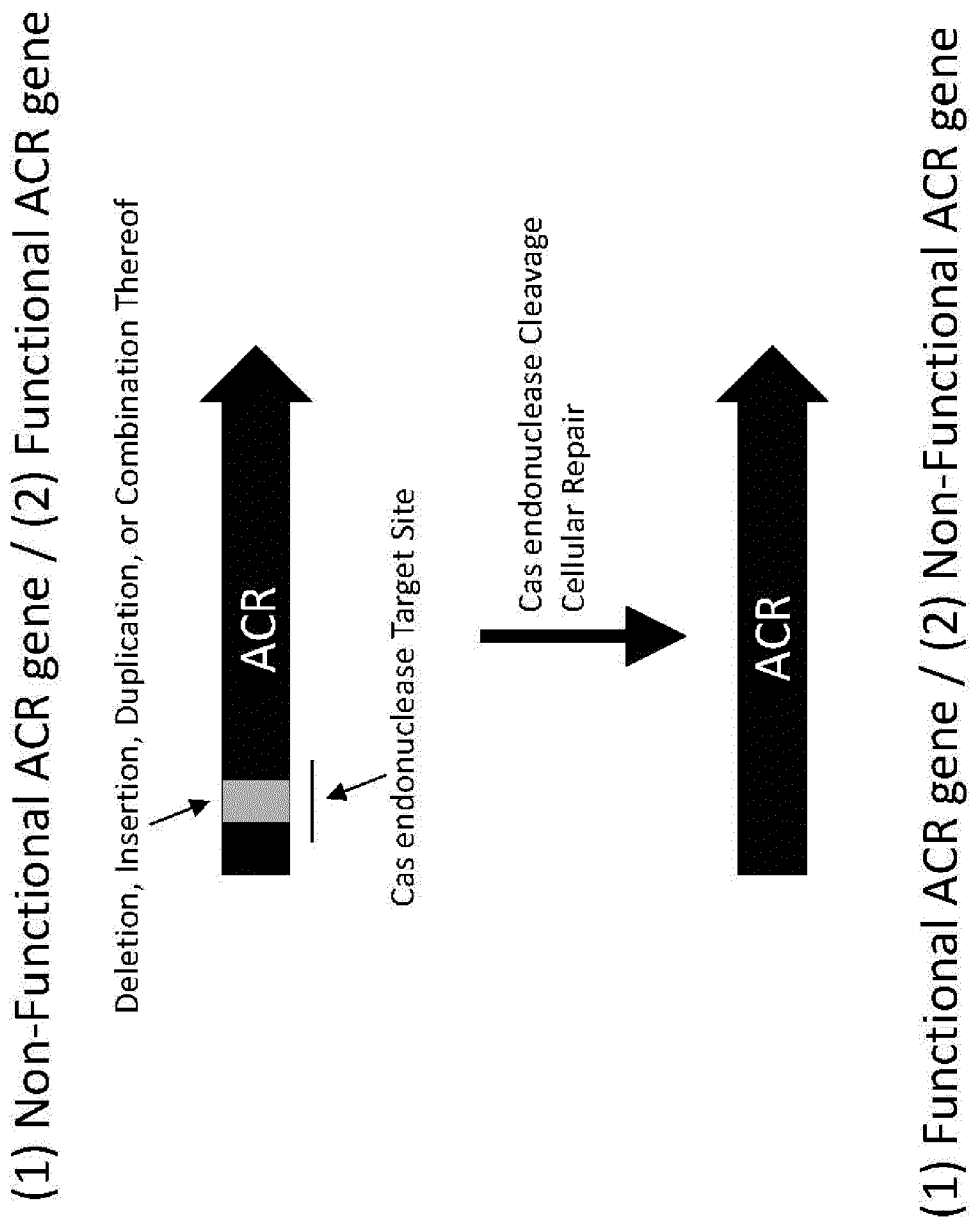

(1) Non-Functional ACR gene / (2) Functional ACR gene
Deletion, Insertion, Duplication, or Combination Thereof
ACR
Cas endonuclease Target Site
Cas endonuclease Cleavage
Cellular Repair
ACR
(1) Functional ACR gene / (2) Non-Functional ACR gene

FIG. 10

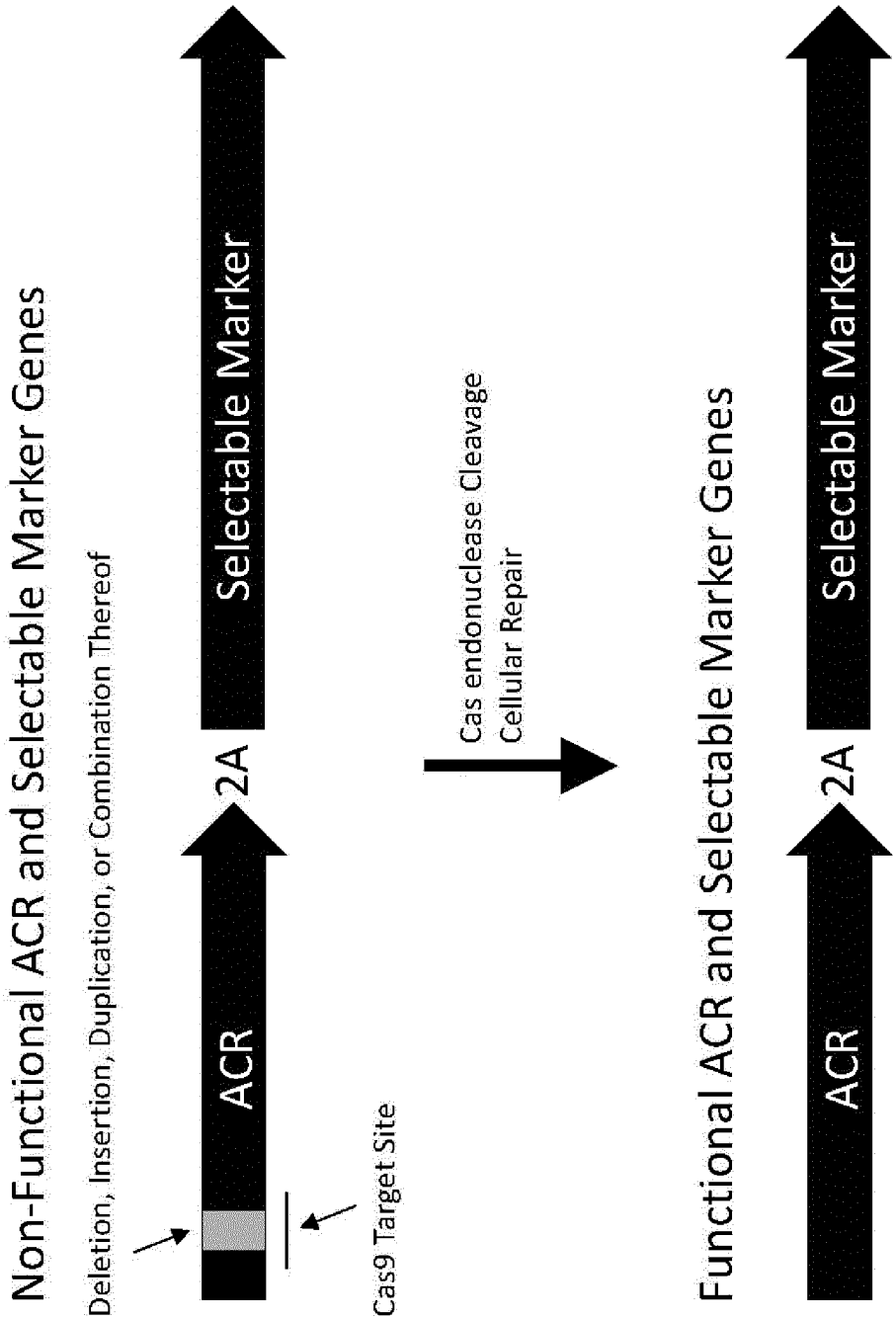
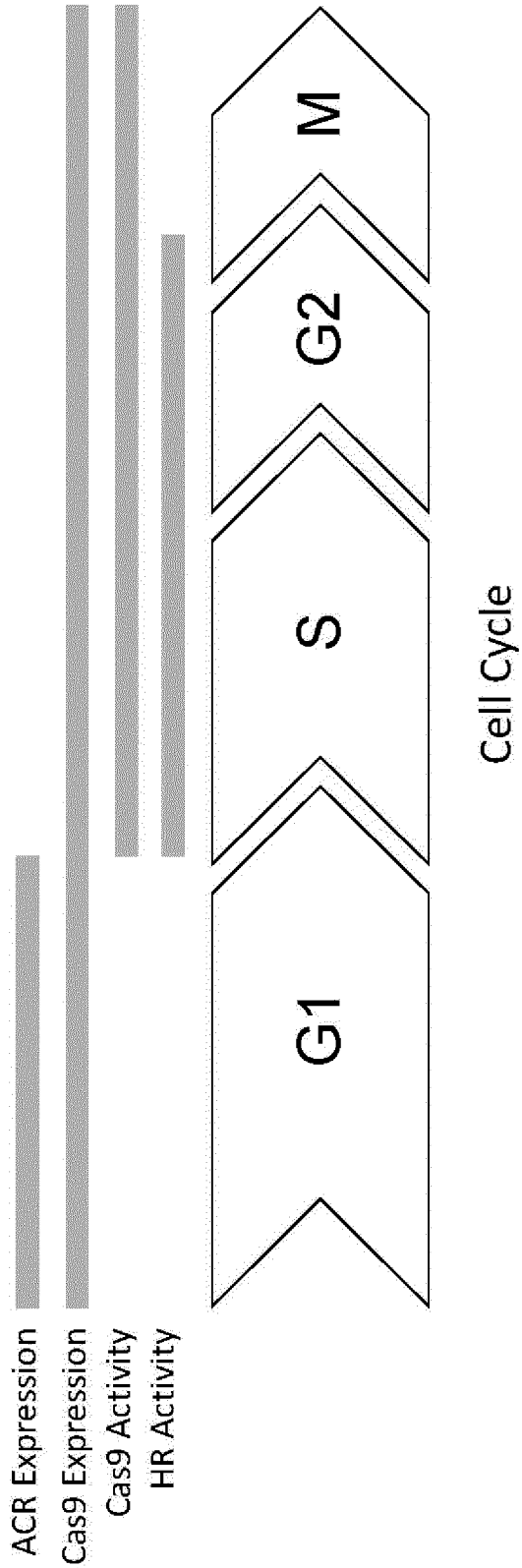

Non-Functional ACR and Selectable Marker Genes
Deletion, Insertion, Duplication, or Combination Thereof
ACR
2A
Selectable Marker
Cas9 Target Site
Cas endonuclease Cleavage
Cellular Repair
Functional ACR and Selectable Marker Genes
ACR
2A
Selectable Marker

FIG. 11


G1
S
G2
M
Cell Cycle
ACR Expression
Cas9 Expression
Cas9 Activity
HR Activity

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	**FOR FURTHER ACTION**	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
P114808PCT		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/EP2018/060510	24 April 2018 (24-04-2018)	24 April 2017 (24-04-2017)
Applicant		
DUPONT NUTRITION BIOSCIENCES APS		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 8 sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**
   a. With regard to the **language,** the international search was carried out on the basis of:
      ☒ the international application in the language in which it was filed
      ☐ a translation of the international application into _____, which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b))
   b. ☐ This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).
   c. ☒ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. ☐ **Certain claims were found unsearchable** (See Box No. II)

3. ☒ **Unity of invention is lacking** (see Box No III)

4. With regard to the **title,**
   ☒ the text is approved as submitted by the applicant
   ☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract,**
   ☒ the text is approved as submitted by the applicant
   ☐ the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority

6. With regard to the **drawings,**
   a. the figure of the **drawings** to be published with the abstract is Figure No. 1
      ☐ as suggested by the applicant
      ☒ as selected by this Authority, because the applicant failed to suggest a figure
      ☐ as selected by this Authority, because this figure better characterizes the invention
   b. ☐ none of the figures is to be published with the abstract

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2018/060510

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:

      [X] in the form of an Annex C/ST.25 text file.

      [ ] on paper or in the form of an image file.

   b. [ ] furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. [ ] furnished subsequent to the international filing date for the purposes of international search only:

      [ ] in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      [ ] on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2018/060510

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.:

1-10

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2018/060510

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/005 C07K14/315 C12N5/10 C12N15/63
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)
C07K C12N

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, BIOSIS, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	RAUCH BENJAMIN J ET AL: "Inhibition of CRISPR-Cas9 with Bacteriophage Proteins", CELL, CELL PRESS, AMSTERDAM, NL, vol. 168, no. 1, 29 December 2016 (2016-12-29), page 150, XP029882136, ISSN: 0092-8674, DOI: 10.1016/J.CELL.2016.12.009	1-10
Y	page 155, left-hand column, line 3 starting from the bottom - page 156, left-hand column, line 28; figure 6	1-10

-/--

[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
14 June 2018	12/09/2018

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	Brenz Verca, Stefano

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.
X	PAWLUK APRIL ET AL: "Naturally Occurring Off-Switches for CRISPR-Cas9", CELL, CELL PRESS, AMSTERDAM, NL, vol. 167, no. 7, 8 December 2016 (2016-12-08), page 1829, XP029850707, ISSN: 0092-8674, DOI: 10.1016/J.CELL.2016.11.017	1-8,10
Y	page 1830, right-hand column, lines 16-37; figure 1C page 1832, right-hand column, lines 7-16; last line - page 1833, left-hand column, line 2; figure 4 page 1836, left-hand column, lines 6-17	1-10
X	WO 2014/093479 A1 (UNIV MONTANA STATE [US]) 19 June 2014 (2014-06-19) paragraphs [0019], [0036], [0042] - [0043], [0078]; claim 35	1-10
Y	Y BILL KIM ET AL: "Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions", NATURE BIOTECHNOLOGY, vol. 35, no. 4, 13 February 2017 (2017-02-13), pages 371-376, XP055484491, US ISSN: 1087-0156, DOI: 10.1038/nbt.3803 figure 1a	5,6
X,P	WO 2017/160689 A1 (UNIV MASSACHUSETTS [US]) 21 September 2017 (2017-09-21)	1-4,6-8, 10
Y,P	page 27, lines 7-14 page 41, line 25 - page 42, line 6 claims 11-49	1-10
X,P	HYNES ALEXANDER P ET AL: "An anti-CRISPR from a virulent streptococcal phage inhibitsStreptococcus pyogenesCas9", NATURE MICROBIOLOGY, NATURE PUBLISHING GROUP UK, LONDON, vol. 2, no. 10, 7 August 2017 (2017-08-07) , pages 1374-1380, XP036429527, DOI: 10.1038/S41564-017-0004-7	1-10
Y,P	the whole document	1-10

-/--

1

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,P	MOE HIROSAWA ET AL: "Cell-type-specific genome editing with a microRNA-responsive CRISPR-Cas9 switch", NUCLEIC ACIDS RESEARCH, vol. 45, no. 13, 19 May 2017 (2017-05-19), pages e118-e118, XP055444821, ISSN: 0305-1048, DOI: 10.1093/nar/gkx309 the whole document	1-10
A	JAMES K NUÑEZ ET AL: "Chemical and Biophysical Modulation of Cas9 for Tunable Genome Engineering", ACS CHEMICAL BIOLOGY, ACS PUBLICATIONS, USA, vol. 11, no. 3, 18 March 2016 (2016-03-18) , pages 681-688, XP002758943, ISSN: 1554-8937, DOI: 10.1021/ACSCHEMBIO.5B01019 [retrieved on 2016-02-09] the whole document	1-10

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2018/060510

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2014093479 A1	19-06-2014	NONE	
WO 2017160689 A1	21-09-2017	NONE	

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-10

   A method for modulating the activity of a Cas endonuclease with a target polynucleotide in a cell, comprising providing an anti-CRISPR (ACR) polypeptide to the cell, wherein the anti-CRISPR polypeptide modulates the activity of the Cas endonuclease in the cell.

   ---

2. claims: 11-33

   A method of increasing the specificity of a Cas endonuclease and guide polynucleotide complex in a cell, comprising introducing an anti-CRISPR (ACR) polypeptide to the cell, wherein the ACR polypeptide interacts with the Cas endonuclease.

   ---

3. claims: 34-53

   A method of increasing site-specific homologous recombination frequency of a donor polynucleotide in a cell, comprising introducing to the cell an anti-CRISPR (ACR) polypeptide to increase the homologous recombination of the donor polynucleotide by a polynucleotide-guided Cas endonuclease.

   ---

4. claims: 54-59

   A cell comprising a Cas endonuclease and an anti-CRISPR (ACR) protein.

   ---

5. claims: 60-65

   A cell comprising a recombinant construct comprising a polynucleotide sequence encoding an anti-CRISPR (ACR) protein, operably linked to a heterologous regulatory expression element.

   ---

6. claim: 66

   A method for gene activation or gene repression in a cell, comprising introducing an anti-CRISPR (ACR) polypeptide and a Cas endonuclease into the cell comprising the gene to be activated or repressed.
