



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

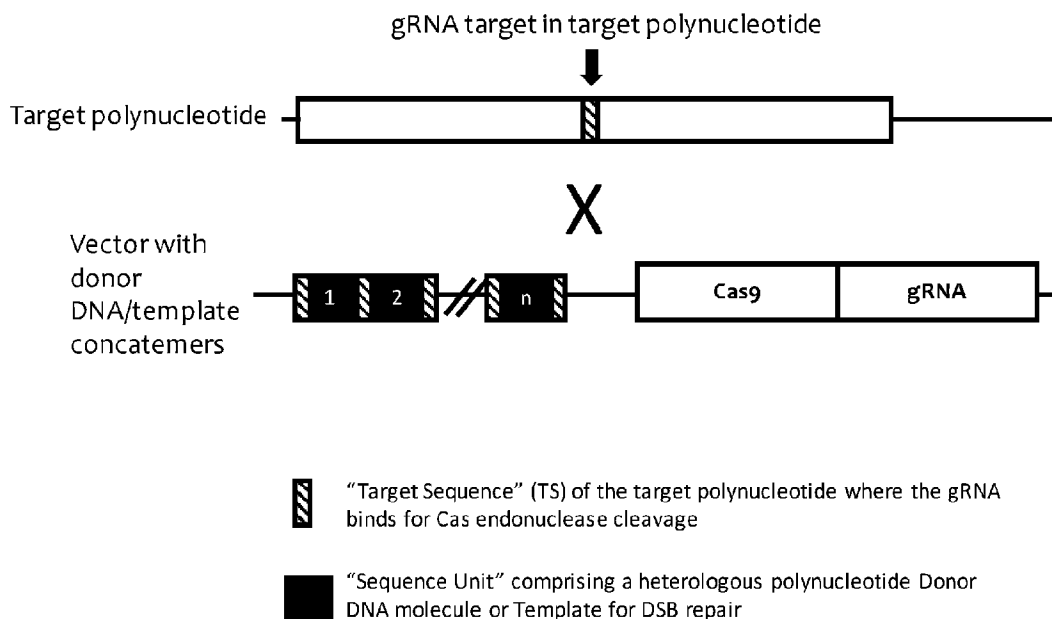
(13) **A1**

(86) Date de dépôt PCT/PCT Filing Date: 2020/07/17
(87) Date publication PCT/PCT Publication Date: 2021/01/28
(85) Entrée phase nationale/National Entry: 2021/12/08
(86) N° demande PCT/PCT Application No.: US 2020/042579
(87) N° publication PCT/PCT Publication No.: 2021/016098
(30) Priorité/Priority: 2019/07/23 (US62/877,359)

(51) Cl.Int./Int.Cl. *C12N 9/22* (2006.01),
C12N 15/10 (2006.01), *C12N 15/82* (2006.01)
(71) Demandeur/Applicant:
PIONEER HI-BRED INTERNATIONAL, INC., US
(72) Inventeurs/Inventors:
DAS, AMIT, US;
KANUGANTI, SATISH, US;
MOHANTY, AMITABH, IN;
SVITASHEV, SERGEI, US;
TAVVA, VENKATA S, US
(74) Agent: TORYS LLP

(54) Titre : STRATEGIE DE CONCEPTION DE DONNEUR POUR L'EDITION DE GENOME CRISPR-CAS9
(54) Title: DONOR DESIGN STRATEGY FOR CRISPR-CAS9 GENOME EDITING

FIG. 1



(57) **Abrégé/Abstract:**

Methods and compositions are provided for the improvement of homology-directed repair of a double strand break in a plant cell, using concatemers of heterologous polynucleotides that are flanked by sequences capable of sequence hybridization with a guide RNA. In some aspects, the double strand break is created by an RNA-guided Cas endonuclease. The homology-directed repair of the double-strand break may include incorporation of a heterologous polynucleotide, for example a gene encoding a trait of agronomic importance. The homology-directed repair of the double-strand break may occur as a result of template-directed repair using a heterologous polynucleotide as a repair template.

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

(19) World Intellectual Property

Organization

International Bureau

(43) International Publication Date

28 January 2021 (28.01.2021)



(10) International Publication Number

WO 2021/016098 A1

(51) International Patent Classification:

C12N 9/22 (2006.01)

C12N 15/82 (2006.01)

C12N 15/10 (2006.01)

(21) International Application Number:

PCT/US2020/042579

(22) International Filing Date:

17 July 2020 (17.07.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/877,359

23 July 2019 (23.07.2019)

US

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

(72) Inventors: **DAS, Amit**; 7250 NW 62nd Avenue, PO Box 552, Johnston, Iowa 50131-1014 (US). **KANUGANTI,**

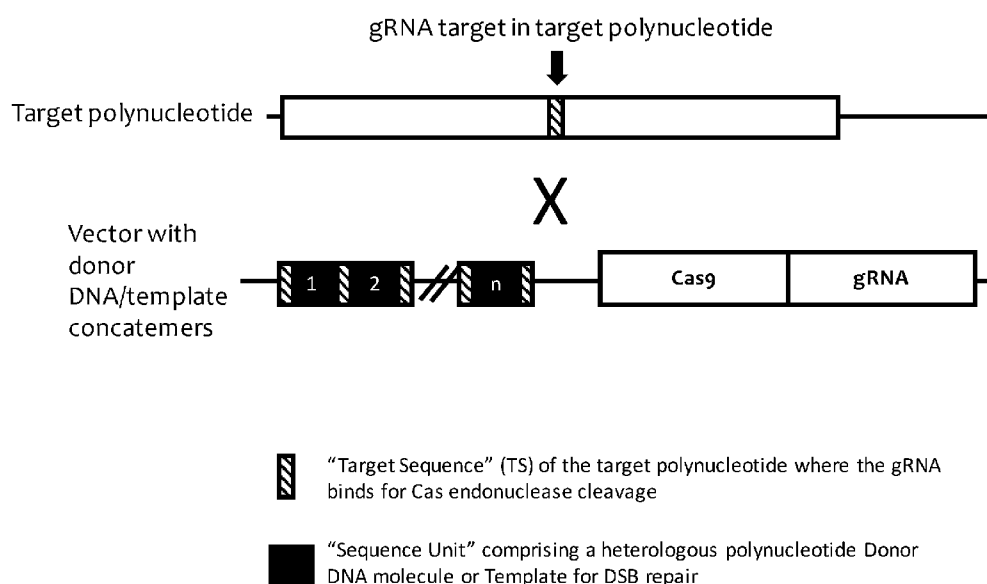
Satish; 7250 NW 62nd Avenue, PO Box 552, Johnston, Iowa 50131-1014 (US). **MOHANTY, Amitabh**; Greenwood Residency, Apt C-217 Hislop Road, Kowkur, Hyderabad TS 500010 (IN). **SVITASHEV, Sergei**; 7250 NW 62nd Avenue, PO Box 552, Johnston, Iowa 50131-1014 (US). **TAVVA, Venkata S**; 7250 NW 62nd Avenue, PO Box 552, Johnston, Iowa 50131-1014 (US).

(74) Agent: **VOSNIDOU, Nancy C.**; 7250 NW 62nd Avenue, PO Box 552, Johnston, Iowa 50131-1014 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, 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,

(54) Title: DONOR DESIGN STRATEGY FOR CRISPR-CAS9 GENOME EDITING

FIG. 1



(57) Abstract: Methods and compositions are provided for the improvement of homology-directed repair of a double strand break in a plant cell, using concatemers of heterologous polynucleotides that are flanked by sequences capable of sequence hybridization with a guide RNA. In some aspects, the double strand break is created by an RNA-guided Cas endonuclease. The homology-directed repair of the double-strand break may include incorporation of a heterologous polynucleotide, for example a gene encoding a trait of agronomic importance. The homology-directed repair of the double-strand break may occur as a result of template-directed repair using a heterologous polynucleotide as a repair template.

[Continued on next page]



WO 2021/016098 A1

WO 2021/016098 A1

SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, 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).

Declarations under Rule 4.17:

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

Published:

- *with international search report (Art. 21(3))*
- *with sequence listing part of description (Rule 5.2(a))*

DONOR DESIGN STRATEGY FOR CRISPR-CAS9 GENOME EDITING**CROSS-REFERENCE TO RELATED APPLICATIONS**

[0001] This application claims the benefit of U.S. Provisional Patent Application Serial No. 62/877,359, filed on 23 July 2019, all of which is incorporated herein by reference in its entirety.

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 7824WOPCT_SequenceListing_ST25.txt created on 13 July 2020 and having a size of 41,654 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 for modifying the genome of a cell.

BACKGROUND

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

[0005] Newer technologies utilizing archaeal or bacterial adaptive immunity systems have been identified, called CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats), which comprise different domains of effector proteins that encompass a variety of

activities (DNA recognition, binding, cleavage, and repair). Repair of a double-strand break can be via Non-Homologous End Joining (NHEJ) or Homology-Directed Repair (HDR)/Homologous Recombination (HR). HDR/HR may be accomplished by several mechanisms, including homologous recombination at a target site, which may further include introduction of a template for template-directed repair, or the introduction of a DNA molecule for targeted integration.

[0006] There remains a need for methods and compositions for the improving the frequency of homology-directed repair of double-strand-break sites.

SUMMARY OF INVENTION

[0007] Provided are methods for repairing a double-strand break in a target polynucleotide, and increasing the frequency of homology directed repair or homologous recombination.

[0008] In one aspect, the method comprises providing to a target polynucleotide a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in the target polynucleotide, and a plurality of sequence units, each sequence unit comprising a heterologous polynucleotide, wherein each sequence unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides, and identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units.

[0009] In one aspect, the method comprises providing to a target polynucleotide a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in the target polynucleotide, and a plurality of sequence units, each sequence unit comprising a heterologous polynucleotide, wherein each sequence unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides, and identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding,

as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units, wherein the heterologous polynucleotide is a donor DNA molecule that is inserted into the double-strand break.

[0010] In one aspect, the method comprises providing to a target polynucleotide a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in the target polynucleotide, and a plurality of sequence units, each sequence unit comprising a heterologous polynucleotide, wherein each sequence unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides, and identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units, wherein the heterologous polynucleotide is a template DNA molecule that directs the repair of the double-strand break.

[0011] In one aspect, the method comprises providing to a target polynucleotide a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in the target polynucleotide, and a plurality of sequence units, each sequence unit comprising a heterologous polynucleotide, wherein each sequence unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides, and identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units, wherein each heterologous polynucleotide is flanked by a set of second flanking sequences, each of which is 1, between 1 and 5, 5, between 5 and 10, 10, between 10 and 15, 15, between 15 and 20, 20, between 20 and 25, or greater than 25 nucleotides in length and share at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% identity to a sequence within 50, 75, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2500, 3000, 3500, 4000, 4500, 5000, or greater than 5000 nucleotides of the double-strand break,

and wherein said set of second flanking sequences is flanked by the set of first flanking sequences.

[0012] In one aspect, the method comprises providing to a target polynucleotide a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in the target polynucleotide, and a plurality of sequence units, each sequence unit comprising a heterologous polynucleotide, wherein each sequence unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides, and identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units, wherein a plurality of different guide RNA molecules are provided, and wherein the second flanking sequences are capable of hybridizing to the plurality of different guide RNA molecules.

[0013] In one aspect, the method comprises providing to a target polynucleotide a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in the target polynucleotide, and a plurality of sequence units, each sequence unit comprising a heterologous polynucleotide, wherein each sequence unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides, and identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units, wherein the target polynucleotide is in a cell.

[0014] In one aspect, the method comprises providing to a target polynucleotide a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in the target polynucleotide, and a plurality of sequence units, each sequence unit comprising a heterologous polynucleotide, wherein each sequence unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides, and identifying at least one nucleotide insertion, deletion, substitution, or

modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units, wherein the target polynucleotide is in a cell, wherein the cell is a plant cell.

[0015] In one aspect, the method comprises providing to a target polynucleotide a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in the target polynucleotide, and a plurality of sequence units, each sequence unit comprising a heterologous polynucleotide, wherein each sequence unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides, and identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units, wherein the plurality of sequence units is stably integrated into the plant cell.

[0016] In one aspect, the method comprises providing to a target polynucleotide a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in the target polynucleotide, and a plurality of sequence units, each sequence unit comprising a heterologous polynucleotide, wherein each sequence unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides, and identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units, wherein the guide RNA molecule is provided via particle bombardment.

[0017] In one aspect, the method comprises providing to a target polynucleotide a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in the target polynucleotide, and a plurality of sequence units, each sequence unit comprising a heterologous polynucleotide, wherein each sequence unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides, and identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding,

as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units, wherein the plurality of sequence units is provided by *Agrobacterium*-mediated transformation.

[0018] In one aspect, the method comprises providing to a target polynucleotide a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in the target polynucleotide, and a plurality of sequence units, each sequence unit comprising a heterologous polynucleotide, wherein each sequence unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides, and identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units, wherein the plurality of sequence units is provided by particle bombardment.

[0019] In one aspect, the method comprises providing to a target polynucleotide a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in the target polynucleotide, and a plurality of sequence units, each sequence unit comprising a heterologous polynucleotide, wherein each sequence unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides, and identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units, wherein the frequency of homologous recombination repair at the double-strand-break site of the target polynucleotide is greater than the rate of non-homologous end joining repair at that same site.

[0020] In one aspect, the method provides a method of altering a phenotypic trait in a plant, comprising providing to a plant cell a set of molecules comprising a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in a target polynucleotide in the plant cell, and a plurality of sequence units, each comprising a heterologous polynucleotide, wherein each unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality

of sequence units with the complex to release the heterologous polynucleotides; identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units; and obtaining a plant from the plant cell, wherein the plant comprises an alteration of at least one phenotypic trait as compared to an isoline plant that was not provided the set of molecules.

[0021] In one aspect, the method provides a method of altering a phenotypic trait in a plant, comprising providing to a plant cell a set of molecules comprising a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in a target polynucleotide in the plant cell, and a plurality of sequence units, each comprising a heterologous polynucleotide, wherein each unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides; identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units; and obtaining a plant from the plant cell, wherein the plant comprises an alteration of at least one phenotypic trait as compared to an isoline plant that was not provided the set of molecules; wherein each heterologous polynucleotide is flanked by a set of second flanking sequences, each of which is 1, between 1 and 5, 5, between 5 and 10, 10, between 10 and 15, 15, between 15 and 20, 20, between 20 and 25, or greater than 25 nucleotides in length and share at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater than 99% identity to a sequence within 50, 75, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2500, 3000, 3500, 4000, 4500, 5000, or greater than 5000 nucleotides of the double-strand break, and wherein said set of second flanking sequences is flanked by the set of first flanking sequences.

[0022] In one aspect, the method provides a method of altering a phenotypic trait in a plant, comprising providing to a plant cell a set of molecules comprising a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in a target polynucleotide in the plant cell, and a plurality of sequence units, each comprising a heterologous polynucleotide, wherein each unit is flanked by a set of first flanking

sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides; identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units; and obtaining a plant from the plant cell, wherein the plant comprises an alteration of at least one phenotypic trait as compared to an isoline plant that was not provided the set of molecules; wherein the cell is a monocot plant cell.

[0023] In one aspect, the method provides a method of altering a phenotypic trait in a plant, comprising providing to a plant cell a set of molecules comprising a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in a target polynucleotide in the plant cell, and a plurality of sequence units, each comprising a heterologous polynucleotide, wherein each unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides; identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units; and obtaining a plant from the plant cell, wherein the plant comprises an alteration of at least one phenotypic trait as compared to an isoline plant that was not provided the set of molecules; wherein the cell is a monocot plant cell, wherein the monocot plant cell is selected from the group consisting of: maize, rice, sorghum, barley, and wheat.

[0024] In one aspect, the method provides a method of altering a phenotypic trait in a plant, comprising providing to a plant cell a set of molecules comprising a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in a target polynucleotide in the plant cell, and a plurality of sequence units, each comprising a heterologous polynucleotide, wherein each unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides; identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target

polynucleotide prior to the providing of the plurality of sequence units; and obtaining a plant from the plant cell, wherein the plant comprises an alteration of at least one phenotypic trait as compared to an isoline plant that was not provided the set of molecules; wherein the cell is a dicot plant cell.

[0025] In one aspect, the method provides a method of altering a phenotypic trait in a plant, comprising providing to a plant cell a set of molecules comprising a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in a target polynucleotide in the plant cell, and a plurality of sequence units, each comprising a heterologous polynucleotide, wherein each unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides; identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units; and obtaining a plant from the plant cell, wherein the plant comprises an alteration of at least one phenotypic trait as compared to an isoline plant that was not provided the set of molecules; wherein the cell is a dicot plant cell, wherein the dicot plant cell is selected from the group consisting of: soy, canola, cotton, sugarcane, and *Arabidopsis*.

[0026] In one aspect, the method provides a method of altering a phenotypic trait in a plant, comprising providing to a plant cell a set of molecules comprising a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in a target polynucleotide in the plant cell, and a plurality of sequence units, each comprising a heterologous polynucleotide, wherein each unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides; identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units; and obtaining a plant from the plant cell, wherein the plant comprises an alteration of at least one phenotypic trait as compared to an isoline plant that was not provided the set of molecules, wherein the phenotypic trait is average yield.

[0027] In one aspect, the method provides a method of altering a phenotypic trait in a plant, comprising providing to a plant cell a set of molecules comprising a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in a target polynucleotide in the plant cell, and a plurality of sequence units, each comprising a heterologous polynucleotide, wherein each unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides; identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units; and obtaining a plant from the plant cell, wherein the plant comprises an alteration of at least one phenotypic trait as compared to an isoline plant that was not provided the set of molecules; further comprising obtaining a tissue, part, or reproductive element of the plant, wherein the tissue, part, or reproductive element comprises the at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide of the plant from which it was obtained.

[0028] In one aspect, the method provides a progeny plant obtained or derived by the method of altering a phenotypic trait in a plant, comprising providing to a plant cell a set of molecules comprising a Cas endonuclease, a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in a target polynucleotide in the plant cell, and a plurality of sequence units, each comprising a heterologous polynucleotide, wherein each unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA; cleaving the plurality of sequence units with the complex to release the heterologous polynucleotides; identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units; and obtaining a plant from the plant cell, wherein the plant comprises an alteration of at least one phenotypic trait as compared to an isoline plant that was not provided the set of molecules; further comprising obtaining a tissue, part, or reproductive element of the plant, wherein the tissue, part, or reproductive element comprises the at least one nucleotide insertion, deletion, substitution, modification, or combination of any of the preceding, of the

sequence of the target polynucleotide of the plant from which it was obtained; wherein the progeny plant comprises said nucleotide insertion, deletion, substitution, modification, or combination thereof.

BRIEF DESCRIPTION OF THE DRAWINGS AND THE SEQUENCE LISTING

[0029] 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.

[0030] **FIG. 1** depicts the concept of a concatemer donor DNA/template DNA, wherein each unit of DNA is flanked by a guide RNA target sequence.

[0031] **FIG. 2** depicts a schematic of a vector construction for a general single-guide approach.

[0032] **FIG. 3** depicts a schematic of a vector construction for a general dual-guide approach.

[0033] **FIG. 4** is the vector diagram for Construct #1 used in the rice GL3 concatemer experiments (tandem copies of the donor/template), used to transform Rice.

[0034] **FIG. 5** is the vector diagram for Construct #2 used in the rice GL3 concatemer single copy repair template control experiment (single donor/template 1 target site) used to transform Rice.

[0035] **FIG. 6** is the vector diagram for Construct #3 used in the rice GL3 concatemer single copy repair template control experiment (single donor/template 2 target sites) used to transform Rice.

[0036] **FIG. 7A** depicts a schematic of a vector construction for using a single donor/template copy with flanking target site sequences, using PAM sequences that are of inverted orientation.

[0037] **FIG. 7B** depicts a schematic of a vector construction for using a single donor/template copy with flanking target site sequences, using PAM sequences that are in the vector in the same orientation.

[0038] **FIG. 7C** depicts a schematic of a vector construction for using a concatemerized donor/template copy with flanking target site sequences, using PAM sequences that are in the vector in the same orientation.

[0039] **FIG. 8** is the vector diagram for Construct #4 used in the control experiment (single donor/template, according to FIG. 7A) used to transform Maize.

[0040] **FIG. 9** is the vector diagram for Construct #5 used in the control experiment (single donor/template, according to FIG. 7B) used to transform Maize.

[0041] **FIG. 10** is the vector diagram for Construct #5 used in the control experiment (concatemer donor/template, according to FIG. 7C) used to transform Maize.

[0042] **FIG. 11A** depicts a schematic of a vector construction for using a single donor/template copy of 200 nt, with flanking target site sequences.

[0043] **FIG. 11B** depicts a schematic of a vector construction for using a concatemerized donor/template copy, each of 200 nt with flanking target site sequences.

[0044] **FIG. 11C** depicts a schematic of a vector construction for using a single donor/template copy of 828 nt, with flanking target site sequences.

[0045] 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 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.

[0046] **SEQID NO:1** is the *Oryza sativa* DNA sequence for the Rice U3 PolIII Chr4 promoter (OSU3POLIII CHR4 PRO).

[0047] **SEQID NO:2** is the *Oryza sativa* DNA sequence for the Rice GL3 gRNA target site sequence (CR1).

[0048] **SEQID NO:3** is the Artificial DNA sequence for the Guide RNA sequence.

[0049] **SEQID NO:4** is the *Zea mays* DNA sequence for the Maize Ubiquitin promoter (UBI1ZM PRO).

[0050] **SEQID NO:5** is the *Zea mays* DNA sequence for the Maize Ubiquitin 5'UTR (UBI1ZM 5UTR).

[0051] **SEQID NO:6** is the *Zea mays* DNA sequence for the Maize Ubiquitin Intron 1 (UBI1ZM INTRON1).

[0052] **SEQID NO:7** is the *Streptococcus pyogenes* DNA sequence for the *S. pyogenes* Cas9 CDS.

[0053] **SEQID NO:8** is the DNA sequence for the PinII terminator.

- [0054] **SEQID NO:9** is the *Oryza sativa* DNA sequence for the Rice GL3 Homology Region 1(HDR-OS-GL3 FRAG2).
- [0055] **SEQID NO:10** is the *Oryza sativa* DNA sequence for the Rice GL3 template.
- [0056] **SEQID NO:11** is the *Oryza sativa* DNA sequence for the Rice GL3 Homology Region 2(HDR-OS-GL3 FRAG3).
- [0057] **SEQID NO:12** is the Cauliflower mosaic virus DNA sequence for the 35S promoter (CAMV35S PRO-V4).
- [0058] **SEQID NO:13** is the Artificial DNA sequence for the HYG-Z5 Yellow N1 selectable marker.
- [0059] **SEQID NO:14** is the Artificial DNA sequence for the Primers used to clone gRNA in to Rice U3 at Slot1 AarI site.
- [0060] **SEQID NO:15** is the Artificial DNA sequence for the Primers used to clone gRNA in to Rice U3 at Slot1 AarI site.
- [0061] **SEQID NO:16** is the Artificial DNA sequence for the BamHI Linker.
- [0062] **SEQID NO:17** is the Artificial DNA sequence for the HindIII Linker.
- [0063] **SEQID NO:18** is the Artificial DNA sequence for the Primer for cloning of the heterologous polynucleotide at NcoI.
- [0064] **SEQID NO:19** is the Artificial DNA sequence for the Primer for cloning of the heterologous polynucleotide at NcoI.
- [0065] **SEQID NO:20** is the Artificial DNA sequence for the Primer for cloning of the heterologous polynucleotide at BamHI.
- [0066] **SEQID NO:21** is the Artificial DNA sequence for the Primer for cloning of the heterologous polynucleotide at BamHI.
- [0067] **SEQID NO:22** is the Artificial DNA sequence for the Primer for cloning of the heterologous polynucleotide at BglII.
- [0068] **SEQID NO:23** is the Artificial DNA sequence for the Primer for cloning of the heterologous polynucleotide at BglII.
- [0069] **SEQID NO:24** is the Artificial DNA sequence for the Primer for cloning of the heterologous polynucleotide at HindIII.
- [0070] **SEQID NO:25** is the Artificial DNA sequence for the Primer for cloning of the heterologous polynucleotide at HindIII.

- [0071] **SEQID NO:26** is the Artificial DNA sequence for the Primer for cloning of the heterologous polynucleotide at StuI.
- [0072] **SEQID NO:27** is the Artificial DNA sequence for the Primer for cloning of the heterologous polynucleotide at StuI.
- [0073] **SEQID NO:28** is the Artificial DNA sequence for the Primer for cloning of the heterologous polynucleotide at StuI.
- [0074] **SEQID NO:29** is the Artificial DNA sequence for the Primer for cloning of the heterologous polynucleotide at StuI.
- [0075] **SEQID NO:30** is the Zea mays DNA sequence for the AXIG1 promoter.
- [0076] **SEQID NO:31** is the Zea mays DNA sequence for the WUS2 morphogenic factor (ALT1).
- [0077] **SEQID NO:32** is the Agrobacterium tumefaciens DNA sequence for the NOS terminator.
- [0078] **SEQID NO:33** is the Zea mays DNA sequence for the PLTP promoter.
- [0079] **SEQID NO:34** is the Zea mays DNA sequence for the PLTP 5' UTR.
- [0080] **SEQID NO:35** is the Zea mays DNA sequence for the ODP2 morphogenic factor (ALT1).
- [0081] **SEQID NO:36** is the Oryza sativa DNA sequence for the T28 terminator.
- [0082] **SEQID NO:37** is the Zea mays DNA sequence for the Ubiquitin promoter.
- [0083] **SEQID NO:38** is the Zea mays DNA sequence for the Ubiquitin 5' UTR.
- [0084] **SEQID NO:39** is the Zea mays DNA sequence for the Ubiquitin intron 1.
- [0085] **SEQID NO:40** is the Simian virus 40 DNA sequence for the T-antigen monopartite nuclear localization signal .
- [0086] **SEQID NO:41** is the Streptococcus pyogenes DNA sequence for the Cas9 Exon 1.
- [0087] **SEQID NO:42** is the Solanum tuberosum DNA sequence for the LS1 Intron2.
- [0088] **SEQID NO:43** is the Streptococcus pyogenes DNA sequence for the Cas9 Exon 2.
- [0089] **SEQID NO:44** is the Agrobacterium tumefaciens DNA sequence for the C-terminal bipartite nuclear localization signal from VirD2 endonuclease.
- [0090] **SEQID NO:45** is the Solanum tuberosum DNA sequence for the PinII terminator.
- [0091] **SEQID NO:46** is the Zea mays DNA sequence for the U6 PolIII Chr8 Promoter.

[0092] **SEQID NO:47** is the Zea mays DNA sequence for the DNA sequence encoding the VT domain of the guide RNA for Zm target site.

[0093] **SEQID NO:48** is the Artificial DNA sequence for the Guide RNA sequence.

[0094] **SEQID NO:49** is the Zea mays DNA sequence for the Zm gRNA target site sequence (CR1).

[0095] **SEQID NO:50** is the Zea mays DNA sequence for the Zm Homology Region 1.

[0096] **SEQID NO:51** is the Artificial DNA sequence for the Zm Target Site template.

[0097] **SEQID NO:52** is the Zea mays DNA sequence for the Zm Homology Region 2.

[0098] **SEQID NO:53** is the Escherichia coli DNA sequence for the NPTII selectable marker.

DETAILED DESCRIPTION

[0099] CRISPR/Cas assisted targeted genome editing can result in editing of a gene sequence to generate “allelic edits” by homologous recombination (HR) via template-dependent repair. Some methods involve single template-dependent repairs of target genomic DNA following DNA double strand breaks. However, the success and frequency of such editing strategies involving a donor or repair template can be limited in the genomes of some types of organisms, such as some plants. An increase in efficiency of template dependent repair or editing of the target genome enables high throughput targeted genome editing, and may play a key role in genetic pathway engineering by targeting multiple genes or multiple targets in one gene.

[0100] The frequency of homologous recombination (HR) or homology-directed repair (HDR) of a double-strand break (DSB) created by a CRISPR endonuclease can be increased by providing a concatemer of tandemly-repeated sequences. The tandem repeats are each flanked by spacers and targets for guide RNAs (same or different, for single- or dual- or multiple-guide approaches), that are cleaved by a Cas endonuclease and released for integration into the DSB, or for template-directed repair of the DSB. The frequency of HR/HDR is increased by providing a plurality of repair templates or integratable sequences at the DSB. The concatemer can be delivered by either gene gun or via the *Agrobacterium* method to the target tissue.

[0101] Presented are novel methods and compositions for increasing the frequency of homology-directed repair of a double-strand break, using a concatemer of donor DNA templates.

[0102] A novel construct design was designed to test the efficacy of template-directed DSB repair by a CRISPR/Cas9 system, using of multiple copies of a donor or template in single

construct that were each flanked by CRISPR/Cas9 targets on either side (concatemer). A “concatemer” is defined herein as a plurality of identical polynucleotides that are flanked by sequences that are targets of a Cas endonuclease-guide RNA complex. In some aspects, such flanking sequences are capable of hybridizing with the guide RNA of the complex.

[0103] With the concatemer design, a Cas enzyme generated DNA double strand breaks at each target in between template copies. The free templates, thus generated, were available in multiple copies for the host DNA repair system.

[0104] There were three major components of the vector plasmid that was introduced into the target cell: Cas9, Guide RNA (gRNA), and repair template: multiple (*e.g.*, four) units in tandem, each unit flanked on either side by the target sequence recognized by the same gRNA.

[0105] The rationale of the strategy aimed to generate more repair templates available during editing process inside the cellular nucleus. Cloning of multiple repair templates (4 units) were excisable inside the cells. While a longer circular plasmid harboring all components may be straightforward to transform in some plant tissues by particle bombardment, releasing of shorter free repair templates upon excision of the same gRNA inside the cells is likely to provide greater number of templates and as well as these may be more accessible to the DNA target site.

[0106] One purpose of the concatemer approach is to increase the *in vivo* load of donor/template molecules by having multiple copies of the donor/template in tandem, without eliciting potential silencing effects due to repetitive elements. The advantages of this method include the availability of more than one copy of the donor or template molecule, *in vivo* release of the donor/template by flanking PAM sequences and guides, and edited templates to avoid targeting by Cas after repair/integration.

[0107] Terms used in the claims and specification are defined as set forth below unless otherwise specified. It must be noted that, as used in the specification and the appended claims, the singular forms “a,” “an” and “the” include plural referents unless the context clearly dictates otherwise.

Definitions

[0108] As used herein, “nucleic acid” means a polynucleotide and includes a single or a double-stranded polymer of deoxyribonucleotide or ribonucleotide bases. Nucleic acids may also include fragments and modified nucleotides. Thus, the terms “polynucleotide”, “nucleic acid sequence”, “nucleotide sequence” and “nucleic acid fragment” are used interchangeably to

denote a polymer of RNA and/or DNA and/or RNA-DNA that is single- or double-stranded, optionally comprising synthetic, non-natural, or altered nucleotide bases. Nucleotides (usually found in their 5'-monophosphate form) are referred to by their single letter designation as follows: "A" for adenosine or deoxyadenosine (for RNA or DNA, respectively), "C" for cytosine or deoxycytosine, "G" for guanosine or deoxyguanosine, "U" for uridine, "T" for deoxythymidine, "R" for purines (A or G), "Y" for pyrimidines (C or T), "K" for G or T, "H" for A or C or T, "I" for inosine, and "N" for any nucleotide.

[0109] 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.

[0110] "Open reading frame" is abbreviated ORF.

[0111] The term "selectively hybridizes" includes reference to hybridization, under stringent hybridization conditions, of a nucleic acid sequence to a specified nucleic acid target sequence to a detectably greater degree (e.g., at least 2-fold over background) than its hybridization to non-target nucleic acid sequences and to the substantial exclusion of non-target nucleic acids. Selectively hybridizing sequences typically have about at least 80% sequence identity, or 90% sequence identity, up to and including 100% sequence identity (i.e., fully complementary) with each other.

[0112] The term "stringent conditions" or "stringent hybridization conditions" includes reference to conditions under which a probe will selectively hybridize to its target sequence in an *in vitro* hybridization assay. Stringent conditions are sequence-dependent and will be different in different circumstances. By controlling the stringency of the hybridization and/or washing conditions, target sequences can be identified which are 100% complementary to the probe (homologous probing). Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). Generally, a probe is less than about 1000 nucleotides in length, optionally less than 500 nucleotides in length. Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salt(s)) at pH 7.0 to 8.3, and at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide.

Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C.

[0113] By “homology” is meant DNA sequences that are similar. For example, a “region of homology to a genomic region” that is found on the donor DNA is a region of DNA that has a similar sequence to a given “genomic region” in the cell or organism genome. A region of homology can be of any length that is sufficient to promote homologous recombination at the cleaved target site. For example, the region of homology can comprise at least 5-10, 5-15, 5-20, 5-25, 5-30, 5-35, 5-40, 5-45, 5- 50, 5-55, 5-60, 5-65, 5- 70, 5-75, 5-80, 5-85, 5-90, 5-95, 5-100, 5-200, 5-300, 5-400, 5-500, 5-600, 5-700, 5-800, 5-900, 5-1000, 5-1100, 5-1200, 5-1300, 5-1400, 5-1500, 5-1600, 5-1700, 5-1800, 5-1900, 5-2000, 5-2100, 5-2200, 5-2300, 5-2400, 5-2500, 5-2600, 5-2700, 5-2800, 5-2900, 5-3000, 5-3100 or more bases in length such that the region of homology has sufficient homology to undergo homologous recombination with the corresponding genomic region. “Sufficient homology” indicates that two polynucleotide sequences have structural similarity such that they are capable of acting as substrates for a homologous recombination reaction. The structural similarity includes overall length of each polynucleotide fragment, as well as the sequence similarity of the polynucleotides. Sequence similarity can be described by the percent sequence identity over the whole length of the sequences, and/or by conserved regions comprising localized similarities such as contiguous nucleotides having 100% sequence identity, and percent sequence identity over a portion of the length of the sequences.

[0114] As used herein, a “genomic region” is a segment of a chromosome in the genome of a cell that is present on either side of the target site or, alternatively, also comprises a portion of the target site. The genomic region can comprise at least 5-10, 5-15, 5-20, 5-25, 5-30, 5-35, 5-40, 5-45, 5- 50, 5-55, 5-60, 5-65, 5- 70, 5-75, 5-80, 5-85, 5-90, 5-95, 5-100, 5-200, 5-300, 5-400, 5-500, 5-600, 5-700, 5-800, 5-900, 5-1000, 5-1100, 5-1200, 5-1300, 5-1400, 5-1500, 5-1600, 5-1700, 5-1800, 5-1900, 5-2000, 5-2100, 5-2200, 5-2300, 5-2400, 5-2500, 5-2600, 5-2700, 5-2800.

5-2900, 5-3000, 5-3100 or more bases such that the genomic region has sufficient homology to undergo homologous recombination with the corresponding region of homology.

[0115] 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.

[0116] “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.

[0117] The term “percentage of sequence identity” refers to the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the results by 100 to yield the percentage of sequence identity. Useful examples of percent sequence identities include, but are not limited to, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%, or any percentage from 50% to 100%. These identities can be determined using any of the programs described herein.

[0118] 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.

[0119] The “Clustal V method of alignment” corresponds to the alignment method labeled Clustal V (described by Higgins and Sharp, (1989) *CABIOS* 5:151-153; Higgins *et al.*, (1992) *Comput Appl Biosci* 8:189-191) and found in the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). For multiple alignments, the default values correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10. Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal method are KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids these parameters are KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. After alignment of the sequences using the Clustal V program, it is possible to obtain a “percent identity” by viewing the “sequence distances” table in the same program. The “Clustal W method of alignment” corresponds to the alignment method labeled Clustal W (described by Higgins and Sharp, (1989) *CABIOS* 5:151-153; Higgins *et al.*, (1992) *Comput Appl Biosci* 8:189-191) and found in the MegAlign™ v6.1 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Default parameters for multiple alignment (GAP PENALTY=10, GAP LENGTH PENALTY=0.2, Delay Divergen Seqs (%)=30, DNA Transition Weight=0.5, Protein Weight Matrix=Gonnet Series, DNA Weight Matrix=IUB). After alignment of the sequences using the Clustal W program, it is possible to obtain a “percent identity” by viewing the “sequence distances” table in the same program. Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using GAP Version 10 (GCG, Accelrys, San Diego, CA) using the following parameters: % identity and % similarity for a nucleotide sequence using a gap creation penalty weight of 50 and a gap length extension penalty weight of 3, and the nwsgapdna.cmp scoring matrix; % identity and % similarity for an

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

[0120] Polynucleotide and polypeptide sequences, variants thereof, and the structural relationships of these sequences can be described by the terms “homology”, “homologous”, “substantially identical”, “substantially similar” and “corresponding substantially” which are used interchangeably herein. These refer to polypeptide or nucleic acid sequences wherein changes in one or more amino acids or nucleotide bases do not affect the function of the molecule, such as the ability to mediate gene expression or to produce a certain phenotype. These terms also refer to modification(s) of nucleic acid sequences that do not substantially alter the functional properties of the resulting nucleic acid relative to the initial, unmodified nucleic acid. These modifications include deletion, substitution, and/or insertion of one or more

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

[0121] 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.

[0122] 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.

[0123] The term "fragment" refers to a contiguous set of nucleotides or amino acids. 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 nucleotides. 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 amino acids. A fragment may or may not exhibit the function of a sequence sharing some percent identity over the length of said fragment.

[0124] The terms “fragment that is functionally equivalent” and “functionally equivalent fragment” are used interchangeably herein. These terms refer to a portion or subsequence of an isolated nucleic acid fragment or polypeptide that displays the same activity or function as the longer sequence from which it derives. In one example, the fragment retains the ability to alter gene expression or produce a certain phenotype whether or not the fragment encodes an active protein. For example, the fragment can be used in the design of genes to produce the desired phenotype in a modified plant. Genes can be designed for use in suppression by linking a nucleic acid fragment, whether or not it encodes an active enzyme, in the sense or antisense orientation relative to a plant promoter sequence.

[0125] “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.

[0126] 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 a cell; that is, not heterologous.

[0127] 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.

[0128] “Coding sequence” refers to a polynucleotide sequence which codes for a specific amino acid sequence. “Regulatory sequences” refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences include, but are not limited to, promoters, translation leader sequences, 5' untranslated sequences, 3' untranslated sequences, introns,

polyadenylation target sequences, RNA processing sites, effector binding sites, and stem-loop structures.

[0129] 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.

[0130] 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.

[0131] 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).

[0132] 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.

[0133] By “domain” it is meant a contiguous stretch of nucleotides (that can be RNA, DNA, and/or RNA-DNA-combination sequence) or amino acids.

[0134] 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.

[0135] 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.

[0136] An “optimized” polynucleotide is a sequence that has been optimized for improved expression in a particular heterologous host cell.

[0137] A “plant-optimized nucleotide sequence” is a nucleotide sequence that has been optimized for expression in plants, particularly for increased expression in plants. A plant-optimized nucleotide sequence includes a codon-optimized gene. A plant-optimized nucleotide sequence can be synthesized by modifying a nucleotide sequence encoding a protein such as, for example, a Cas 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.

[0138] 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.

[0139] 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.

[0140] “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).

[0141] “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.

[0142] “RNA transcript” refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or pre-mRNA. A RNA transcript is referred to as the mature RNA or mRNA when it is a RNA sequence derived from post-transcriptional processing of the primary transcript pre-mRNA. “Messenger RNA” or “mRNA” refers to the RNA that is without introns and that can be translated into protein by the cell. “cDNA” refers to a DNA that is complementary to, and synthesized from, an mRNA template using the enzyme reverse transcriptase. The cDNA can be single-stranded or converted into double-stranded form using the Klenow fragment of DNA polymerase I. “Sense” RNA refers to RNA transcript that includes the mRNA and can be translated into protein within a cell or *in vitro*. “Antisense RNA” refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA, and that blocks the expression of a target gene (see, e.g., U.S. Patent No. 5,107,065). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5’ non-coding sequence, 3’ non-coding sequence, introns,

or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes. The terms "complement" and "reverse complement" are used interchangeably herein with respect to mRNA transcripts, and are meant to define the antisense RNA of the message.

[0143] 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.

[0144] 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.

[0145] 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, an 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*.

[0146] 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.

[0147] 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. In one aspect, a “Donor DNA cassette” comprises a heterologous polynucleotide to be inserted at the double-strand break site created by a double-strand-break inducing agent (*e.g.* a Cas endonuclease and guide RNA complex), that is operably linked to a noncoding expression regulatory element. In some aspects, the Donor DNA cassette further comprises polynucleotide sequences that are homologous to the target site, that flank the polynucleotide of interest operably linked to a noncoding expression regulatory element.

[0148] 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.

[0149] 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.

[0150] 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.

[0151] 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).

[0152] “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.

[0153] “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.

[0154] As used herein, an “effector” or “effector protein” is a protein that encompasses an activity including recognizing, binding to, and/or cleaving or nicking a polynucleotide target. An effector, or effector protein, may also be an endonuclease. The “effector complex” of a CRISPR system includes Cas proteins involved in crRNA and target recognition and binding. Some of the component Cas proteins may additionally comprise domains involved in target polynucleotide cleavage.

[0155] The term “Cas protein” refers to a polypeptide encoded by a Cas (CRISPR-associated) gene. A Cas protein includes but is not limited to: 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” or “Cas effector protein”, that when in complex with a suitable polynucleotide component, is capable of recognizing, binding to, and optionally nicking or cleaving all or part of a specific polynucleotide target sequence. A Cas endonuclease described herein comprises one or more nuclease domains. The endonucleases of the disclosure may include those having one or more RuvC 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% and 85%, at least 85%, between 85% and 90%, at least 90%, between 90% and 95%, at least 95%, between 95% and 96%, at least 96%, between 96% and 97%, at least 97%, between 97% and 98%, at least 98%, between 98% and 99%, at least 99%, between 99% and 100%, or 100% sequence identity with at least 50, between 50 and 100, at least 100, between 100 and 150, at least 150, between 150 and 200, at least 200, between 200 and 250, at least 250, between 250 and 300, at least 300, between 300 and 350, at least 350, between 350 and 400, at least 400, between 400 and 450, at least 500, or greater than 500 contiguous amino acids of a native Cas protein, and retains at least partial activity.

[0156] A “Cas endonuclease” may comprise domains that enable it to function as a double-strand-break-inducing agent. A “Cas endonuclease” may also comprise one or more modifications or mutations that abolish or reduce its ability to cleave a double-strand polynucleotide (dCas). In some aspects, the Cas endonuclease molecule may retain the ability to nick a single-strand polynucleotide (for example, a D10A mutation in a Cas9 endonuclease molecule) (nCas9).

[0157] 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 of functional part of a Cas3 HD domain, a complete of functional part of a Cas3 Helicase domain, complete of functional part of a Cascade protein (such as but not limiting to a Cas5, Cas5d, Cas7 and Cas8b1).

[0158] The terms “functional variant”, “variant that is functionally equivalent” and “functionally equivalent variant” of a Cas endonuclease or Cas effector protein are used interchangeably herein, and refer to a variant of the Cas effector protein disclosed herein in which the ability to recognize, bind to, and optionally unwind, nick or cleave all or part of a target sequence is retained.

[0159] 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.

[0160] The terms “cascade” and “cascade complex” are used interchangeably herein and include reference to a multi-subunit protein complex that can assemble with a polynucleotide

forming a polynucleotide-protein complex (PNP). Cascade is a PNP that relies on the polynucleotide for complex assembly and stability, and for the identification of target nucleic acid sequences. Cascade functions as a surveillance complex that finds and optionally binds target nucleic acids that are complementary to a variable targeting domain of the guide polynucleotide.

[0161] The terms "cleavage-ready Cascade", "crCascade", "cleavage-ready Cascade complex", "crCascade complex", "cleavage-ready Cascade system", "CRC" and "crCascade system", are used interchangeably herein and include reference to a multi-subunit protein complex that can assemble with a polynucleotide forming a polynucleotide-protein complex (PNP), wherein one of the cascade proteins is a Cas endonuclease capable of recognizing, binding to, and optionally unwinding, nicking, or cleaving all or part of a target sequence.

[0162] The terms "5'-cap" and "7-methylguanylate (m7G) cap" are used interchangeably herein. A 7-methylguanylate residue is located on the 5' terminus of messenger RNA (mRNA) in eukaryotes. RNA polymerase II (Pol II) transcribes mRNA in eukaryotes. Messenger RNA capping occurs generally as follows: The most terminal 5' phosphate group of the mRNA transcript is removed by RNA terminal phosphatase, leaving two terminal phosphates. A guanosine monophosphate (GMP) is added to the terminal phosphate of the transcript by a guanylyl transferase, leaving a 5'-5' triphosphate-linked guanine at the transcript terminus. Finally, the 7-nitrogen of this terminal guanine is methylated by a methyl transferase.

[0163] The terminology "not having a 5'-cap" herein is used to refer to RNA having, for example, a 5'-hydroxyl group instead of a 5'-cap. Such RNA can be referred to as "uncapped RNA", for example. Uncapped RNA can better accumulate in the nucleus following transcription, since 5'-capped RNA is subject to nuclear export. One or more RNA components herein are uncapped.

[0164] 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).

[0165] 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.

[0166] 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.

[0167] 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.

[0168] 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.

[0169] 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.

[0170] 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”, “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).

[0171] 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”, “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.

[0172] The terms “target site”, “target sequence”, “target site sequence”, “target DNA”, “target locus”, “genomic target site”, “genomic target sequence”, “genomic target locus”, “target polynucleotide”, 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.

[0173] 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. The Cas endonuclease may not successfully recognize a target DNA sequence if the target DNA sequence is not followed by a PAM sequence. The sequence and length of a PAM herein can differ depending on the Cas 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.

[0174] An “altered target site”, “altered target sequence”, “modified target site”, “modified target sequence” are used interchangeably herein and refer to a target sequence as disclosed herein that comprises at least one alteration when compared to non-altered target sequence. Such “alterations” include, for example: (i) replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, or (iv) any combination of (i) – (iii).

[0175] A “modified nucleotide” or “edited nucleotide” refers to a nucleotide sequence of interest that comprises at least one alteration when compared to its non-modified nucleotide sequence. Such “alterations” include, for example: (i) replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, or (iv) any combination of (i) – (iii).

[0176] 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.

[0177] 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.

[0178] 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.

[0179] 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.

[0180] 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.

[0181] 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.

[0182] A “plant element” or “plant part” 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), 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. 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" or "part" 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.

[0183] "Progeny" comprises any subsequent generation of a plant.

[0184] 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.

[0185] The term "dicotyledonous" or "dicot" refers to the subclass of angiosperm plants also known 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.

[0186] As used herein, a "male sterile plant" is a plant that does not produce male gametes that are viable or otherwise capable of fertilization. As used herein, a "female sterile plant" is a plant that does not produce female gametes that are viable or otherwise capable of fertilization. It is recognized that male-sterile and female-sterile plants can be female-fertile and male-fertile, respectively. It is further recognized that a male fertile (but female sterile) plant can

produce viable progeny when crossed with a female fertile plant and that a female fertile (but male sterile) plant can produce viable progeny when crossed with a male fertile plant.

[0187] The term “non-conventional yeast” herein refers to any yeast that is not a *Saccharomyces* (e.g., *S. cerevisiae*) or *Schizosaccharomyces* yeast species. (see “Non-Conventional Yeasts in Genetics, Biochemistry and Biotechnology: Practical Protocols”, K. Wolf, K.D. Breunig, G. Barth, Eds., Springer-Verlag, Berlin, Germany, 2003).

[0188] The term “crossed” or “cross” or “crossing” in the context of this disclosure means the fusion of gametes via pollination to produce progeny (i.e., cells, seeds, or plants). The term encompasses both sexual crosses (the pollination of one plant by another) and selfing (self-pollination, i.e., when the pollen and ovule (or microspores and megaspores) are from the same plant or genetically identical plants).

[0189] The term “introgression” refers to the transmission of a desired allele of a genetic locus from one genetic background to another. For example, introgression of a desired allele at a specified locus can be transmitted to at least one progeny plant via a sexual cross between two parent plants, where at least one of the parent plants has the desired allele within its genome. Alternatively, for example, transmission of an allele can occur by recombination between two donor genomes, e.g., in a fused protoplast, where at least one of the donor protoplasts has the desired allele in its genome. The desired allele can be, e.g., a transgene, a modified (mutated or edited) native allele, or a selected allele of a marker or QTL.

[0190] 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. Any phenotypic differences between the two groups may thus be attributed solely to the treatment and not to any inherency of the plant's endogenous genetic makeup.

[0191] “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.

[0192] A “polynucleotide of interest” includes any nucleotide sequence that

[0193] In some aspects, a “polynucleotide of interest” encodes a protein or polypeptide that is “of interest” for a particular purpose, e.g. a selectable marker. In some aspects a trait

(“phenotypic trait”) or polynucleotide “of interest” is one that improves a desirable phenotype of a plant, particularly a crop plant, i.e. a trait of agronomic interest. Polynucleotides of interest: include, but are not limited to, polynucleotides encoding important traits for agronomics, herbicide-resistance, insecticidal resistance, disease resistance, nematode resistance, herbicide resistance, microbial resistance, fungal resistance, viral resistance, fertility or sterility, grain characteristics, commercial products, phenotypic marker, or any other trait of agronomic or commercial importance. A polynucleotide of interest may additionally be utilized in either the sense or anti-sense orientation. Further, more than one polynucleotide of interest may be utilized together, or “stacked”, to provide additional benefit. In some aspects, a “polynucleotide of interest” may encode a gene expression regulatory element, for example a promoter, intron, terminator, 5’UTR, 3’UTR, or other noncoding sequence. In some aspects, a “polynucleotide of interest” may comprise a DNA sequences that encodes for an RNA molecule, for example a functional RNA, siRNA, miRNA, or a guide RNA that is capable of interacting with a Cas endonuclease to bind to a target polynucleotide sequence.

[0194] A “complex trait locus” includes a genomic locus that has multiple transgenes genetically linked to each other.

[0195] 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.

[0196] "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.

[0197] The terms "decreased," "fewer," "slower" and "increased" "faster" "enhanced" "greater" as used herein refers to a decrease or increase in a characteristic of the modified plant element or resulting plant compared to an unmodified plant element or resulting plant. For example, a decrease in a characteristic may be at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, between 5% and 10%, at least 10%, between 10% and 20%, at least 15%, at least 20%, between 20% and 30%, at least 25%, at least 30%, between 30% and 40%, at least 35%, at least 40%, between 40% and 50%, at least 45%, at least 50%, between 50% and 60%, at least about 60%, between 60% and 70%, between 70% and 80%, at least 75%, at least about 80%, between 80% and 90%, at least about 90%, between 90% and 100%, at least 100%, between 100% and 200%, at least 200%, at least about 300%, at least about 400%) or more lower than the untreated control and an increase may be at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, between 5% and 10%, at least 10%, between 10% and 20%, at least 15%, at least 20%, between 20% and 30%, at least 25%, at least 30%, between 30% and 40%, at least 35%, at least 40%, between 40% and 50%, at least 45%, at least 50%, between 50% and 60%, at least about 60%, between 60% and 70%, between 70% and 80%, at least 75%, at least about 80%, between 80% and 90%, at least about 90%, between 90% and 100%, at least 100%, between 100% and 200%, at least 200%, at least about 300%, at least about 400% or more higher than the untreated control.

[0198] 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.

[0199] The meaning of abbreviations is as follows: "sec" means second(s), "min" means minute(s), "h" means hour(s), "d" means day(s), "μL" means microliter(s), "mL" means milliliter(s), "L" means liter(s), "μM" means micromolar, "mM" means millimolar, "M" means molar, "mmol" means millimole(s), "μmole" or "umole" mean micromole(s), "g" means gram(s), "μg" or "ug" means microgram(s), "ng" means nanogram(s), "U" means unit(s), "bp" means base pair(s) and "kb" means kilobase(s).

Double-Strand-Break (DSB) Inducing Agents

[0200] Double-strand breaks induced by double-strand-break-inducing agents, such as endonucleases that cleave the phosphodiester bond within a polynucleotide chain, can result in

the induction of DNA repair mechanisms, including the non-homologous end-joining pathway, and homologous recombination. Endonucleases include a range of different enzymes, including restriction endonucleases (see e.g. Roberts et al., (2003) *Nucleic Acids Res* 1:418-20), Roberts et al., (2003) *Nucleic Acids Res* 31:1805-12, and Belfort et al., (2002) in *Mobile DNA II*, pp. 761-783, Eds. Craigie et al., (ASM Press, Washington, DC)), meganucleases (see e.g., WO 2009/114321; Gao et al. (2010) *Plant Journal* 1:176-187), TAL effector nucleases or TALENs (see e.g., US20110145940, Christian, M., T. Cermak, et al. 2010. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186(2): 757-61 and Boch et al., (2009), *Science* 326(5959): 1509-12), zinc finger nucleases (see e.g. Kim, Y. G., J. Cha, et al. (1996). "Hybrid restriction enzymes: zinc finger fusions to FokI cleavage"), and CRISPR-Cas endonucleases (see e.g. WO2007/025097 application published March 1, 2007).

[0201] In addition to the double-strand break inducing agents, site-specific base conversions can also be achieved to engineer one or more nucleotide changes to create one or more EMEs described herein into the genome. These include for example, a site-specific base edit mediated by an C•G to T•A or an A•T to G•C base editing deaminase enzymes (Gaudelli et al., Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage." *Nature* (2017); Nishida et al. "Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems." *Science* 353 (6305) (2016); Komor et al. "Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage." *Nature* 533 (7603) (2016):420-4.

[0202] Any double-strand-break or -nick or -modification inducing agent may be used for the methods described herein, including for example but not limited to: Cas endonucleases, recombinases, TALENs, zinc finger nucleases, restriction endonucleases, meganucleases, and deaminases.

CRISPR systems and Cas endonucleases

[0203] Methods and compositions are provided for polynucleotide modification 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).

[0204] 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.

[0205] 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.

[0206] 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.

[0207] 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.

[0208] 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 single- or double-strand-break in the genomic sequence, and optionally 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.

[0209] 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).

[0210] Once a double-strand break is induced in the genome, cellular DNA repair mechanisms are activated to repair the break.

Double-Strand-Break Repair and Polynucleotide Modification

[0211] A double-strand-break-inducing agent, such a guided Cas endonuclease can 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, for example via nonhomologous end-joining (NHEJ) or Homology-Directed Repair (HDR) processes which can lead to modifications at the target site. 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). NHEJ is often error-prone and can introduce small mutations in the target site. In plants, NHEJ is often the preferred pathway by which DSBs are remediated.

[0212] Modification of a target polynucleotide includes any one or more of the following: insertion of at least one nucleotide, deletion of at least one nucleotide, chemical alteration of at least one nucleotide, replacement of at least one nucleotide, or mutation of at least one nucleotide. In some aspects, the DNA repair mechanism creates an imperfect repair of the double-strand break, resulting in a change of a nucleotide at the break site. In some aspects, a

polynucleotide template may be provided to the break site, wherein the repair results in a template-directed repair of the break. In some aspects, a donor polynucleotide may be provided to the break site, wherein the repair results in the incorporation of the donor polynucleotide into the break site.

Homology-Directed Repair and Homologous Recombination

[0213] Homology-directed repair (HDR) is a mechanism in cells to repair double-stranded and single stranded DNA breaks. Homology-directed repair includes homologous recombination (HR) and single-strand annealing (SSA) (Lieber. 2010 *Annu. Rev. Biochem.* 79:181-211). The most common form of HDR is called homologous recombination (HR), which has the longest sequence homology requirements between the donor and acceptor DNA. Other forms of HDR include single-stranded annealing (SSA) and breakage-induced replication, and these require shorter sequence homology relative to HR. Homology-directed repair at nicks (single-stranded breaks) can occur via a mechanism distinct from HDR at double-strand breaks (Davis and Maizels. *PNAS* (0027-8424), 111 (10), p. E924-E932).

[0214] By “homology” is meant DNA sequences that are similar. For example, a “region of homology to a genomic region” that is found on the donor DNA is a region of DNA that has a similar sequence to a given “genomic region” in the cell or organism genome. A region of homology can be of any length that is sufficient to promote homologous recombination at the cleaved target site. For example, the region of homology can comprise at least 5-10, 5-15, 5-20, 5-25, 5-30, 5-35, 5-40, 5-45, 5-50, 5-55, 5-60, 5-65, 5-70, 5-75, 5-80, 5-85, 5-90, 5-95, 5-100, 5-200, 5-300, 5-400, 5-500, 5-600, 5-700, 5-800, 5-900, 5-1000, 5-1100, 5-1200, 5-1300, 5-1400, 5-1500, 5-1600, 5-1700, 5-1800, 5-1900, 5-2000, 5-2100, 5-2200, 5-2300, 5-2400, 5-2500, 5-2600, 5-2700, 5-2800, 5-2900, 5-3000, 5-3100 or more bases in length such that the region of homology has sufficient homology to undergo homologous recombination with the corresponding genomic region. “Sufficient homology” indicates that two polynucleotide sequences share structural similarity to act as substrates for a homologous recombination reaction. The structural similarity includes overall length of each polynucleotide fragment, as well as the sequence similarity of the polynucleotides. Sequence similarity can be described by the percent sequence identity over the whole length of the sequences, and/or by conserved regions comprising localized similarities such as contiguous nucleotides having 100% sequence identity, and percent sequence identity over a portion of the length of the sequences.

[0215] The amount of homology or sequence identity shared by a target and a donor polynucleotide can vary and includes total lengths and/or regions having unit integral values in the ranges of about 1-20 bp, 20-50 bp, 50-100 bp, 75-150 bp, 100-250 bp, 150-300 bp, 200-400 bp, 250-500 bp, 300-600 bp, 350-750 bp, 400-800 bp, 450-900 bp, 500-1000 bp, 600-1250 bp, 700-1500 bp, 800-1750 bp, 900-2000 bp, 1-2.5 kb, 1.5-3 kb, 2-4 kb, 2.5-5 kb, 3-6 kb, 3.5-7 kb, 4-8 kb, 5-10 kb, or up to and including the total length of the target site. These ranges include every integer within the range, for example, the range of 1-20 bp includes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 bps. The amount of homology can also be described by percent sequence identity over the full aligned length of the two polynucleotides which includes percent sequence identity of about at least 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100%. Sufficient homology includes any combination of polynucleotide length, global percent sequence identity, and optionally conserved regions of contiguous nucleotides or local percent sequence identity, for example sufficient homology can be described as a region of 75-150 bp having at least 80% sequence identity to a region of the target locus. Sufficient homology can also be described by the predicted ability of two polynucleotides to specifically hybridize under high stringency conditions, see, for example, Sambrook *et al.*, (1989) *Molecular Cloning: A Laboratory Manual*, (Cold Spring Harbor Laboratory Press, NY); *Current Protocols in Molecular Biology*, Ausubel *et al.*, Eds (1994) Current Protocols, (Greene Publishing Associates, Inc. and John Wiley & Sons, Inc.); and, Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, (Elsevier, New York).

[0216] DNA double-strand breaks can be an effective factor to stimulate homologous recombination pathways (Puchta *et al.*, (1995) *Plant Mol Biol* 28:281-92; Tzfira and White, (2005) *Trends Biotechnol* 23:567-9; Puchta, (2005) *J Exp Bot* 56:1-14). Using DNA-breaking agents, a two- to nine-fold increase of homologous recombination was observed between artificially constructed homologous DNA repeats in plants (Puchta *et al.*, (1995) *Plant Mol Biol* 28:281-92). In maize protoplasts, experiments with linear DNA molecules demonstrated enhanced homologous recombination between plasmids (Lyznik *et al.*, (1991) *Mol Gen Genet* 230:209-18).

[0217] Alteration of the genome of a prokaryotic and eukaryotic cell or organism cell, for example, through homologous recombination (HR), is a powerful tool for genetic engineering. Homologous recombination has been demonstrated in plants (Halfter *et al.*, (1992) *Mol Gen Genet* 231:186-93) and insects (Dray and Gloor, 1997, *Genetics* 147:689-99). Homologous recombination has also been accomplished in other organisms. For example, at least 150-200 bp of homology was required for homologous recombination in the parasitic protozoan *Leishmania* (Papadopoulou and Dumas, (1997) *Nucleic Acids Res* 25:4278-86). In the filamentous fungus *Aspergillus nidulans*, gene replacement has been accomplished with as little as 50 bp flanking homology (Chaveroche *et al.*, (2000) *Nucleic Acids Res* 28:e97). Targeted gene replacement has also been demonstrated in the ciliate *Tetrahymena thermophila* (Gaertig *et al.*, (1994) *Nucleic Acids Res* 22:5391-8). In mammals, homologous recombination has been most successful in the mouse using pluripotent embryonic stem cell lines (ES) that can be grown in culture, transformed, selected and introduced into a mouse embryo (Watson *et al.*, 1992, *Recombinant DNA*, 2nd Ed., Scientific American Books distributed by WH Freeman & Co.).

Improving the Probability of HDR in DSB Repair

[0218] Methods and compositions for encouraging the repair of a double strand break via HDR are contemplated.

[0219] In some aspects, the fraction of HR reads relative to the number of total mutant reads (NHEJ + HR) is at least 2, 3, 4, 5, 6, 7, 8, 9, 10, between 10 and 15, 15, between 15 and 20, 20, between 20 and 25, 25, between 25 and 30, 30, between 30 and 40, 40, between 40 and 50, 50, between 50 and 60, 60, between 60 and 70, 70, between 70 and 80, 80, between 80 and 90, 90, between 90 and 100, 100, between 100 and 125, 125, between 125 and 150, greater than 150, or infinitely greater than that observed for a single cleavage strategy.

[0220] In some aspects, the percent of HR reads relative to the number of total mutant reads (NHEJ + HR) is at least 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%.

Genomic Sequence Targeting

[0221] The compositions and methods described herein can be used for genomic sequence targeting, for example a gene or a regulatory element.

[0222] In general, DNA targeting can be performed by cleaving one or both strands at a specific polynucleotide sequence in a cell with a Cas endonuclease associated with a suitable guide polynucleotide component. Once a single or double-strand break is induced in the DNA, the cell's DNA repair mechanism is activated to repair the break via nonhomologous end-joining (NHEJ) or Homology-Directed Repair (HDR) processes which can lead to modifications at the target site.

[0223] Double-strand break repair (for example, at a target site) can be categorized by the mechanism of repair and/or the outcome generated. Non-homologous end joining of a double-strand break that results in an “indel” (insertion or deletion), in the absence of any introduced heterologous polynucleotide, is termed “SDN1” (for Site-Directed Nuclease). Homology-directed repair of a double-strand break that results in a modification of one or a few nucleotides at the target site, in the presence of an introduced heterologous polynucleotide that serves as a “template” for the repair, is termed “SDN2”. Homologous recombination at a double-strand break that results in the insertion of an introduced heterologous polynucleotide at the target site is termed “SDN3”. HDR/HR can be facilitated by the presence of “homology regions” on the donor/template (sequences that share a high percent identity, for example greater than 90%) to DNA sequences on both sides of the double-strand break in the target site.

[0224] In some aspects, the methods and compositions described herein improve the probability of a non-NHEJ repair mechanism outcome at a DSB. In one aspect, an increase of the (HDR or HR) to NHEJ repair ratio is effected.

[0225] The length of the DNA sequence at the target site can vary, and includes, for example, target sites that are at least 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more than 30 nucleotides in length. It is further possible that the target site can be palindromic, that is, the sequence on one strand reads the same in the opposite direction on the complementary strand. The nick/cleavage site can be within the target sequence or the nick/cleavage site could be outside of the target sequence. In another variation, the cleavage could occur at nucleotide positions immediately opposite each other to produce a blunt end cut or, in other cases, the incisions could be staggered to produce single-stranded overhangs, also

called “sticky ends”, which can be either 5' overhangs, or 3' overhangs. Active variants of genomic target sites can also be used. Such active variants can comprise at least 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the given target site, wherein the active variants retain biological activity and hence are capable of being recognized and cleaved by an Cas endonuclease.

[0226] Assays to measure the single or double-strand break of a target site by an endonuclease are known in the art and generally measure the overall activity and specificity of the agent on DNA substrates comprising recognition sites.

[0227] A targeting method herein can be performed in such a way that two or more DNA target sites are targeted in the method, for example. Such a method can optionally be characterized as a multiplex method. Two, three, four, five, six, seven, eight, nine, ten, or more target sites can be targeted at the same time in certain embodiments. A multiplex method is typically performed by a targeting method herein in which multiple different RNA components are provided, each designed to guide a guide polynucleotide/Cas endonuclease complex to a unique DNA target site.

Genomic Sequence Editing

[0228] The process for editing a genomic sequence combining DSB and modification templates generally comprises: introducing into a host cell a DSB-inducing agent, or a nucleic acid encoding a DSB-inducing agent, that recognizes a target sequence in the chromosomal sequence and is able to induce a DSB in the genomic sequence, and at least one polynucleotide modification template comprising at least one nucleotide alteration when compared to the nucleotide sequence to be edited. The polynucleotide modification template can further comprise nucleotide sequences flanking the at least one nucleotide alteration, in which the flanking sequences are substantially homologous to the chromosomal region flanking the DSB. Genome editing using DSB-inducing agents, such as Cas-gRNA complexes, has been described, for example in US20150082478 published on 19 March 2015, WO2015026886 published on 26 February 2015, WO2016007347 published 14 January 2016, and WO/2016/025131 published on 18 February 2016.

[0229] Some uses for guide RNA/Cas endonuclease systems have been described (see for example: US20150082478 A1 published 19 March 2015, WO2015026886 published 26 February 2015, and US20150059010 published 26 February 2015) and include but are not

limited to modifying or replacing nucleotide sequences of interest (such as a regulatory elements), insertion of polynucleotides of interest, gene drop-out, 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.

[0230] Proteins may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known. For example, amino acid sequence variants of the protein(s) can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations include, for example, Kunkel, (1985) *Proc. Natl. Acad. Sci. USA* 82:488-92; Kunkel *et al.*, (1987) *Meth Enzymol* 154:367-82; U.S. Patent No. 4,873,192; Walker and Gaastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan Publishing Company, New York) and the references cited therein. Guidance regarding amino acid substitutions not likely to affect biological activity of the protein is found, for example, in the model of Dayhoff *et al.*, (1978) *Atlas of Protein Sequence and Structure* (Natl Biomed Res Found, Washington, D.C.). Conservative substitutions, such as exchanging one amino acid with another having similar properties, may be preferable. Conservative deletions, insertions, and amino acid substitutions are not expected to produce radical changes in the characteristics of the protein, and the effect of any substitution, deletion, insertion, or combination thereof can be evaluated by routine screening assays. Assays for double-strand-break-inducing activity are known and generally measure the overall activity and specificity of the agent on DNA substrates comprising target sites.

[0231] Described herein are methods for genome editing with Cleavage Ready Cascade (crCascade) Complexes. Following characterization of the guide RNA and PAM sequence, components of the cleavage ready Cascade (crCascade) complex and associated CRISPR RNA (crRNA) may be utilized to modify chromosomal DNA in other organisms including plants. To facilitate optimal expression and nuclear localization (for eukaryotic cells), the genes comprising the crCascade may be optimized as described in WO2016186953 published 24 November 2016, and then delivered into cells as DNA expression cassettes by methods known in the art. The components necessary to comprise an active crCascade complex may also be delivered as RNA with or without modifications that protect the RNA from degradation or as mRNA capped or uncapped (Zhang, Y. *et al.*, 2016, *Nat. Commun.* 7:12617) or Cas protein guide polynucleotide

complexes (WO2017070032 published 27 April 2017), or any combination thereof.

Additionally, a part or part(s) of the crCascade complex and crRNA may be expressed from a DNA construct while other components are delivered as RNA with or without modifications that protect the RNA from degradation or as mRNA capped or uncapped (Zhang et al. 2016 *Nat. Commun.* 7:12617) or Cas protein guide polynucleotide complexes (WO2017070032 published 27 April 2017) or any combination thereof. To produce crRNAs *in-vivo*, tRNA derived elements may also be used to recruit endogenous RNases to cleave crRNA transcripts into mature forms capable of guiding the crCascade complex to its DNA target site, as described, for example, in WO2017105991 published 22 June 2017. crCascade nickase complexes may be utilized separately or concertedly to generate a single or multiple DNA nicks on one or both DNA strands. Furthermore, the cleavage activity of the Cas endonuclease may be deactivated by altering key catalytic residues in its cleavage domain (Sinkunas, T. *et al.*, 2013, *EMBO J.* 32:385-394) resulting in a RNA guided helicase that may be used to enhance homology-directed repair, induce transcriptional activation, or remodel local DNA structures. Moreover, the activity of the Cas cleavage and helicase domains may both be knocked-out and used in combination with other DNA cutting, DNA nicking, DNA binding, transcriptional activation, transcriptional repression, DNA remodeling, DNA deamination, DNA unwinding, DNA recombination enhancing, DNA integration, DNA inversion, and DNA repair agents.

[0232] The transcriptional direction of the tracrRNA for the CRISPR-Cas system (if present) and other components of the CRISPR-Cas system (such as variable targeting domain, crRNA repeat, loop, anti-repeat) can be deduced as described in WO2016186946 published 24 November 2016, and WO2016186953 published 24 November 2016.

[0233] As described herein, once the appropriate guide RNA requirement is established, the PAM preferences for each new system disclosed herein may be examined. If the cleavage ready Cascade (crCascade) complex results in degradation of the randomized PAM library, the crCascade complex can be converted into a nickase by disabling the ATPase dependent helicase activity either through mutagenesis of critical residues or by assembling the reaction in the absence of ATP as described previously (Sinkunas, T. *et al.*, 2013, *EMBO J.* 32:385-394). Two regions of PAM randomization separated by two protospacer targets may be utilized to generate a double-stranded DNA break which may be captured and sequenced to examine the PAM sequences that support cleavage by the respective crCascade complex.

[0234] In one embodiment, the invention describes a method for modifying a target site in the genome of a cell, the method comprising introducing into a cell at least one Cas endonuclease and guide RNA, and identifying at least one cell that has a modification at the target site.

[0235] The nucleotide to be edited can be located within or outside a target site recognized and cleaved by a Cas endonuclease. In one embodiment, the at least one nucleotide modification is not a modification at a target site recognized and cleaved by a Cas endonuclease. In another embodiment, there are at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 900 or 1000 nucleotides between the at least one nucleotide to be edited and the genomic target site.

[0236] A knock-out may be produced by an indel (insertion or deletion of nucleotide bases in a target DNA sequence through NHEJ), or by specific removal of sequence that reduces or completely destroys the function of sequence at or near the targeting site.

[0237] A guide polynucleotide/Cas endonuclease induced targeted mutation can occur in a nucleotide sequence that is located within or outside a genomic target site that is recognized and cleaved by the Cas endonuclease.

[0238] The method for editing a nucleotide sequence in the genome of a cell can be a method without the use of an exogenous selectable marker by restoring function to a non-functional gene product.

[0239] In one embodiment, the invention describes a method for modifying a target site in the genome of a cell, the method comprising introducing into a cell at least one PGEN described herein and at least one donor DNA, wherein said donor DNA comprises a polynucleotide of interest, and optionally, further comprising identifying at least one cell that said polynucleotide of interest integrated in or near said target site.

[0240] In one aspect, the methods disclosed herein may employ homologous recombination (HR) to provide integration of the polynucleotide of interest at the target site.

[0241] Various methods and compositions can be employed to produce a cell or organism having a polynucleotide of interest inserted in a target site via activity of a CRISPR-Cas system component described herein. In one method described herein, a polynucleotide of interest is introduced into the organism cell via a donor DNA construct. As used herein, “donor DNA” is a DNA construct that comprises a polynucleotide of interest to be inserted into 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.

[0242] 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).

[0243] The amount of homology or sequence identity shared by a target and a donor polynucleotide can vary and includes total lengths and/or regions having unit integral values in the ranges of about 1-20 bp, 20-50 bp, 50-100 bp, 75-150 bp, 100-250 bp, 150-300 bp, 200-400 bp, 250-500 bp, 300-600 bp, 350-750 bp, 400-800 bp, 450-900 bp, 500-1000 bp, 600-1250 bp, 700-1500 bp, 800-1750 bp, 900-2000 bp, 1-2.5 kb, 1.5-3 kb, 2-4 kb, 2.5-5 kb, 3-6 kb, 3.5-7 kb, 4-8 kb, 5-10 kb, or up to and including the total length of the target site. These ranges include every integer within the range, for example, the range of 1-20 bp includes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 bps. The amount of homology can also be described by percent sequence identity over the full aligned length of the two polynucleotides which includes percent sequence identity of about at least 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100%. Sufficient homology includes any combination of polynucleotide length, global percent sequence identity, and optionally conserved regions of contiguous nucleotides or local percent sequence identity, for example sufficient homology can be described as a region of 75-150 bp having at least 80% sequence identity to a region of the target locus. Sufficient homology can also be described by the predicted ability of two polynucleotides to specifically hybridize under high stringency conditions, see, for example, Sambrook *et al.*, (1989) *Molecular Cloning: A Laboratory Manual*, (Cold Spring Harbor Laboratory Press, NY); *Current Protocols in Molecular Biology*, Ausubel *et al.*, Eds (1994) Current Protocols, (Greene Publishing Associates, Inc. and John Wiley & Sons, Inc.); and, Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, (Elsevier, New York).

[0244] Episomal DNA molecules can also be ligated into the double-strand break, for example, integration of T-DNAs into chromosomal double-strand breaks (Chilton and Que, (2003) *Plant Physiol* 133:956-65; Salomon and Puchta, (1998) *EMBO J.* 17:6086-95). Once the sequence around the double-strand breaks is altered, for example, by exonuclease activities involved in the maturation of double-strand breaks, gene conversion pathways can restore the original structure if a homologous sequence is available, such as a homologous chromosome in non-dividing somatic cells, or a sister chromatid after DNA replication (Molinier *et al.*, (2004) *Plant Cell* 16:342-52). Ectopic and/or epigenic DNA sequences may also serve as a DNA repair template for homologous recombination (Puchta, (1999) *Genetics* 152:1173-81).

[0245] In one embodiment, the disclosure comprises a method for editing a nucleotide sequence in the genome of a cell, the method comprising introducing into at least one PGEN described herein, and a polynucleotide modification template, wherein said polynucleotide modification template comprises at least one nucleotide modification of said nucleotide sequence, and optionally further comprising selecting at least one cell that comprises the edited nucleotide sequence.

[0246] The guide polynucleotide/Cas endonuclease system can be used in combination with at least one polynucleotide modification template to allow for editing (modification) of a genomic nucleotide sequence of interest. (See also US20150082478, published 19 March 2015 and WO2015026886 published 26 February 2015).

[0247] Polynucleotides of interest and/or traits can be stacked together in a complex trait locus as described in WO2012129373 published 27 September 2012, and in WO2013112686, published 01 August 2013. The guide polynucleotide/Cas9 endonuclease system described herein provides for an efficient system to generate double-strand breaks and allows for traits to be stacked in a complex trait locus.

[0248] A guide polynucleotide/Cas system as described herein, mediating gene targeting, can be used in methods for directing heterologous gene insertion and/or for producing complex trait loci comprising multiple heterologous genes in a fashion similar as disclosed in WO2012129373 published 27 September 2012, where instead of using a double-strand break inducing agent to introduce a gene of interest, a guide polynucleotide/Cas system as disclosed herein is used. By inserting independent transgenes within 0.1, 0.2, 0.3, 0.4, 0.5, 1.0, 2, or even 5 centimorgans (cM) from each other, the transgenes can be bred as a single genetic locus (see, for

example, US20130263324 published 03 October 2013 or WO2012129373 published 14 March 2013). After selecting a plant comprising a transgene, plants comprising (at least) one transgenes can be crossed to form an F1 that comprises both transgenes. In progeny from these F1 (F2 or BC1) 1/500 progeny would have the two different transgenes recombined onto the same chromosome. The complex locus can then be bred as single genetic locus with both transgene traits. This process can be repeated to stack as many traits as desired.

[0249] Further uses for guide RNA/Cas endonuclease systems have been described (See for example: US20150082478 published 19 March 2015, WO2015026886 published 26 February 2015, US20150059010 published 26 February 2015, WO2016007347 published 14 January 2016, and PCT application WO2016025131 published 18 February 2016) and include but are not limited to modifying or replacing nucleotide sequences of interest (such as a regulatory elements), insertion of polynucleotides of interest, gene knock-out, gene-knock in, modification of splicing sites and/or introducing alternate splicing sites, modifications of nucleotide sequences encoding a protein of interest, amino acid and/or protein fusions, and gene silencing by expressing an inverted repeat into a gene of interest.

[0250] Resulting characteristics from the gene editing compositions and methods described herein may be evaluated. Chromosomal intervals that correlate with a phenotype or trait of interest can be identified. A variety of methods well known in the art are available for identifying chromosomal intervals. The boundaries of such chromosomal intervals are drawn to encompass markers that will be linked to the gene controlling the trait of interest. In other words, the chromosomal interval is drawn such that any marker that lies within that interval (including the terminal markers that define the boundaries of the interval) can be used as a marker for a particular trait. In one embodiment, the chromosomal interval comprises at least one QTL, and furthermore, may indeed comprise more than one QTL. Close proximity of multiple QTLs in the same interval may obfuscate the correlation of a particular marker with a particular QTL, as one marker may demonstrate linkage to more than one QTL. Conversely, e.g., if two markers in close proximity show co-segregation with the desired phenotypic trait, it is sometimes unclear if each of those markers identifies the same QTL or two different QTL. The term “quantitative trait locus” or “QTL” refers to a region of DNA that is associated with the differential expression of a quantitative phenotypic trait in at least one genetic background, e.g., in at least one breeding population. The region of the QTL encompasses or is closely linked to the gene or genes that

affect the trait in question. An “allele of a QTL” can comprise multiple genes or other genetic factors within a contiguous genomic region or linkage group, such as a haplotype. An allele of a QTL can denote a haplotype within a specified window wherein said window is a contiguous genomic region that can be defined, and tracked, with a set of one or more polymorphic markers. A haplotype can be defined by the unique fingerprint of alleles at each marker within the specified window.

Recombinant Constructs and Transformation of Cells

[0251] The disclosed guide polynucleotides, Cas endonucleases, polynucleotide modification templates, donor DNAs, guide polynucleotide/Cas endonuclease systems disclosed herein, and any one combination thereof, optionally further comprising one or more polynucleotide(s) of interest, can be introduced into a cell. Cells include, but are not limited to, human, non-human, animal, bacterial, fungal, insect, yeast, non-conventional yeast, and plant cells as well as plants and seeds produced by the methods described herein.

[0252] Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989). Transformation methods are well known to those skilled in the art and are described *infra*.

[0253] Vectors and constructs include circular plasmids, and linear polynucleotides, comprising a polynucleotide of interest and optionally other components including linkers, adapters, regulatory or analysis. In some examples a recognition site and/or target site can be comprised within an intron, coding sequence, 5' UTRs, 3' UTRs, and/or regulatory regions.

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

[0254] The invention further provides expression constructs for expressing in a prokaryotic or eukaryotic cell/organism a guide RNA/Cas system that is capable of recognizing, binding to, and optionally nicking, unwinding, or cleaving all or part of a target sequence.

[0255] In one embodiment, the expression constructs of the disclosure comprise a promoter operably linked to a nucleotide sequence encoding a Cas gene (or plant optimized, including a Cas endonuclease gene described herein) and a promoter operably linked to a guide RNA of the present disclosure. The promoter is capable of driving expression of an operably linked nucleotide sequence in a prokaryotic or eukaryotic cell/organism.

[0256] Nucleotide sequence modification of the guide polynucleotide, VT domain and/or CER domain can be selected from, but not limited to, the group consisting of a 5' cap, a 3' polyadenylated tail, a riboswitch sequence, a stability control sequence, a sequence that forms a dsRNA duplex, a modification or sequence that targets the guide poly nucleotide to a subcellular location, a modification or sequence that provides for tracking, a modification or sequence that provides a binding site for proteins, a Locked Nucleic Acid (LNA), a 5-methyl dC nucleotide, a 2,6-Diaminopurine nucleotide, a 2'-Fluoro A nucleotide, a 2'-Fluoro U nucleotide; a 2'-O-Methyl RNA nucleotide, a phosphorothioate bond, linkage to a cholesterol molecule, linkage to a polyethylene glycol molecule, linkage to a spacer 18 molecule, a 5' to 3' covalent linkage, or any combination thereof. These modifications can result in at least one additional beneficial feature, wherein the additional beneficial feature is selected from the group of a modified or regulated stability, a subcellular targeting, tracking, a fluorescent label, a binding site for a protein or protein complex, modified binding affinity to complementary target sequence, modified resistance to cellular degradation, and increased cellular permeability.

[0257] A method of expressing RNA components such as gRNA in eukaryotic cells for performing Cas9-mediated DNA targeting has been to use RNA polymerase III (Pol III) promoters, which allow for transcription of RNA with precisely defined, unmodified, 5'- and 3'-ends (DiCarlo *et al.*, *Nucleic Acids Res.* 41: 4336-4343; Ma *et al.*, *Mol. Ther. Nucleic Acids* 3:e161). This strategy has been successfully applied in cells of several different species including maize and soybean (US20150082478 published 19 March 2015). Methods for expressing RNA components that do not have a 5' cap have been described (WO2016/025131 published 18 February 2016).

[0258] Various methods and compositions can be employed to obtain a cell or organism having a polynucleotide of interest inserted in a target site for a Cas endonuclease. Such methods can employ homologous recombination (HR) to provide integration of the polynucleotide of interest at the target site. In one method described herein, a polynucleotide of interest is introduced into the organism cell via a donor DNA construct.

[0259] 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.

[0260] 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).

[0261] The amount of homology or sequence identity shared by a target and a donor polynucleotide can vary and includes total lengths and/or regions having unit integral values in the ranges of about 1-20 bp, 20-50 bp, 50-100 bp, 75-150 bp, 100-250 bp, 150-300 bp, 200-400 bp, 250-500 bp, 300-600 bp, 350-750 bp, 400-800 bp, 450-900 bp, 500-1000 bp, 600-1250 bp, 700-1500 bp, 800-1750 bp, 900-2000 bp, 1-2.5 kb, 1.5–3 kb, 2-4 kb, 2.5-5 kb, 3-6 kb, 3.5-7 kb, 4-8 kb, 5-10 kb, or up to and including the total length of the target site. These ranges include every integer within the range, for example, the range of 1-20 bp includes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 bps. The amount of homology can also be described by percent sequence identity over the full aligned length of the two polynucleotides which includes percent sequence identity at least of about 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, between 98% and 99%, 99%, between 99% and 100%, or 100%. Sufficient homology includes any combination of polynucleotide length, global percent sequence identity, and optionally conserved regions of contiguous nucleotides or local percent sequence identity, for example sufficient homology can be described as a region of 75-150 bp having at least 80% sequence identity to a region of the target locus. Sufficient homology can also be described by the predicted ability of two polynucleotides to specifically hybridize under high stringency conditions, see, for example, Sambrook *et al.*, (1989) *Molecular Cloning: A Laboratory Manual*, (Cold Spring Harbor Laboratory Press, NY); *Current Protocols in Molecular Biology*, Ausubel *et al.*, Eds (1994) *Current Protocols*, (Greene Publishing Associates, Inc. and John Wiley & Sons, Inc.); and, Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, (Elsevier, New York).

[0262] The structural similarity between a given genomic region and the corresponding region of homology found on the donor DNA can be any degree of sequence identity that allows for homologous recombination to occur. For example, the amount of homology or sequence

identity shared by the “region of homology” of the donor DNA and the “genomic region” of the organism genome can be at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, such that the sequences undergo homologous recombination

[0263] The region of homology on the donor DNA can have homology to any sequence flanking the target site. While in some instances the regions of homology share significant sequence homology to the genomic sequence immediately flanking the target site, it is recognized that the regions of homology can be designed to have sufficient homology to regions that may be further 5' or 3' to the target site. The regions of homology can also have homology with a fragment of the target site along with downstream genomic regions

[0264] In one embodiment, the first region of homology further comprises a first fragment of the target site and the second region of homology comprises a second fragment of the target site, wherein the first and second fragments are dissimilar.

Polynucleotides of Interest

[0265] Polynucleotides of interest are further described herein and include polynucleotides reflective of the commercial markets and interests of those involved in the development of the crop. Crops and markets of interest change, and as developing nations open up world markets, new crops and technologies will emerge also. In addition, as our understanding of agronomic traits and characteristics such as yield and heterosis increase, the choice of genes for genetic engineering will change accordingly.

[0266] General categories of polynucleotides of interest include, for example, genes of interest involved in information, such as zinc fingers, those involved in communication, such as kinases, and those involved in housekeeping, such as heat shock proteins. More specific polynucleotides of interest include, but are not limited to, genes involved in traits of agronomic interest such as but not limited to: crop yield, grain quality, crop nutrient content, starch and carbohydrate quality and quantity as well as those affecting kernel size, sucrose loading, protein quality and quantity, nitrogen fixation and/or utilization, fatty acid and oil composition, genes encoding proteins conferring resistance to abiotic stress (such as drought, nitrogen, temperature, salinity, toxic metals or trace elements, or those conferring resistance to toxins such as pesticides and herbicides), genes encoding proteins conferring resistance to biotic stress (such as attacks by

fungi, viruses, bacteria, insects, and nematodes, and development of diseases associated with these organisms).

[0267] Agronomically important traits such as oil, starch, and protein content can be genetically altered in addition to using traditional breeding methods. Modifications include increasing content of oleic acid, saturated and unsaturated oils, increasing levels of lysine and sulfur, providing essential amino acids, and also modification of starch. Hordothionin protein modifications are described in U.S. Patent Nos. 5,703,049, 5,885,801, 5,885,802, and 5,990,389.

[0268] Polynucleotide sequences of interest may encode proteins involved in providing disease or pest resistance. By "disease resistance" or "pest resistance" is intended that the plants avoid the harmful symptoms that are the outcome of the plant-pathogen interactions. Pest resistance genes may encode resistance to pests that have great yield drag such as rootworm, cutworm, European Corn Borer, and the like. Disease resistance and insect resistance genes such as lysozymes or cecropins for antibacterial protection, or proteins such as defensins, glucanases or chitinases for antifungal protection, or *Bacillus thuringiensis* endotoxins, protease inhibitors, collagenases, lectins, or glycosidases for controlling nematodes or insects are all examples of useful gene products. Genes encoding disease resistance traits include detoxification genes, such as against fumonisin (U.S. Patent No. 5,792,931); avirulence (avr) and disease resistance (R) genes (Jones *et al.* (1994) *Science* 266:789; Martin *et al.* (1993) *Science* 262:1432; and Mindrinos *et al.* (1994) *Cell* 78:1089); and the like. Insect resistance genes may encode resistance to pests that have great yield drag such as rootworm, cutworm, European Corn Borer, and the like. Such genes include, for example, *Bacillus thuringiensis* toxic protein genes (U.S. Patent Nos. 5,366,892; 5,747,450; 5,736,514; 5,723,756; 5,593,881; and Geiser *et al.* (1986) *Gene* 48:109); and the like.

[0269] An "herbicide resistance protein" or a protein resulting from expression of an "herbicide resistance-encoding nucleic acid molecule" includes proteins that confer upon a cell the ability to tolerate a higher concentration of an herbicide than cells that do not express the protein, or to tolerate a certain concentration of an herbicide for a longer period of time than cells that do not express the protein. Herbicide resistance traits may be introduced into plants by genes coding for resistance to herbicides that act to inhibit the action of acetolactate synthase (ALS, also referred to as acetohydroxyacid synthase, AHAS), in particular the sulfonylurea (UK: sulphonylurea) type herbicides, genes coding for resistance to herbicides that act to inhibit the

action of glutamine synthase, such as phosphinothricin or basta (e.g., the *bar* gene), glyphosate (e.g., the EPSP synthase gene and the GAT gene), HPPD inhibitors (e.g., the HPPD gene) or other such genes known in the art. See, for example, US Patent Nos. 7,626,077, 5,310,667, 5,866,775, 6,225,114, 6,248,876, 7,169,970, 6,867,293, and 9,187,762. The *bar* gene encodes resistance to the herbicide basta, the *nptII* gene encodes resistance to the antibiotics kanamycin and geneticin, and the ALS-gene mutants encode resistance to the herbicide chlorsulfuron.

[0270] Furthermore, it is recognized that the polynucleotide of interest may also comprise antisense sequences complementary to at least a portion of the messenger RNA (mRNA) for a targeted gene sequence of interest. Antisense nucleotides are constructed to hybridize with the corresponding mRNA. Modifications of the antisense sequences may be made as long as the sequences hybridize to and interfere with expression of the corresponding mRNA. In this manner, antisense constructions having 70%, 80%, or 85% sequence identity to the corresponding antisense sequences may be used. Furthermore, portions of the antisense nucleotides may be used to disrupt the expression of the target gene. Generally, sequences of at least 50 nucleotides, 100 nucleotides, 200 nucleotides, or greater may be used.

[0271] In addition, the polynucleotide of interest may also be used in the sense orientation to suppress the expression of endogenous genes in plants. Methods for suppressing gene expression in plants using polynucleotides in the sense orientation are known in the art. The methods generally involve transforming plants with a DNA construct comprising a promoter that drives expression in a plant operably linked to at least a portion of a nucleotide sequence that corresponds to the transcript of the endogenous gene. Typically, such a nucleotide sequence has substantial sequence identity to the sequence of the transcript of the endogenous gene, generally greater than about 65% sequence identity, about 85% sequence identity, or greater than about 95% sequence identity. See U.S. Patent Nos. 5,283,184 and 5,034,323.

[0272] The polynucleotide of interest can also be an expression regulatory element, such as but not limited to a promoter, enhancer, intron, terminator, or UTR (untranslated regulatory sequence). A UTR may be present at either the 5' end or the 3' end of a coding or noncoding sequence. Other examples of polynucleotides of interest include genes encoding for ribonucleotide molecules, for example mRNA, siRNA, or other ribonucleotides. The regulatory element or RNA molecule may be endogenous to the cell in which the genetic modification occurs, or it may be heterologous to the cell.

[0273] The polynucleotide of interest can also be a phenotypic marker. A phenotypic marker is screenable or a selectable marker that includes visual markers and selectable markers whether it is a positive or negative selectable marker. Any phenotypic marker can be used. Specifically, a selectable or screenable marker comprises a DNA segment that allows one to identify, or select for or against a molecule or a cell that comprises it, often under particular conditions. These markers can encode an activity, such as, but not limited to, production of RNA, peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds or compositions and the like.

[0274] Examples of selectable markers include, but are not limited to, DNA segments that comprise restriction enzyme sites; DNA segments that encode products which provide resistance against otherwise toxic compounds including antibiotics, such as, spectinomycin, ampicillin, kanamycin, tetracycline, Basta, neomycin phosphotransferase II (NEO) and hygromycin phosphotransferase (HPT)); DNA segments that encode products which are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); DNA segments that encode products which can be readily identified (e.g., phenotypic markers such as β -galactosidase, GUS; fluorescent proteins such as green fluorescent protein (GFP), cyan (CFP), yellow (YFP), red (RFP), and cell surface proteins); the generation of new primer sites for PCR (e.g., the juxtaposition of two DNA sequence not previously juxtaposed), the inclusion of DNA sequences not acted upon or acted upon by a restriction endonuclease or other DNA modifying enzyme, chemical, etc.; and, the inclusion of a DNA sequences required for a specific modification (e.g., methylation) that allows its identification.

[0275] Additional selectable markers include genes that confer resistance to herbicidal compounds, such as sulphonylureas, glufosinate ammonium, bromoxynil, imidazolinones, and 2,4-dichlorophenoxyacetate (2,4-D). See for example, Acetolactase synthase (ALS) for resistance to sulfonylureas, imidazolinones, triazolopyrimidine sulfonamides, pyrimidinylsalicylates and sulphonylaminocarbonyl-triazolinones (Shaner and Singh, 1997, *Herbicide Activity: Toxicol Biochem Mol Biol* 69-110); glyphosate resistant 5-enolpyruvylshikimate-3-phosphate (EPSPS) (Saroja *et al.* 1998, *J. Plant Biochemistry & Biotechnology* Vol 7:65-72);

[0276] Polynucleotides of interest includes genes that can be stacked or used in combination with other traits, such as but not limited to herbicide resistance or any other trait

described herein. Polynucleotides of interest and/or traits can be stacked together in a complex trait locus as described in US20130263324 published 03 Oct 2013 and in WO/2013/112686, published 01 August 2013.

[0277] A polypeptide of interest includes any protein or polypeptide that is encoded by a polynucleotide of interest described herein.

[0278] Further provided are methods for identifying at least one plant cell, comprising in its genome, a polynucleotide of interest integrated at the target site. A variety of methods are available for identifying those plant cells with insertion into the genome at or near to the target site. Such methods can be viewed as directly analyzing a target sequence to detect any change in the target sequence, including but not limited to PCR methods, sequencing methods, nuclease digestion, Southern blots, and any combination thereof. See, for example, US20090133152 published 21 May 2009. The method also comprises recovering a plant from the plant cell comprising a polynucleotide of interest integrated into its genome. The plant may be sterile or fertile. It is recognized that any polynucleotide of interest can be provided, integrated into the plant genome at the target site, and expressed in a plant.

Expression Elements

[0279] Any polynucleotide encoding a Cas protein, other CRISPR system component, or other polynucleotide disclosed herein may be functionally linked to a heterologous expression element, to facilitate transcription or regulation in a host cell. Such expression elements include but are not limited to: promoter, leader, intron, and terminator. Expression elements may be “minimal” – meaning a shorter sequence derived from a native source, that still functions as an expression regulator or modifier. Alternatively, an expression element may be “optimized” – meaning that its polynucleotide sequence has been altered from its native state in order to function with a more desirable characteristic in a particular host cell (for example, but not limited to, a bacterial promoter may be “maize-optimized” to improve its expression in corn plants). Alternatively, an expression element may be “synthetic” – meaning that it is designed in silico and synthesized for use in a host cell. Synthetic expression elements may be entirely synthetic, or partially synthetic (comprising a fragment of a naturally-occurring polynucleotide sequence).

[0280] It has been shown that certain promoters are able to direct RNA synthesis at a higher rate than others. These are called “strong promoters”. Certain other promoters have been shown to direct RNA synthesis at higher levels only in particular types of cells or tissues and are

often referred to as “tissue specific promoters”, or “tissue-preferred promoters” if the promoters direct RNA synthesis preferably in certain tissues but also in other tissues at reduced levels.

[0281] 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.

[0282] 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.

[0283] Tissue-preferred promoters can be utilized to target enhanced expression within a particular plant tissue. Tissue-preferred promoters include, for example, WO2013103367 published 11 July 2013, Kawamata *et al.*, (1997) *Plant Cell Physiol* 38:792-803; Hansen *et al.*, (1997) *Mol Gen Genet* 254:337-43; Russell *et al.*, (1997) *Transgenic Res* 6:157-68; Rinehart *et al.*, (1996) *Plant Physiol* 112:1331-41; Van Camp *et al.*, (1996) *Plant Physiol* 112:525-35; Canevascini *et al.*, (1996) *Plant Physiol* 112:513-524; Lam, (1994) *Results Probl Cell Differ* 20:181-96; and Guevara-Garcia *et al.*, (1993) *Plant J* 4:495-505. Leaf-preferred promoters include, for example, Yamamoto *et al.*, (1997) *Plant J* 12:255-65; Kwon *et al.*, (1994) *Plant Physiol* 105:357-67; Yamamoto *et al.*, (1994) *Plant Cell Physiol* 35:773-8; Gotor *et al.*, (1993) *Plant J* 3:509-18; Orozco *et al.*, (1993) *Plant Mol Biol* 23:1129-38; Matsuoka *et al.*, (1993) *Proc. Natl. Acad. Sci. USA* 90:9586-90; Simpson *et al.*, (1958) *EMBO J* 4:2723-9; Timko *et al.*, (1988) *Nature* 318:57-8. Root-preferred promoters include, for example, Hire *et al.*, (1992) *Plant Mol Biol* 20:207-18 (soybean root-specific glutamine synthase gene); Miao *et al.*, (1991) *Plant Cell* 3:11-22 (cytosolic glutamine synthase (GS)); Keller and Baumgartner, (1991) *Plant Cell* 3:1051-61 (root-specific control element in the GRP 1.8 gene of French bean); Sanger *et al.*, (1990) *Plant Mol Biol* 14:433-43 (root-specific promoter of *A. tumefaciens* mannopine synthase (MAS)); Bogusz *et al.*, (1990) *Plant Cell* 2:633-41 (root-specific promoters isolated from *Parasponia andersonii* and *Trema tomentosa*); Leach and Aoyagi, (1991) *Plant Sci* 79:69-76 (*A. rhizogenes* rolC and rolD root-inducing genes); Teeri *et al.*, (1989) *EMBO J* 8:343-50 (*Agrobacterium* wound-induced TR1' and TR2' genes); VfENOD-GRP3 gene promoter (Kuster *et al.*, (1995) *Plant Mol Biol* 29:759-72); and rolB promoter (Capano *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.

[0284] Seed-preferred promoters include both seed-specific promoters active during seed development, as well as seed-germinating promoters active during seed germination. See, Thompson *et al.*, (1989) *BioEssays* 10:108. Seed-preferred promoters include, but are not limited to, Cim1 (cytokinin-induced message); cZ19B1 (maize 19 kDa zein); and milps (myo-inositol-1-phosphate synthase); and for example those disclosed in WO2000011177 published 02 March 2000 and U.S. Patent 6,225,529. For dicots, seed-preferred promoters include, but are not limited to, bean β -phaseolin, napin, β -conglycinin, soybean lectin, cruciferin, and the like. For monocots, seed-preferred promoters include, but are not limited to, maize 15 kDa zein, 22 kDa zein, 27 kDa gamma zein, waxy, shrunken 1, shrunken 2, globulin 1, oleosin, and nuc1. See also, WO2000012733 published 09 March 2000, where seed-preferred promoters from *END1* and *END2* genes are disclosed.

[0285] Chemical inducible (regulated) promoters can be used to modulate the expression of a gene in a prokaryotic and eukaryotic cell or organism through the application of an exogenous chemical regulator. The promoter may be a chemical-inducible promoter, where application of the chemical induces gene expression, or a chemical-repressible promoter, where application of the chemical represses gene expression. Chemical-inducible promoters include, but are not limited to, the maize In2-2 promoter, activated by benzene sulfonamide herbicide safeners (De Veylder *et al.*, (1997) *Plant Cell Physiol* 38:568-77), the maize GST promoter (GST-II-27, WO1993001294 published 21 January 1993), activated by hydrophobic electrophilic compounds used as pre-emergent herbicides, and the tobacco PR-1a promoter (Ono *et al.*, (2004) *Biosci Biotechnol Biochem* 68:803-7) activated by salicylic acid. Other chemical-regulated promoters include steroid-responsive promoters (see, for example, the glucocorticoid-inducible promoter (Schena *et al.*, (1991) *Proc. Natl. Acad. Sci. USA* 88:10421-5; McNellis *et al.*, (1998) *Plant J* 14:247-257); tetracycline-inducible and tetracycline-repressible promoters (Gatz *et al.*, (1991) *Mol Gen Genet* 227:229-37; U.S. Patent Nos. 5,814,618 and 5,789,156).

[0286] 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.*

[0287] 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.

[0288] Another example of an inducible promoter useful in plant cells, is the ZmCAS1 promoter, described in US20130312137 published 21 November 2013.

[0289] 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.

Developmental Genes (Morphogenic Factors)

[0290] Morphogenic factors (also called “developmental genes” or “dev genes”, which are used synonymously throughout) are polynucleotides that act to enhance the rate, efficiency, and/or efficacy of targeted polynucleotide modification by a number of mechanisms, some of which are related to the capability of stimulating growth of a cell or tissue, including but not limited to promoting progression through the cell cycle, inhibiting cell death, such as apoptosis, stimulating cell division, and/or stimulating embryogenesis. The polynucleotides can fall into several categories, including but not limited to, cell cycle stimulatory polynucleotides, developmental polynucleotides, anti-apoptosis polynucleotides, hormone polynucleotides, transcription factors, or silencing constructs targeted against cell cycle repressors or pro-apoptotic factors. Methods and compositions for rapid and efficient transformation of plants by transforming cells of plant explants with an expression construct comprising a heterologous nucleotide encoding a morphogenic factor are described in US Patent Application Publication No. US2017/0121722 (published 04 May 2017).

[0291] A morphogenic factor (gene or protein) may be involved in plant metabolism, organ development, stem cell development, cell growth stimulation, organogenesis, somatic embryogenesis initiation, accelerated somatic embryo maturation, initiation and/or development of the apical meristem, initiation and/or development of shoot meristem, or a combination thereof.

[0292] In some aspects, the morphogenic factor is a molecule selected from one or more of the following categories: 1) cell cycle stimulatory polynucleotides including plant viral replicase genes such as RepA, cyclins, E2F, proliferin, cdc2 and cdc25; 2) developmental polynucleotides such as Lec1, Kn1 family, WUSCHEL, Zwiller, BBM, Aintegumenta (ANT), FUS3, and members of the Knotted family, such as Kn1, STM, OSH1, and SbH1; 3) anti-apoptosis polynucleotides such as CED9, Bcl2, Bcl-X(L), Bcl-W, A1, McL-1, Mac1, Boo, and Bax-inhibitors; 4) hormone polynucleotides such as IPT, TZS, and CKI-1; and 5) silencing constructs targeted against cell cycle repressors, such as Rb, CKI, prohibitin, and wee1, or stimulators of apoptosis such as APAF-1, bad, bax, CED-4, and caspase-3, and repressors of plant developmental transitions, such as Pickle and WD polycomb genes including FIE and Medea. The polynucleotides can be silenced by any known method such as antisense, RNA interference, cosuppression, chimerplasty, or transposon insertion.

[0293] In some aspects, the morphogenic factor is a member of the WUS/WOX gene family (WUS1, WUS2, WUS3, WOX2A, WOX4, WOX5, or WOX9) see US patents 7,348,468 and 7,256,322 and United States Patent Application publications 20170121722 and 20070271628; Laux et al. (1996) Development 122:87-96; and Mayer et al. (1998) Cell 95:805-815; van der Graaff et al., 2009, Genome Biology 10:248; Dolzblasz et al., 2016, Mol. Plant 19:1028-39. The Wuschel protein, designated hereafter as WUS, plays a key role in the initiation and maintenance of the apical meristem, which contains a pool of pluripotent stem cells (Endrizzi, et al., (1996) Plant Journal 10:967-979; Laux, et al., (1996) Development 122:87-96; and Mayer, et al., (1998) Cell 95:805-815). Modulation of WUS/WOX is expected to modulate plant and/or plant tissue phenotype including plant metabolism, organ development, stem cell development, cell growth stimulation, organogenesis, somatic embryogenesis initiation, accelerated somatic embryo maturation, initiation and/or development of the apical meristem, initiation and/or development of shoot meristem, or a combination thereof. WUS encodes a novel homeodomain protein which presumably functions as a transcriptional regulator (Mayer, et al., (1998) Cell 95:805-815). The stem cell population of Arabidopsis shoot meristems is believed to be maintained by a regulatory loop between the CLAVATA (CLV) genes which promote organ initiation and the WUS gene which is required for stem cell identity, with the CLV genes repressing WUS at the transcript level, and WUS expression being sufficient to induce meristem cell identity and the expression of the stem cell marker CLV3 (Brand, et al.,

(2000) *Science* 289:617-619; Schoof, et al., (2000) *Cell* 100:635-644). Expression of *Arabidopsis* WUS can induce stem cells in vegetative tissues, which can differentiate into somatic embryos (Zuo, et al. (2002) *Plant J* 30:349-359). Also of interest in this regard would be a MYB118 gene (see U.S. Patent 7,148,402), MYB115 gene (see Wang et al. (2008) *Cell Research* 224-235), a BABYBOOM gene (BBM; see Boutilier et al. (2002) *Plant Cell* 14:1737-1749), or a CLAVATA gene (see, for example, U.S. Patent 7,179,963).

[0294] In some embodiments, the morphogenic factor or protein is a member of the AP2/ERF family of proteins. The AP2/ERF family of proteins is a plant-specific class of putative transcription factors that regulate a wide variety of developmental processes and are characterized by the presence of an AP2 DNA binding domain that is predicted to form an amphipathic alpha helix that binds DNA (PFAM Accession PF00847). The AP2 domain was first identified in APETALA2, an *Arabidopsis* protein that regulates meristem identity, floral organ specification, seed coat development, and floral homeotic gene expression. The AP2/ERF proteins have been subdivided into distinct subfamilies based on the presence of conserved domains. Initially, the family was divided into two subfamilies based on the number of DNA binding domains, with the ERF subfamily having one DNA binding domain, and the AP2 subfamily having 2 DNA binding domains. As more sequences were identified, the family was subsequently subdivided into five subfamilies: AP2, DREB, ERF, RAV, and others. (Sakuma *et al.* (2002) *Biochem Biophys Res Comm* 290:998-1009).

[0295] Members of the APETALA2 (AP2) family of proteins function in a variety of biological events, including but not limited to, development, plant regeneration, cell division, embryogenesis, and morphogenic (see, *e.g.*, Riechmann and Meyerowitz (1998) *Biol Chem* 379:633-646; Saleh and Pagés (2003) *Genetika* 35:37-50 and Database of Arabidopsis Transcription Factors at daft.cbi.pku.edu.cn). The AP2 family includes, but is not limited to, AP2, ANT, Glossy15, AtBBM, BnBBM, and maize ODP2/BBM.

[0296] Other morphogenic factors useful in the present disclosure include, but are not limited to, Ovule Development Protein 2 (ODP2) polypeptides, and related polypeptides, *e.g.*, Babyboom (BBM) protein family proteins. In an aspect, the polypeptide comprising the two AP2-DNA binding domains is an ODP2, BBM2, BMN2, or BMN3 polypeptide. The ODP2 polypeptides of the disclosure contain two predicted APETALA2 (AP2) domains and are members of the AP2 protein family (PFAM Accession PF00847). The AP2 family of putative

transcription factors has been shown to regulate a wide range of developmental processes, and the family members are characterized by the presence of an AP2 DNA binding domain. This conserved core is predicted to form an amphipathic alpha helix that binds DNA. The AP2 domain was first identified in APETALA2, an Arabidopsis protein that regulates meristem identity, floral organ specification, seed coat development, and floral homeotic gene expression. The AP2 domain has now been found in a variety of proteins. The ODP2 polypeptides share homology with several polypeptides within the AP2 family, e.g., see FIG. 1 of US 8,420,893, which is incorporated herein by reference in its entirety, provides an alignment of the maize and rice ODP2 polypeptides with eight other proteins having two AP2 domains. A consensus sequence of all proteins appearing in the alignment of US 8420893 is also provided in FIG. 1 therein.

[0297] In some embodiments, the morphogenic factor is a babyboom (BBM) polypeptide, which is a member of the AP2 family of transcription factors. The BBM protein from *Arabidopsis* (AtBBM) is preferentially expressed in the developing embryo and seeds and has been shown to play a central role in regulating embryo-specific pathways. Overexpression of AtBBM has been shown to induce spontaneous formation of somatic embryos and cotyledon-like structures on seedlings. See, Boutilier *et al.* (2002) *The Plant Cell* 14:1737-1749. The maize BBM protein also induces embryogenesis and promotes transformation (See, U.S. Patent No. 7,579,529, which is herein incorporated by reference in its entirety). Thus, BBM polypeptides stimulate proliferation, induce embryogenesis, enhance the regenerative capacity of a plant, enhance transformation, and as demonstrated herein, enhance rates of targeted polynucleotide modification. As used herein "regeneration" refers to a morphogenic response that results in the production of new tissues, organs, embryos, whole plants or parts of whole plants that are derived from a single cell or a group of cells. Regeneration may proceed indirectly via a callus phase or directly, without an intervening callus phase. "Regenerative capacity" refers to the ability of a plant cell to undergo regeneration.

[0298] Other morphogenic factors useful in the present disclosure include, but are not limited to, LEC1 (Lotan *et al.*, 1998, *Cell* 93:1195-1205), LEC2 (Stone *et al.*, 2008, *PNAS* 105:3151-3156; Belide *et al.*, 2013, *Plant Cell Tiss. Organ Cult* 113:543-553), KN1/STM (Sinha *et al.*, 1993, *Genes Dev* 7:787-795), the IPT gene from *Agrobacterium* (Ebinuma and Komamine, 2001, *In vitro Cell. Dev Biol – Plant* 37:103-113), MONOPTEROS-DELTA (Ckurshumova *et*

al., 2014, New Phytol. 204:556-566), the Agrobacterium AV-6b gene (Wabiko and Minemura 1996, Plant Physiol. 112:939-951), the combination of the Agrobacterium IAA-h and IAA-m genes (Endo et al., 2002, Plant Cell Rep., 20:923-928), the Arabidopsis SERK gene (Hecht et al., 2001, Plant Physiol. 127:803-816), the Arabidopsis AGL15 gene (Harding et al., 2003, Plant Physiol. 133:653-663), and the FUSCA gene (Castle and Meinke, Plant Cell 6:25-41), and the PICKLE gene (Ogas et al., 1999, PNAS 96:13839-13844).

[0299] The morphogenic factor can be derived from a monocot. In various aspects, the morphogenic factor is derived from barley, maize, millet, oats, rice, rye, *Setaria* sp., sorghum, sugarcane, switchgrass, triticale, turfgrass, or wheat.

[0300] The morphogenic factor can be derived from a dicot. The morphogenic factor can be derived from kale, cauliflower, broccoli, mustard plant, cabbage, pea, clover, alfalfa, broad bean, tomato, cassava, soybean, canola, alfalfa, sunflower, safflower, tobacco, *Arabidopsis*, or cotton.

[0301] The present disclosure encompasses isolated or substantially purified polynucleotide or polypeptide morphogenic factor compositions.

[0302] The morphogenic factor 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 of the morphogenic proteins can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence 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. Pat. 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.), herein incorporated by reference. Conservative substitutions, such as exchanging one amino acid with another having similar properties, may be optimal.

[0303] In some embodiments, polynucleotides or polypeptides having homology to a known morphogenic factor and/or sharing conserved functional domains can be identified by screening sequence databases using programs such as BLAST, or using standard nucleic acid

hybridization techniques known in the art, for example as described in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes*, Part I, Chapter 2 (Elsevier, NY); Ausubel *et al.*, eds. (1995) *Current Protocols in Molecular Biology*, Chapter 2 (Greene Publishing and Wiley-Interscience, NY); and, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, NY).

[0304] In some aspects, the morphogenic factor is selected from the group consisting of: SEQ ID NOs:1-5, 11-16, 22, and 23-47. In some aspects, the morphogenic protein is selected from the group consisting of: SEQ ID NOs: 6-10, 17-21, and 48-73.

[0305] In some aspects, a plurality of morphogenic factors is selected. When multiple morphogenic factors are used, the polynucleotides encoding each of the factors can be present on the same expression cassette or on separate expression cassettes. Likewise, the polynucleotide(s) encoding the morphogenic factor(s) and the polynucleotide encoding the double-strand break-inducing agent can be located on the same or different expression cassettes. When two or more factors are coded for by separate expression cassettes, the expression cassettes can be provided to the organism simultaneously or sequentially.

[0306] In some aspects, the expression of the morphogenic factor is transient. In some aspects, the expression of the morphogenic factor is constitutive. In some aspects, the expression of the morphogenic factor is specific to a particular tissue or cell type. In some aspects, the expression of the morphogenic factor is temporally regulated. In some aspects, the expression of the morphogenic factor is regulated by an environmental condition, such as temperature, time of day, or other factor. In some aspects, the expression of the morphogenic factor is stable. In some aspects, expression of the morphogenic factor is controlled. The controlled expression may be a pulsed expression of the morphogenic factor for a particular period of time. Alternatively, the morphogenic factor may be expressed in only some transformed cells and not expressed in others. The control of expression of the morphogenic factor can be achieved by a variety of methods as disclosed herein.

Helper Plasmids

[0307] *Agrobacterium*, a natural plant pathogen, has been widely used for the transformation of dicotyledonous plants and more recently for transformation of monocotyledonous plants. The advantage of the *Agrobacterium*-mediated gene transfer system is

that it offers the potential to regenerate transgenic cells at relatively high frequencies without a significant reduction in plant regeneration rates. Moreover, the process of DNA transfer to the plant genome is well characterized relative to other DNA delivery methods. DNA transferred via *Agrobacterium* is less likely to undergo any major rearrangements than is DNA transferred via direct delivery, and it integrates into the plant genome often in single or low copy numbers.

[0308] The most commonly used *Agrobacterium*-mediated gene transfer system is a binary transformation vector system where the *Agrobacterium* has been engineered to include a disarmed, or nononcogenic, Ti helper plasmid, which encodes the *vir* functions necessary for DNA transfer, and a much smaller separate plasmid called the binary vector plasmid, which carries the transferred DNA, or the T-DNA region. The T-DNA is defined by sequences at each end, called T-DNA borders, which play an important role in the production of T-DNA and in the transfer process.

[0309] Binary vectors are vectors in which the virulence genes are placed on a different plasmid than the one carrying the T-DNA region (Bevan, 1984, Nucl. Acids. Res. 12: 8711-8721). The development of T-DNA binary vectors has made the transformation of plant cells easier as they do not require recombination. The finding that some of the virulence genes exhibited gene dosage effects (Jin et al., J. Bacteriol. (1987) 169:4417-4425) led to the development of a superbinary vector, which carried additional virulence genes (Komari, T., et al., Plant Cell Rep. (1990), 9:303-306). These early superbinary vectors carried a large “*vir*” fragment (~14.8 kbp) from the hypervirulence Ti plasmid, pTiBo542, which had been introduced into a standard binary vector (*ibid*). The superbinary vectors resulted in vastly improved plant transformation. For example, Hiei, Y., et al. (Plant J. (1994) 6:271-282) described efficient transformation of rice by *Agrobacterium*, and subsequently there were reports of using this system for maize, barley and wheat (Ishida, Y., et al., Nat. Biotech. (1996) 14:745-750; Tingay, S., et al., Plant J. (1997) 11:1369-1376; and Cheng, M., et al., Plant Physiol. (1997) 115:971-980; see also U.S. Pat. No. 5,591,616 to Hiei et al). Examples of prior superbinary vectors include pTOK162 (Japanese Patent Appl. (Kokai) No. 4-222527, EP-A-504,869, EP-A-604,662, and U.S. Pat. No. 5,591,616) and pTOK233 (see Komari, T., *ibid*; and Ishida, Y., et al., *ibid*).

[0310] The present disclosure comprises methods and compositions utilizing superbinary vectors comprising *vir* genes. In various aspects, the present disclosure provides a vector

comprising: (a) an origin of replication for propagation and stable maintenance in *Escherichia coli*; (b) an origin of replication for propagation and stable maintenance in *Agrobacterium* spp.; (c) a selectable marker gene; and (d) *Agrobacterium* spp. virulence genes *virB1-B11*; *virC1-C2*; *virD1-D2*; and *virG* genes. In an aspect, the vector further comprises *Agrobacterium* spp. virulence genes *virA*, *virD3*, *virD4*, *virD5*, *virE1*, *virE2*, *virE3*, *virH*, *virH1*, *virH2*, *virK*, *virL*, *virM*, *virP*, or *virQ*, or combinations thereof. In an aspect, the vector comprises *Agrobacterium* sp. virulence genes *virB1-B11*, *virC1-C2*; *virD1-D2*, and *virG* genes. In another aspect, the vector comprises *Agrobacterium* sp. virulence genes *virA*, *virB1-B11*, *virC1-C2*; *virD1-D5*, *virE1-E3*, *virG*, and *virJ* genes.

[0311] *Agrobacteria* with helper plasmids, such as pVIR9, pVIR7, or pVIR10, can significantly improve the transient protein expression, transient T-DNA delivery, somatic embryo phenotypes, transformation frequencies, recovery of quality events, and usable quality events in different plant lines (WO2017078836A1, published 11 May 2017).

[0312] VIR genes are also used for the improvement of transformation with *Ochrobactrum*, for example as disclosed in US20180216123, published 02 August 2018.

Introduction of System Components into a Cell

[0313] The methods and compositions described herein do not depend on a particular method for introducing a sequence into an organism or cell, only that the polynucleotide or polypeptide gains access to the interior of at least one cell of the organism. Introducing includes reference to the incorporation of a nucleic acid into a eukaryotic or prokaryotic cell where the nucleic acid may be incorporated into the genome of the cell, and includes reference to the transient (direct) provision of a nucleic acid, protein or ribonucleoprotein complex to the cell.

[0314] Methods for introducing polynucleotides or polypeptides or a polynucleotide-protein complex into cells or organisms are known in the art including, but not limited to, microinjection, electroporation, stable transformation methods, transient transformation methods, ballistic particle acceleration (particle bombardment), whiskers mediated transformation, *Agrobacterium*-mediated transformation, direct gene transfer, viral-mediated introduction, transfection, transduction, cell-penetrating peptides, mesoporous silica nanoparticle (MSN)-mediated direct protein delivery, topical applications, sexual crossing, sexual breeding, and any combination thereof. General methods for the introduction of polynucleotides into a cell for

transformation, for example *Agrobacterium*-mediated transformation, *Ochrobactrum*-mediated transformation, and particle bombardment-mediated transformation of cells are known in the art.

[0315] For example, the guide polynucleotide (guide RNA, crNucleotide + tracrNucleotide, guide DNA and/or guide RNA-DNA molecule) can be introduced into a cell directly (transiently) as a single stranded or double stranded polynucleotide molecule. The guide RNA (or crRNA + tracrRNA) can also be introduced into a cell indirectly by introducing a recombinant DNA molecule comprising a heterologous nucleic acid fragment encoding the guide RNA (or crRNA + tracrRNA), operably linked to a specific promoter that is capable of transcribing the guide RNA (crRNA+tracrRNA molecules) in said cell. The specific promoter can be, but is not limited to, a RNA polymerase III promoter, which allow for transcription of RNA with precisely defined, unmodified, 5'- and 3'-ends (Ma *et al.*, 2014, *Mol. Ther. Nucleic Acids* 3:e161; DiCarlo *et al.*, 2013, *Nucleic Acids Res.* 41: 4336-4343; WO2015026887, published 26 February 2015). Any promoter capable of transcribing the guide RNA in a cell can be used and includes a heat shock /heat inducible promoter operably linked to a nucleotide sequence encoding the guide RNA.

[0316] Protocols for introducing polynucleotides, polypeptides or polynucleotide-protein complexes into eukaryotic cells, such as plants or plant cells are known and include microinjection (Crossway *et al.*, (1986) *Biotechniques* 4:320-34 and U.S. Patent No. 6,300,543), meristem transformation (U.S. Patent No. 5,736,369), electroporation (Riggs *et al.*, (1986) *Proc. Natl. Acad. Sci. USA* 83:5602-6, *Agrobacterium*-mediated transformation (U.S. Patent Nos. 5,563,055 and 5,981,840), whiskers mediated transformation (Ainley *et al.* 2013, *Plant Biotechnology Journal* 11:1126-1134; Shaheen A. and M. Arshad 2011 Properties and Applications of Silicon Carbide (2011), 345-358 Editor(s): Gerhardt, Rosario. Publisher: InTech, Rijeka, Croatia. CODEN: 69PQBP; ISBN: 978-953-307-201-2), direct gene transfer (Paszkowski *et al.*, (1984) *EMBO J* 3:2717-22), and ballistic particle acceleration (U.S. Patent Nos. 4,945,050; 5,879,918; 5,886,244; 5,932,782; Tomes *et al.*, (1995) "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment" in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, ed. Gamborg & Phillips (Springer-Verlag, Berlin); McCabe *et al.*, (1988) *Biotechnology* 6:923-6; Weissinger *et al.*, (1988) *Ann Rev Genet* 22:421-77; Sanford *et al.*, (1987) *Particulate Science and Technology* 5:27-37 (onion); Christou *et al.*, (1988) *Plant Physiol* 87:671-4 (soybean); Finer and McMullen, (1991) *In vitro Cell Dev Biol* 27P:175-82

(soybean); Singh *et al.*, (1998) *Theor Appl Genet* 96:319-24 (soybean); Datta *et al.*, (1990) *Biotechnology* 8:736-40 (rice); Klein *et al.*, (1988) *Proc. Natl. Acad. Sci. USA* 85:4305-9 (maize); Klein *et al.*, (1988) *Biotechnology* 6:559-63 (maize); U.S. Patent Nos. 5,240,855; 5,322,783 and 5,324,646; Klein *et al.*, (1988) *Plant Physiol* 91:440-4 (maize); Fromm *et al.*, (1990) *Biotechnology* 8:833-9 (maize); Hooykaas-Van Slogteren *et al.*, (1984) *Nature* 311:763-4; U.S. Patent No. 5,736,369 (cereals); Bytebier *et al.*, (1987) *Proc. Natl. Acad. Sci. USA* 84:5345-9 (*Liliaceae*); De Wet *et al.*, (1985) in *The Experimental Manipulation of Ovule Tissues*, ed. Chapman *et al.*, (Longman, New York), pp. 197-209 (pollen); Kaeppler *et al.*, (1990) *Plant Cell Rep* 9:415-8 and Kaeppler *et al.*, (1992) *Theor Appl Genet* 84:560-6 (whisker-mediated transformation); D'Halluin *et al.*, (1992) *Plant Cell* 4:1495-505 (electroporation); Li *et al.*, (1993) *Plant Cell Rep* 12:250-5; Christou and Ford (1995) *Annals Botany* 75:407-13 (rice) and Osjoda *et al.*, (1996) *Nat Biotechnol* 14:745-50 (maize via *Agrobacterium tumefaciens*).

[0317] 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.

[0318] The methods provided herein rely upon the use of bacteria-mediated and/or biolistic-mediated gene transfer to produce regenerable plant cells. Bacterial strains useful in the methods of the disclosure include, but are not limited to, a disarmed *Agrobacteria*, an *Ochrobactrum* bacteria or a *Rhizobiaceae* bacteria. Standard protocols for particle bombardment (Finer and McMullen, 1991, *In Vitro Cell Dev. Biol. – Plant* 27:175-182), *Agrobacterium*-mediated transformation (Jia *et al.*, 2015, *Int J. Mol. Sci.* 16:18552-18543; US2017/0121722 incorporated herein by reference in its entirety), or *Ochrobactrum*-mediated transformation (US2018/0216123 incorporated herein by reference in its entirety) can be used with the methods and compositions of the disclosure.

[0319] 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.

[0320] Nucleic acids and proteins can be provided to a cell by any method including methods using molecules to facilitate the uptake of anyone or all components of a guided Cas system (protein and/or nucleic acids), such as cell-penetrating peptides and nanocarriers. See also US20110035836 published 10 February 2011, and EP2821486A1 published 07 January 2015.

[0321] Other methods of introducing polynucleotides into a prokaryotic and eukaryotic cell or organism or plant part can be used, including plastid transformation methods, and the methods for introducing polynucleotides into tissues from seedlings or mature seeds.

[0322] Stable transformation is intended to mean that the nucleotide construct introduced into an organism integrates into a genome of the organism and is capable of being inherited by the progeny thereof. Transient transformation is intended to mean that a polynucleotide is introduced into the organism and does not integrate into a genome of the organism or a polypeptide is introduced into an organism. Transient transformation indicates that the introduced composition is only temporarily expressed or present in the organism.

[0323] A variety of methods are available to identify those cells having an altered genome at or near a target site without using a screenable marker phenotype. Such methods can be viewed as directly analyzing a target sequence to detect any change in the target sequence, including but not limited to PCR methods, sequencing methods, nuclease digestion, Southern blots, and any combination thereof.

Cells and Organisms

[0324] The presently disclosed polynucleotides and polypeptides can be introduced into a cell. Cells include, but are not limited to, human, non-human, animal, mammalian, bacterial, protist, fungal, insect, yeast, non-conventional yeast, and plant cells, as well as plants and seeds produced by the methods described herein. In some aspects, the cell of the organism is a reproductive cell, a somatic cell, a meiotic cell, a mitotic cell, a stem cell, or a pluripotent stem cell. Any cell from any organism may be used with the compositions and methods described herein, including monocot and dicot plants, and plant elements.

Animal Cells

[0325] The presently disclosed polynucleotides and polypeptides can be introduced into an animal cell. Animal cells can include, but are not limited to: an organism of a phylum

including chordates, arthropods, mollusks, annelids, cnidarians, or echinoderms; or an organism of a class including mammals, insects, birds, amphibians, reptiles, or fishes. In some aspects, the animal is human, mouse, *C. elegans*, rat, fruit fly (*Drosophila* spp.), zebrafish, chicken, dog, cat, guinea pig, hamster, chicken, Japanese ricefish, sea lamprey, pufferfish, tree frog (*e.g.*, *Xenopus* spp.), monkey, or chimpanzee. Particular cell types that are contemplated include haploid cells, diploid cells, reproductive cells, neurons, muscle cells, endocrine or exocrine cells, epithelial cells, muscle cells, tumor cells, embryonic cells, hematopoietic cells, bone cells, germ cells, somatic cells, stem cells, pluripotent stem cells, induced pluripotent stem cells, progenitor cells, meiotic cells, and mitotic cells. In some aspects, a plurality of cells from an organism may be used.

[0326] The compositions and methods described herein may be used to edit the genome of an animal cell in various ways. In one aspect, it may be desirable to delete one or more nucleotides. In another aspect, it may be desirable to insert one or more nucleotides. In one aspect, it may be desirable to replace one or more nucleotides. In another aspect, it may be desirable to modify one or more nucleotides via a covalent or non-covalent interaction with another atom or molecule.

[0327] Genome modification may be used to effect a genotypic and/or phenotypic change on the target organism. Such a change is preferably related to an improved phenotype of interest or a physiologically-important characteristic, the correction of an endogenous defect, or the expression of some type of expression marker. In some aspects, the phenotype of interest or physiologically-important characteristic is related to the overall health, fitness, or fertility of the animal, the ecological fitness of the organism, or the relationship or interaction of the organism with other organisms in its environment.

[0328] Cells that have been genetically modified using the compositions or methods described herein may be transplanted to a subject for purposes such as gene therapy, *e.g.* to treat a disease, or as an antiviral, antipathogenic, or anticancer therapeutic, for the production of genetically modified organisms in agriculture, or for biological research.

Plant Cells and Plants

[0329] Examples of monocot plants that can be used include, but are not limited to, corn (*Zea mays*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (*e.g.*, pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*),

foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*), wheat (*Triticum* species, for example *Triticum aestivum*, *Triticum monococcum*), sugarcane (*Saccharum spp.*), oats (*Avena*), barley (*Hordeum*), switchgrass (*Panicum virgatum*), pineapple (*Ananas comosus*), banana (*Musa spp.*), palm, ornamentals, turfgrasses, and other grasses.

[0330] Examples of dicot plants that can be used include, but are not limited to, soybean (*Glycine max*), *Brassica* species (for example but not limited to: oilseed rape or Canola) (*Brassica napus*, *B. campestris*, *Brassica rapa*, *Brassica juncea*), alfalfa (*Medicago sativa*), tobacco (*Nicotiana tabacum*), *Arabidopsis* (*Arabidopsis thaliana*), sunflower (*Helianthus annuus*), cotton (*Gossypium arboreum*, *Gossypium barbadense*), and peanut (*Arachis hypogaea*), tomato (*Solanum lycopersicum*), and potato (*Solanum tuberosum*).

[0331] Additional plants that can be used include safflower (*Carthamus tinctorius*), sweet potato (*Ipomoea batatas*), 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.

[0332] Vegetables that can be used include tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus spp.*), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea (*Rhododendron spp.*), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa spp.*), tulips (*Tulipa spp.*), daffodils (*Narcissus spp.*), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum.

[0333] Conifers that may be used include pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*); Douglas fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow cedar (*Chamaecyparis nootkatensis*).

[0334] In certain embodiments of the disclosure, a fertile plant is a plant that produces viable male and female gametes and is self-fertile. Such a self-fertile plant can produce a progeny plant without the contribution from any other plant of a gamete and the genetic material comprised therein. Other embodiments of the disclosure can involve the use of a plant that is not self-fertile because the plant does not produce male gametes, or female gametes, or both, that are viable or otherwise capable of fertilization.

[0335] The present disclosure finds use in the breeding of plants comprising one or more introduced traits, or edited genomes.

[0336] A non-limiting example of how two traits can be stacked into the genome at a genetic distance of, for example, 5 cM from each other is described as follows: A first plant comprising a first transgenic target site integrated into a first DSB target site within the genomic window and not having the first genomic locus of interest is crossed to a second transgenic plant, comprising a genomic locus of interest at a different genomic insertion site within the genomic window and the second plant does not comprise the first transgenic target site. About 5% of the plant progeny from this cross will have both the first transgenic target site integrated into a first DSB target site and the first genomic locus of interest integrated at different genomic insertion sites within the genomic window. Progeny plants having both sites in the defined genomic window can be further crossed with a third transgenic plant comprising a second transgenic target site integrated into a second DSB target site and/or a second genomic locus of interest within the defined genomic window and lacking the first transgenic target site and the first genomic locus of interest. Progeny are then selected having the first transgenic target site, the first genomic locus of interest and the second genomic locus of interest integrated at different genomic insertion sites within the genomic window. Such methods can be used to produce a transgenic plant comprising a complex trait locus having at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31 or more transgenic target sites integrated into DSB target sites and/or genomic loci of interest integrated at different sites within the genomic window. In such a manner, various complex trait loci can be generated.

[0337] 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, applications, 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

[0338] 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.

[0339] A novel construct design was developed to test the efficacy of template-directed repair of double-strand breaks (DSBs) generated by a CRISPR/Cas9 system, using of multiple copies of a donor template in single construct (donor DNA concatemer) that were each flanked by CRISPR/Cas9 targets on either side. With this “concatemer” design, a Cas enzyme-generated DNA DSB can be created at each target in between template copies. The freed templates, thus generated, were available in multiple copies for the host DNA repair system.

[0340] Although plant cells were used for these examples, any cell may be used, such as but not limited to human cells, murine cells, other mammalian cells, insect cells, other animal cells, fungal cells, protist cells, bacterial cells, and archaeobacterial cells.

Example 1: Vector Construction

[0341] The three components of the vector introduced into the host cell included the Cas endonuclease (Cas9 was used as an exemplary Cas endonuclease; any Cas endonuclease may be used), at least one cognate guide RNA, and a heterologous DNA molecule that comprised either a template for directed repair of the double-strand break (DSB), or a polynucleotide for insertion at the DSB site.

[0342] The rationale of the strategy aimed to generate more repair templates/donor molecules available during editing process inside the cellular nucleus. Unique aspects included the combination of cloning of multiple copies of the repair templates/donor DNA molecules which were excisable inside the cells. These multiple copies are present as tandemly concatenated repeats in a “sequence unit” (or “unit”).

[0343] Each individual “sequence unit” comprises a heterologous polynucleotide (a donor DNA molecule for insertion at the DSB, or a polynucleotide template for template-directed repair of the DSB) that is flanked by sequence identical to a sequence at the cleavage site of the target polynucleotide (“target site”) where a guide RNA is capable of hybridizing, to

create a “unit” comprising, in this order: a Target Site sequence, a Donor/Template sequence, a Target Site sequence. Each sequence unit also may further comprise a PAM sequence that is capable of being recognized by a Cas endonuclease, such that the target sequences that are capable of hybridizing to the guide RNA or guide RNAs that is/are provided with the Cas endonuclease become cleaved after recognition by the Cas endonuclease and binding of the guide RNA.

[0344] The heterologous polynucleotide may optionally be flanked inside of the flanking target sites by another set of polynucleotides that share homology to some region of the target site (“homology regions”, or “HR”, or “HDR” regions) that is not the cleavage site, to create a “unit” comprising, in this order: a Target Site sequence, a Homology Region, a Donor/Template sequence, a Homology Region, a Target Site sequence.

[0345] Homology Regions are at least 10 nucleotides in length and share at least 80% sequence identity with a region of the target polynucleotide that is not the cleavage site of the target polynucleotide. The flanking Homology Regions, if present, may be identical to each other or non-identical to each other.

[0346] While a longer circular plasmid harboring all components can be delivered to plant tissue by gene gun or other methods such as but not limited to *Agrobacterium* infection, *Ochrobactrum* infection, vacuum infiltration, viral infection, electroporation, etc., releasing of shorter free repair templates upon excision of the same gRNA inside the cells is likely to provide greater number of templates and as well as these may be more accessible to the DNA target site.

[0347] The repair templates were synthesized with flanking target sequence at 5' end and cloned in gateway compatible entry vector. Other components (encoding Cas9 and gRNA) were initially cloned in entry vectors and finally all entry vectors were cloned into a single binary vector backbone. The binary clone was sequenced by NGS and plasmid DNA was isolated and purified before transformation.

Example 2: Transformation, Regeneration, and Selection

[0348] In the following examples, plant cells were transformed by methods known in the art. Exemplary transformation methods include particle bombardment and *Agrobacterium*-mediated transformation.

Rice

[0349] Seeds from two rice inbred lines (Line A and Line B) were sterilized in 75% ethanol for 2-3 minutes and washed thoroughly with water and incubated in 4% sodium hypochlorite for 10 minutes. The seeds were then washed 5 times with water and dried completely at room temperature. The dried seeds were inoculated on callus inducing media and the plates were incubated at 28 °C in light for 5-7 days. After that the proliferating calli obtained from rice seeds were placed on osmotic media for 4 hours before being bombarded with DNA:gold particles.

[0350] Sufficient amount of gold particles (amount of gold particles depends on the number of bombardments) were weighed and placed in 2.0 ml Eppendorf tubes. One ml of 100% ethanol was added to the tube and sonicated for 30 sec before centrifuging for 1 min. The pellet containing the gold particles was resuspended in 1 ml of 100% ethanol, vortexed for 30 seconds and centrifuged again. This step was repeated twice before resuspending the pellet in 1 ml of sterile water. Fifty µl of gold particle suspension was aliquoted into Eppendorf tubes and stored at 4 °C.

[0351] Five µg of DNA, 50 µl of 2.5 mM CaCl₂ and 20 µl of 0.1 M spermidine were added to 50 µl of gold particle suspension; vortexed for 1-2 minutes and allowed the mixture to settle down for 5 minutes. The tubes were centrifuged for 2 minutes before discarding the supernatant. The pellet was resuspended in 40 µl of 100% ethanol and mixed gently by vortexing and 5 µl of sample was quickly dispensed onto macrocarrier disks and dried completely.

[0352] Macro carrier disk carrying DNA:gold particle prep were loaded onto macro carrier disk holder and stopping screen was placed on top of the disk. Manufacturer's instructions were followed to deliver DNA:gold particles onto tissue samples which were placed on osmotic medium. After bombardment, the tissue samples were kept on the same osmotic medium for 24 h at 32 °C in dark.

[0353] After 24 hours post bombardment, the samples were sub-cultured on to resting media and kept in dark at 28 °C for 5 days. The cultures were then transferred to selection media containing Hygromycin as selectable agent. After 3-4 selection cycles, proliferating hygromycin resistant and Zs-Yellow positive callus variants were sub-cultured onto regeneration media and then to rooting and hardening media to obtain stable lines. Each independent line was transferred

to an individual pot in greenhouse and samples were collected to perform molecular and phenotypic analysis.

Maize

[0354] Ears of a maize (*Zea mays* L.) cultivar were surface-sterilized for 15-20 min in 20% (v/v) bleach (5.25% sodium hypochlorite) plus 1 drop of Tween 20 followed by 3 washes in sterile water. Immature embryos (IEs) were isolated from ears and were placed in 2 ml of the *Agrobacterium* infection medium with acetosyringone solution. The optimal size of the embryos varies based on the inbred, but for transformation with WUS2 and ODP2 a wide size range of immature embryo sizes could be used. The solution was drawn off and 1 ml of *Agrobacterium* suspension was added to the embryos and the tube vortexed for 5-10 sec. The microfuge tube was allowed to stand for 5 min in the hood. The suspension of *Agrobacterium* and embryos were poured onto 710I co-cultivation medium. Any embryos left in the tube were transferred to the plate using a sterile spatula. The *Agrobacterium* suspension was drawn off and the embryos placed axis side down on the media. The plate was sealed with Parafilm M® film (moisture resistant flexible plastic, available at Bemis Company, Inc., 1 Neenah Center 4th floor, PO Box 669, Neenah, WI 54957) and incubated in the dark at 21 °C for 1-3 days of co-cultivation.

[0355] Embryos were transferred to resting medium (605T medium) without selection. Three to seven days later, they were transferred to either a selection medium for event selection, or to maturation medium (289Q medium) supplemented with a selective agent.

[0356] It is contemplated that other bacterium-mediated transformation methods can be used, for example, with *Ochrobactrum*.

[0357] Sixteen days later, embryos with healthy somatic embryos generated in Example 2 were moved onto a regeneration medium.

[0358] In one example, embryos were treated with *Agrobacterium* and one day later selected embryos were moved onto 605T medium (no selection for the first week), 605T medium with 0.1 mg/l ethametsulfuron with AA (early selection with AA) or 605T medium with 0.1 mg/l ethametsulfuron (early selection with no AA), respectively. For the next transfer, selected embryos were moved onto their respective maturation media. For the final transfer to rooting medium, selected plantlets of individual events were moved. For this experiment, the total elapsed time from *Agrobacterium* infection to the greenhouse was 48 days.

[0359] In another example, embryos were treated with *Agrobacterium* in liquid for 5 minutes and then co-cultured for one day on 710I medium. At this point, selected embryos were moved onto 605T medium, 605T medium with 0.1 mg/l ethametsulfuron with AA or or 605T medium with 0.1 mg/l ethametsulfuron, respectively. Twelve days later, the embryos on 605T were split onto either 289Q medium with 0.1 mg/l imazapyr or onto 289Q medium with 0.5 mg/l imazapyr. The embryos from both the 605T medium with 0.1 mg/l ethametsulfuron with AA and 605T medium with 0.1 mg/l ethametsulfuron were moved onto 289Q (no further selection). After maturation, healthy plantlets (events) were transferred to rooting medium 13158H, with selected events being moved from the above maturation treatments, respectively.

Soy

[0360] Standard protocols for particle bombardment (Finer and McMullen, 1991, In Vitro Cell Dev. Biol. – Plant 27:175-182), *Agrobacterium*-mediated transformation (Jia et al., 2015, Int J. Mol. Sci. 16:18552-18543; US20170121722 incorporated herein by reference in its entirety), or *Ochrobactrum*-mediated transformation (US20180216123 incorporated herein by reference in its entirety) for soybean can be used with the methods of the disclosure.

[0361] Soybean transformation is done essentially as described by Paz *et al.* ((2006) Plant Cell Rep 25:206-213) and US Patent 7,473,822. Mature seed from soybean lines are surface-sterilized for 16 hrs using chlorine gas, produced by mixing 3.5 mL of 12 N HCl with 100 mL of commercial bleach (5.25% sodium hypochloride), as described by Di *et al.* ((1996) Plant Cell Rep 15:746-750). Disinfected seeds are soaked in sterile distilled water at room temperature for 16 hrs (100 seeds in a 25×100 mm petri dish).

[0362] A volume of 10 mL of *Ochrobactrum haywardense* H1 NRRL Deposit B-67078 further containing vector PHP70365 (SEQID NO: 106) suspension at OD₆₀₀=0.5 in infection medium containing 300 µM acetosyringone is added to the soaked seeds. The seeds are then split by cutting longitudinally along the hilum to separate the cotyledons, and the seed coats, primary shoots, and embryonic axes are removed in *Ochrobactrum haywardense* H1 NRRL Deposit B-67078 suspension, thereby generating half-seed explants. The half-seed explants are placed flat side down in a deep plate with 4 mL fresh *Ochrobactrum*/infection media with no overlapping of cotyledons. The plates are sealed with parafilm ("Parafilm M" VWR Cat#52858), then sonicated (Sonicator-VWR model 50T) for 30 seconds. After sonication, half-seed explants are transferred to a single layer of autoclaved sterile filter paper (VWR#415/Catalog # 28320-020) onto co-

cultivation solid medium (18-22 explants per plate; flat side down). The plates are sealed with Micropore tape (Catalog # 1530-0, 3M, St. Paul, MN)) and incubated under dim light (5-10 $\mu\text{E}/\text{m}^2/\text{s}$, cool white fluorescent lamps) for 16 hrs at 21°C for 5 days.

[0363] Regeneration methods are carried out according to those disclosed in WO2017040343A1 (published 09 March 2017). After co-cultivation, the half-seed explants are washed in liquid shoot induction (SI) medium once then the explants are cultured on shoot induction medium solidified with 0.7% agar in the absence of selection. The base of the explant (*i.e.*, the part of the explant from where the embryonic axis was removed) is embedded in the medium, facing upwards. Shoot induction is carried out in a Percival Biological Incubator at 24°C with a photoperiod of 18hrs and a light intensity of 130-160 $\mu\text{E}/\text{m}^2/\text{s}$. After 14 days, the explants are transferred to fresh shoot induction medium containing 3 mg/L bialaphos. The half seed explants are transferred to fresh medium every two weeks. After four weeks of culture on shoot induction medium, explants are transferred to shoot elongation (SE) medium containing 5 mg/L bialaphos (Table 10). Six to ten weeks later, elongated shoots (>1-2 cm) are isolated and transferred to rooting medium (Table 10) containing 1 mg/L bialaphos.

Canola

[0364] *Agrobacterium*-mediated transformation and regeneration is performed as described in (De Block, M., et al. (1989). "Transformation of Brassica napus and Brassica oleracea Using Agrobacterium tumefaciens and the Expression of the bar and neo Genes in the Transgenic Plants." Plant Physiology 91(2): 694-701).

Example 3: HDR of a double-strand break in rice

[0365] *Grain length3* (*GL3*) gene (Locus ID: Os03g44500) was selected to test the concatemer strategy, as one example. Any genomic locus of any organism could be used.

[0366] *GL3* (Grain length3) encodes a Ser/Thr protein phosphatase with Kelch-like repeat domain (OsPPKL1). A rare allele, called *qgl3*, leads to a long grain phenotype by an aspartate to glutamate transition in a conserved AVLDT motif of the second Kelch domain in OsPPKL1. Two SNPs are important in *GL3.1-WY3*: aspartic acid to glutamic acid (D364E; 1092C-A) and histidine to tyrosine (H499Y; 1495C-T) substitutions. Three inbred lines (Lines A, B, and C) were identified that comprised the second SNP ((H499Y; 1495C-T). A strategy was designed to edit native *GL3.1* allele to mimic *GL3.1-WY3* by "C" to "A" substitution (GAC-

GAA) resulting aspartic acid to glutamic acid (D364E; 1092C-A). A 240bp repair template with 3 SNPs- Editing SNP for Aspartate to Glutamate (GAC to GAA), Edit1 T to A (Ala) & Edit2 A to G (Thr) for heterogeneity in donor. Each repair template consisted of the 240bp donor/template flanked by the Target Site guide targets, to release the individual fragments *in vivo*.

[0367] A single nucleotide edit was designed in the GL3 gene to substitute one amino acid. Two additional bases covering seed sequence of the gRNA target were modified without altering the encoding pattern (codon degeneracy), to generate heterogeneity in HDR product, so that the product remained stable after the HDR process.

[0368] The objective of the experiment was to enhance efficiency and/or frequency homology directed repair (HDR) of DSBs generated using Cas9/gRNA technology. It can be used to test for enhanced efficiency in SDN2 (template-directed repair of the double-strand break site) as well as SDN3 (integration of a heterologous polynucleotide at the double-strand break site). FIGs. 2 and 3 depict the general vector design.

[0369] Rice calli were transformed as described above, with Construct #1 (for the concatemer repair template), Construct #2 (for the single copy repair template with one flanking homology region), or Construct #3 (for the single copy repair template with two flanking homology regions), as depicted in FIGs. 4, 5, and 6, respectively.

[0370] Sequences for the constructs included those described in Table 1.

Table 1: Sequences used in the rice experiments

SEQID	Description	Type	Organism
1	U3 PolIII Chr4 promoter (OSU3POLIII CHR4 PRO)	DNA	Oryza sativa
2	DNA sequence encoding the VT domain of the guide RNA for the GL3 target site	DNA	Oryza sativa
3	Guide RNA sequence	DNA	Artificial
4	Ubiquitin promoter (UBI1ZM PRO)	DNA	Zea mays
5	Ubiquitin 5'UTR (UBI1ZM 5UTR)	DNA	Zea mays
6	Ubiquitin Intron 1 (UBI1ZM INTRON1)	DNA	Zea mays
7	Cas9 CDS	DNA	Streptococcus pyogenes
8	PinII terminator	DNA	Solanum tuberosum
9	GL3 Homology Region 1(HDR-OS-GL3 FRAG2)	DNA	Oryza sativa
10	GL3 target site donor/template	DNA	Oryza sativa

11	GL3 Homology Region 2(HDR-OS-GL3 FRAG3)	DNA	Oryza sativa
12	35S promoter (CAMV35S PRO-V4)	DNA	Cauliflower mosaic virus
13	HYG-Z5 Yellow N1 selectable marker	DNA	Artificial
14	Primers used to clone gRNA in to Rice U3 at Slot1 Aar1 site	DNA	Artificial
15	Primers used to clone gRNA in to Rice U3 at Slot1 Aar1 site	DNA	Artificial
16	BamH1 Linker	DNA	Artificial
17	HindIII Linker	DNA	Artificial
18	Primer for cloning of the heterologous polynucleotide at Nco1	DNA	Artificial
19	Primer for cloning of the heterologous polynucleotide at Nco1	DNA	Artificial
20	Primer for cloning of the heterologous polynucleotide at BamH1	DNA	Artificial
21	Primer for cloning of the heterologous polynucleotide at BamH1	DNA	Artificial
22	Primer for cloning of the heterologous polynucleotide at BglII	DNA	Artificial
23	Primer for cloning of the heterologous polynucleotide at BglII	DNA	Artificial
24	Primer for cloning of the heterologous polynucleotide at HindIII	DNA	Artificial
25	Primer for cloning of the heterologous polynucleotide at HindIII	DNA	Artificial
26	Primer for cloning of the heterologous polynucleotide at Stu1	DNA	Artificial
27	Primer for cloning of the heterologous polynucleotide at Stu1	DNA	Artificial
28	Primer for cloning of the heterologous polynucleotide at Stu1	DNA	Artificial
29	Primer for cloning of the heterologous polynucleotide at Stu1	DNA	Artificial

[0371] The construct with the concatemer comprised 4 tandem repeats of the donor/template. The “Target Site” was a DNA sequence identical to the site sequence of the polynucleotide targeted for cleavage by the Cas endonuclease, and which hybridized with the guide RNA VT domain sequence (provided in this example as SEQID NO:2).

[0372] Following transformation, the repair concatemer was cleaved by Cas9/gRNA(s) complexes, releasing free single units of repair template which facilitated enhancement of HDR by providing higher number of repair templates which were used for repair of genomic DNA double stranded breaks generated by the same Cas9/gRNA(s) complexes. In our experiments, we used a single guide to generate DSBs and evaluated frequencies of HDR events using the concatemer and single donor repair templates. DNA constructs were delivered by particle bombardment to the plant tissues.

[0373] NGS analysis was performed to identify all three edits in the target region and to confirm perfect HDR events (edits). The edits were confirmed in two generations – T0 and T1.

[0374] As shown in Table 2, HDR-based SDN2 efficiency was significantly improved when a concatemerized donor/repair template was used, compared to single copy repair template.

The HDR efficiency was increased up to 3.2% in Line A and 7.7% in Line B (both T0 mono-allelic and chimeric), while single copy repair template failed to produce HDR-based edited events. Line C had too few variants screened to produce reliable results.

Table 2: HDR efficiency of concatemer donor/template DNA of rice CRISPR/Cas constructs

Template	Inbred Germplasm	Construct	Variants screened	Variants with HDR (3'/5' assay)	Efficiency of HDR (%)
Concatemer repair template	Line A	1	93	3	3.2%
	Line B	1	26	2	7.7%
	Line C	1	9	0	0%
Single copy repair template	Line A	2	42	0	0%

[0375] Table 3A shows the genomic DNA sequence of the Rice GL3 locus that was targeted for cleavage, and the positions of the SNPS that were created.

Table 3A: Rice GL3 locus DNA template for DSB repair (SNPs indicated with bold underlined font), with the corresponding Wild Type locus nucleotides at the SNP position indicated below

Fragment of the template (heterologous polynucleotide in the construct) that shows the SNPs (First 50 nucleotides of SEQID NO:10)	TCTATCTGGTTTCAGTGCTAGAA <u>A</u> AC <u>G</u> GC <u>A</u> AGCTGGAGTCTGGTGCGACACAA
Wild Type Locus	C A T

[0376] Table 3B shows the T0 genomic variants that were generated using the concatemer repair template.

Table 3B: Target site NGS reads for concatemer-generated variants in rice

Inbred Line	Target site feature represented in HDR read/allele	Target site read- NGS
A	*	AGAAAC
	*	GCAGCT
		GCA [+T] GCT
		GCTGCT
		[+T] G [-CAG] CT
		GCAGCT
A	*	AGAAAC
		AGA [+C] A [+C] A [-C]
		GCA [+T] GCT

	*	GCAGCT
		GCAGCT
A	*	AGAAAC
		AGA [+C] A [+C] A [-C]
	*	GCAGCT
		GC [+T] tGCT
B	*	AGAAAC
		AGA [+C] A [+C] A [-C]
	*	GCAGCT
		GC [+T] tGCT
B	*	AGAAAC
		AGA [+C] A [+C] A [-C]
		AGA [+C] A [-AC]
		AGAAAC
		GctGCT
		[-G] CA [+GT] GCT
		GCAGCT

[0377] Table 4 shows NGS sequence alignment of wild type and SDN2 variants of Rice Line A that showed editing of the target SNP along with two additional nucleotide edits.

Table 4: SDN2 variants of Rice Line A

Sample	Sequence
Variant 1	AACGGCA
Variant 2	AACGGCA
Variant 3	AACGGCA
Line A WT	CACAGCT
WT Reference Line	CACAGCT

[0378] Table 5 shows NGS sequence alignment of wild type and SDN2 variants of Rice Line B that showed editing of the target SNP along with two additional nucleotide edits.

Table 5: SDN2 variants of Rice Line B

Sample	Sequence
Variant 1	AACGGCA
Variant 2	AACGGCA
Line B WT	CACAGCT

[0379] These results demonstrate that the frequency of HDR of a DSB is improved by providing multiple copies of a donor DNA fragment or template DNA to the DSB site as part of a concatemer, which is cleaved on either side of each donor/template by a guide RNA that

recognizes and directs a Cas endonuclease to cleave target site sequences that flank each donor/template unit in the concatemer.

Example 4: HDR of a double-strand break in maize

[0380] The objective of the experiment was to enhance SDN2 efficiency using multiple (five) copies of donor DNA interspersed with gRNA target sites, similar to the approach described above. FIGs. 7A, 7B, and 7C depict the general vector designs.

[0381] In this example, edit of a single nucleotide at one maize target site was demonstrated to create 4 SNPs.

[0382] The objective of the experiment was to enhance homology dependent repair (HDR) efficiency using Cas9/gRNA technology. It can be used to test for enhanced efficiency in SDN2 (template-directed repair of the double-strand break site) as well as SDN3 (integration of a heterologous polynucleotide at the double-strand break site). FIGs. 7A, 7B, and 7C depict the general vector designs.

[0383] Maize embryos were transformed as described above using *Agrobacterium*-mediated delivery of T-DNA, with Construct #4 (single-copy of donor/template, with inverted target site sequences), Construct #5 (single-copy of donor/template, with monodirectional target site sequences), and Construct #6 (concatemer construct with monodirectional target site sequences), as depicted in FIGs. 8, 9, and 10, respectively.

[0384] Each vector comprised morphogenic factors (WUS and ODP2 under Axig and PLTP promoters, respectively), *cas9* driven by the Ubiquitin promoter, gRNA for the Target Site (TS) sequence operably linked to the maize U6 polymerase III promoter, and a selectable marker gene – neomycin phosphotransferase II (NptII) under the regulation of a Ubiquitin promoter.

[0385] Sequences for the constructs included those described in Table 6.

Table 6: Sequences used in the maize experiments

SEQ ID	Description	Type	Organism
30	AXIG1 promoter	DNA	<i>Zea mays</i>
31	WUS2 morphogenic factor (ALT1)	DNA	<i>Zea mays</i>
32	NOS terminator	DNA	<i>Agrobacterium tumefaciens</i>
33	PLTP promoter	DNA	<i>Zea mays</i>
34	PLTP 5' UTR	DNA	<i>Zea mays</i>
35	ODP2 morphogenic factor (ALT1)	DNA	<i>Zea mays</i>
36	T28 terminator	DNA	<i>Oryza sativa</i>
37	Ubiquitin promoter	DNA	<i>Zea mays</i>
38	Ubiquitin 5' UTR	DNA	<i>Zea mays</i>
39	Ubiquitin intron 1	DNA	<i>Zea mays</i>
40	T-antigen monopartite nuclear localization signal	DNA	<i>Simian virus 40</i>
41	Cas9 Exon 1	DNA	<i>Streptococcus pyogenes</i>
42	LS1 Intron2	DNA	<i>Solanum tuberosum</i>
43	Cas9 Exon 2	DNA	<i>Streptococcus pyogenes</i>
44	C-terminal bipartite nuclear localization signal from VirD2 endonuclease	DNA	<i>Agrobacterium tumefaciens</i>
45	PinII terminator	DNA	<i>Solanum tuberosum</i>
46	U6 PolIII Chr8 Promoter	DNA	<i>Zea mays</i>
47	DNA sequence encoding the VT domain of the guide RNA for Zm target site	DNA	<i>Zea mays</i>
48	Guide RNA sequence	DNA	<i>Artificial</i>
49	Zm gRNA target site sequence (CR1)	DNA	<i>Zea mays</i>
50	Zm Homology Region 1	DNA	<i>Zea mays</i>
51	Zm Target Site donor/template	DNA	<i>Artificial</i>
52	Zm Homology Region 2	DNA	<i>Zea mays</i>
53	NPTII selectable marker	DNA	<i>Solanum tuberosum</i>

[0386] The construct with the concatemer comprised 5 tandem repeats of the donor/template. The “Target Site” was a DNA sequence identical to the target site sequence of the polynucleotide targeted for cleavage by the Cas endonuclease, and which hybridized with the guide RNA VT domain sequence (provided in this example as SEQID NO:47).

[0387] Following transformation, the repair concatemer was cleaved by Cas9/gRNA(s) complexes, releasing free single units of repair template which facilitated enhancement of HDR by providing higher number of repair templates which were used for repair of genomic DNA double stranded breaks generated by the same Cas9/gRNA(s) complexes..

[0388] NGS analysis was performed to identify edits in the target region and to confirm perfect HDR. Table 7 shows the genomic DNA sequence of the maize locus that was targeted for cleavage, and the positions of the SNPS that were created.

Table 7: Maize locus DNA template for DSB repair (SNPs indicated with bold underlined font), with the corresponding Wild Type locus nucleotides at the SNP positions indicated below

(the cut site in the corresponding genomic locus target site is indicated by a slash (/), and the PAM site is indicated by a box)

Template (heterologous polynucleotide in the construct) that shows the SNPs (SEQID NO:51)	GCATAATGAGG <u>G</u> <u>T</u> <u>A</u> <u>G</u> <u>T</u> GGA / TGA <u>T</u> <u>G</u>
Wild Type Locus	A C A G

[0389] As shown in Table 8, template-directed repair efficiency was significantly improved (more than doubled) when a concatemer repair template was used as a donor molecule, compared to single copy repair template. There was no significant difference attributable to the target site sequence orientation.

Table 8: Results from the maize experiments with concatemer donor/template DNA

Experiment	Construct ID	# T0 plants analyzed	# T0 plants with 4 nt edits	% editing frequency
single donor/template, PAMs opposite direction	4	752	9	1.2%
single donor/template, PAMs same direction	5	752	13	1.7%
concatemer donor/template, PAMs same direction	6	752	25	3.3%

[0390] These examples demonstrate that the frequency of homologous recombination/homology directed repair of a double-strand break in a target polynucleotide is increased when the template or donor DNA molecule is presented as multiple copies in a concatemer, wherein each copy is flanked by a sequence sharing homology to the target site and capable of hybridizing with a guide RNA that forms a complex with a Cas endonuclease to effect cleavage and release of the donors/templates.

Example 5: Higher copy number of the template, as well as longer templates, improve HDR outcomes

[0391] Further experiments were conducted to investigate outcomes with Agrobacterium-mediated transformation of vectors using either multiple copies of a template (FIG. 11B) or a template of longer length (FIG. 11C), as compared to a control of a single copy of a donor of 200 nt. In all cases, the donor polynucleotide was flanked by regions of homology to the target site. The experiment was done in a large number of embryos to validate statistical analyses.

[0392] As shown in Table 9, increasing the number of templates increased SDN2 frequency by two-fold. Longer donor DNA segments increased SDN2 frequency by 50%. Additionally, longer DNA generated a higher frequency of complete edits (fewer chimeric plants, better transmission).

Table 9: Results from the maize experiments with concatemer donor/template DNA

Template DNA	# T0 plants analyzed	EDITS by qPCR	EDITS by NGS		
			Total Edits, ≥20%	20-40%	>40%
200 nt	1789	45 (2.5%)	22 (1.3%)	15 (68%)	7 (32%)
5x 200 nt	1750	70 (4.0%)	44 (2.6%)	28 (63%)	16 (37%)
828 nt	1680	49 (2.9%)	34 (1.9%)	12 (35%)	22 (65%)

WE CLAIM:

1. A method of alteration of a target polynucleotide, comprising:
 - (a) providing to the target polynucleotide:
 - (i) a Cas endonuclease,
 - (ii) a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in the target polynucleotide, and
 - (iii) a plurality of sequence units, wherein each sequence unit comprises a heterologous polynucleotide, wherein each sequence unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA of (a)(ii);
 - (b) cleaving the plurality of sequence units of (a)(iii) with the complex of (a)(ii), releasing the heterologous polynucleotides;
 - (c) identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units of (a)(iii).
2. The method of Claim 1, wherein the heterologous polynucleotide is a donor DNA molecule that is inserted into the double-strand break.
3. The method of Claim 1, wherein the heterologous polynucleotide is a template DNA molecule that directs the repair of the double-strand break.
4. The method of Claim 1, wherein each heterologous polynucleotide is flanked by a set of second flanking sequences, each of which is at least 10 nucleotides in length and share at least 80% identity to a sequence near the target polynucleotide, and wherein said set of second flanking sequences is flanked by the set of first flanking sequences.
5. The method of Claim 1, wherein a plurality of different guide RNA molecules are provided in (a)(ii), and wherein the first flanking sequences of (a)(iii) are capable of hybridizing to the plurality of different guide RNA molecules.
6. The method of Claim 1, wherein the target polynucleotide is in a cell.
7. The method of Claim 6, wherein the cell is a plant cell.

8. The method of Claim 1, wherein the plurality of sequence units is stably integrated into the plant cell.
9. The method of Claim 1, wherein the guide RNA molecule is provided via particle bombardment.
10. The method of Claim 1, wherein the Cas endonuclease and guide RNA are provided as a ribonucleoprotein complex.
11. The method of Claim 1, wherein the frequency of homologous recombination repair at the double-strand-break site of the target polynucleotide is greater than the rate of non-homologous end joining repair at that same site.
12. A method of altering a phenotypic trait in a plant, comprising:
 - (a) providing to a plant cell a set of molecules comprising:
 - (i) a Cas endonuclease,
 - (ii) a guide RNA molecule that forms a complex with the Cas endonuclease to create a double-strand break in a target polynucleotide in the plant cell, and
 - (iii) a plurality of sequence units, each comprising a heterologous polynucleotide, wherein each unit is flanked by a set of first flanking sequences, each of which is capable of hybridization with the guide RNA of (a)(ii);
 - (b) cleaving the plurality of sequence units of (a)(iii) with the complex of (a)(ii), releasing the heterologous polynucleotides;
 - (c) identifying at least one nucleotide insertion, deletion, substitution, or modification of the sequence of the target polynucleotide, or any combination of the preceding, as compared to the sequence of the target polynucleotide prior to the providing of the plurality of sequence units of (a)(iii); and
 - (d) obtaining a plant from the plant cell;

wherein the plant comprises an alteration of at least one phenotypic trait as compared to an isoline plant that was not provided the set of molecules of (a).
13. The method of Claim 12, wherein each heterologous polynucleotide is flanked by a set of second flanking sequences, each of which is at least 10 nucleotides in length and share at

- least 80% identity to a sequence near the target site, and wherein said set of second flanking sequences is flanked by the set of first flanking sequences.
14. The method of Claim 1 or Claim 12, wherein the cell is a monocot plant cell.
 15. The method of Claim 13, wherein the monocot plant cell is selected from the group consisting of: maize, rice, sorghum, barley, and wheat.
 16. The method of Claim 1 or Claim 12, wherein the cell is a dicot plant cell.
 17. The method of Claim 16, wherein the dicot plant cell is selected from the group consisting of: soy, canola, cotton, sugarcane, and *Arabidopsis*.
 18. The method of Claim 12, wherein the phenotypic trait is average yield.
 19. The method of Claim 12, further comprising obtaining a tissue, part, or reproductive element of the plant of (c), wherein the tissue, part, or reproductive element comprises the at least one nucleotide insertion, deletion, substitution, or modification of the sequence, or any of the preceding, of the target polynucleotide of the plant from which it was obtained.
 20. A progeny plant obtained or derived from the method of Claim 19, wherein the progeny plant comprises said nucleotide insertion, deletion, substitution, modification, or any combination of the preceding.

FIG. 1

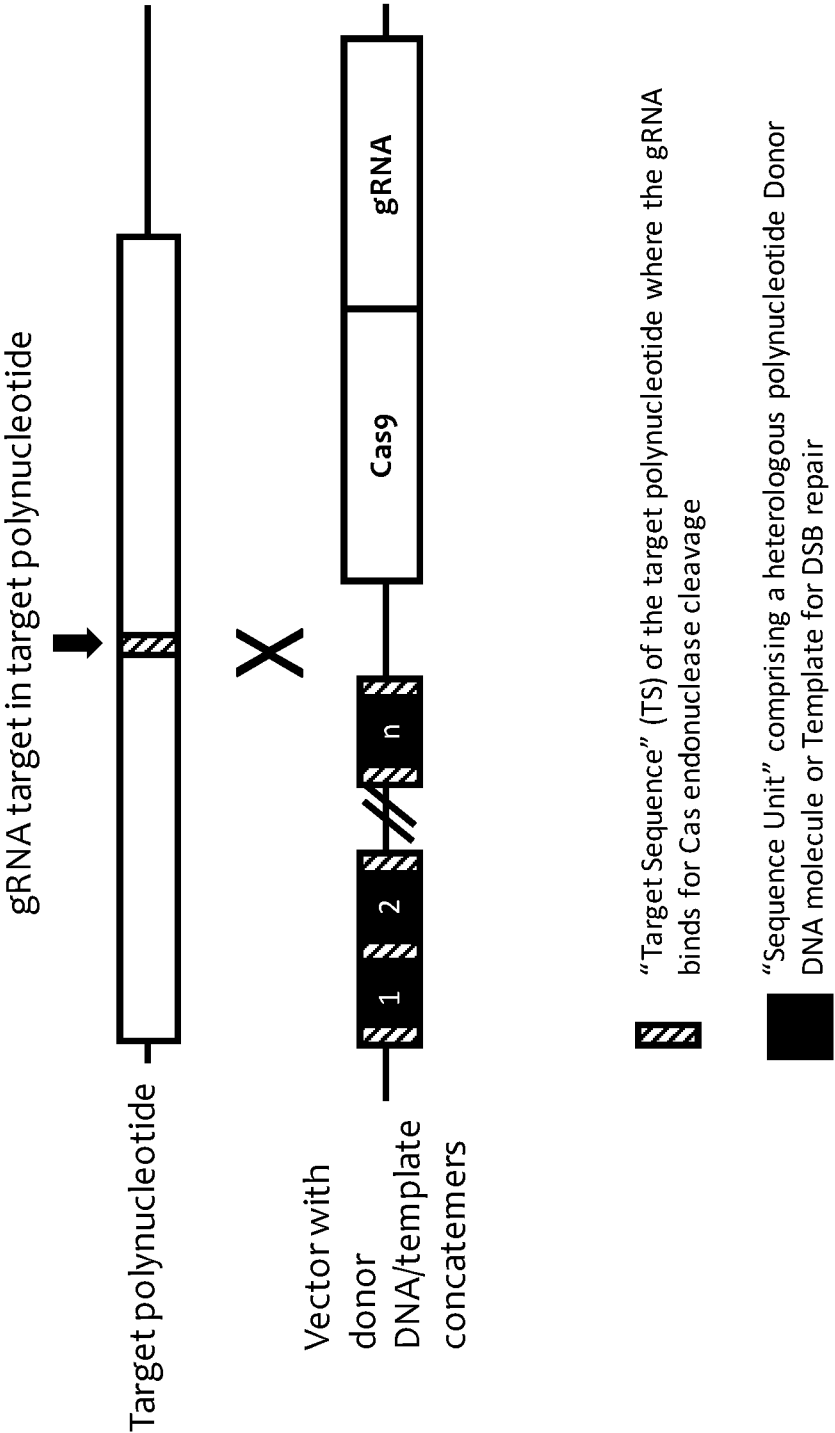


FIG. 2

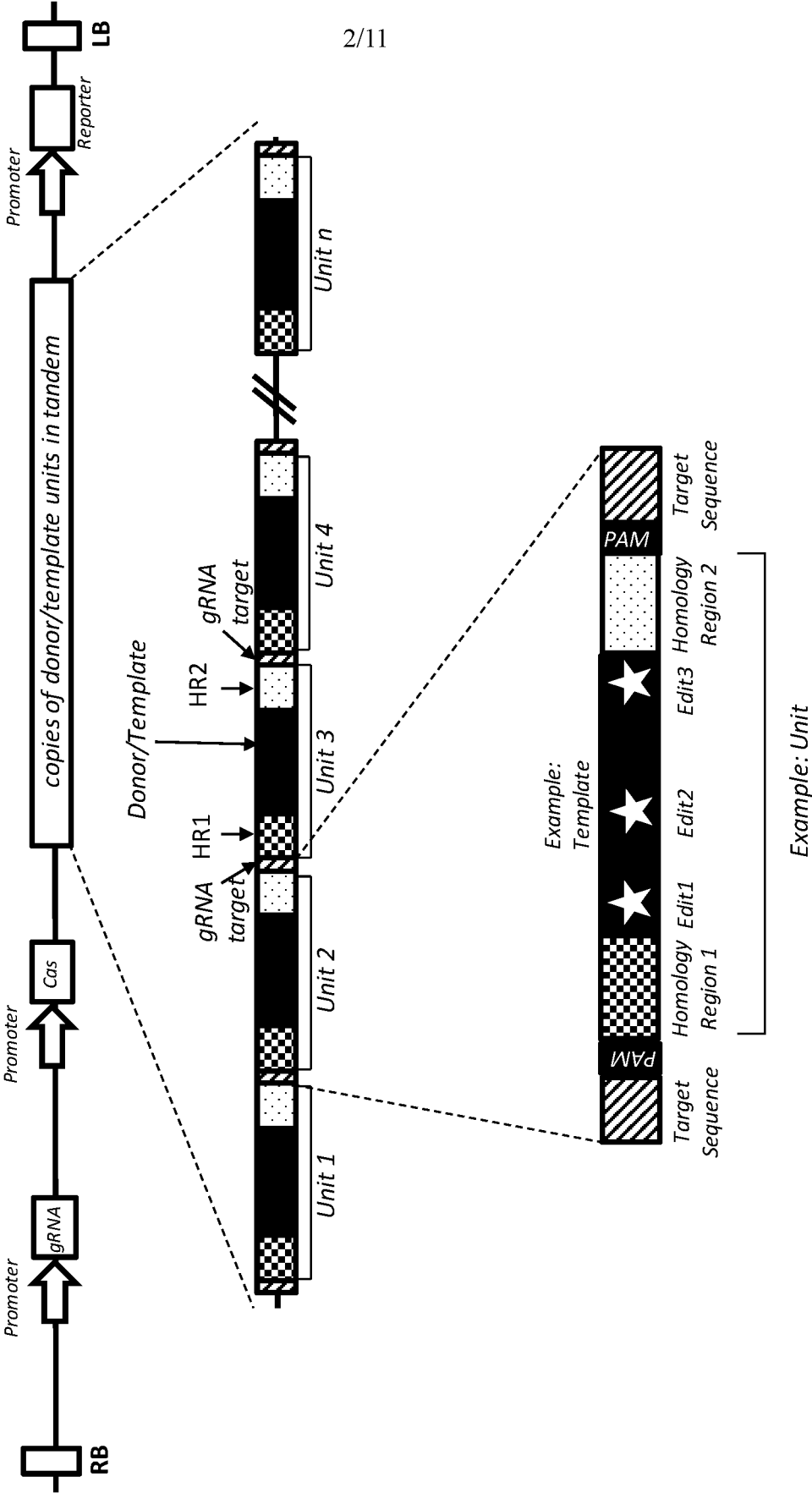
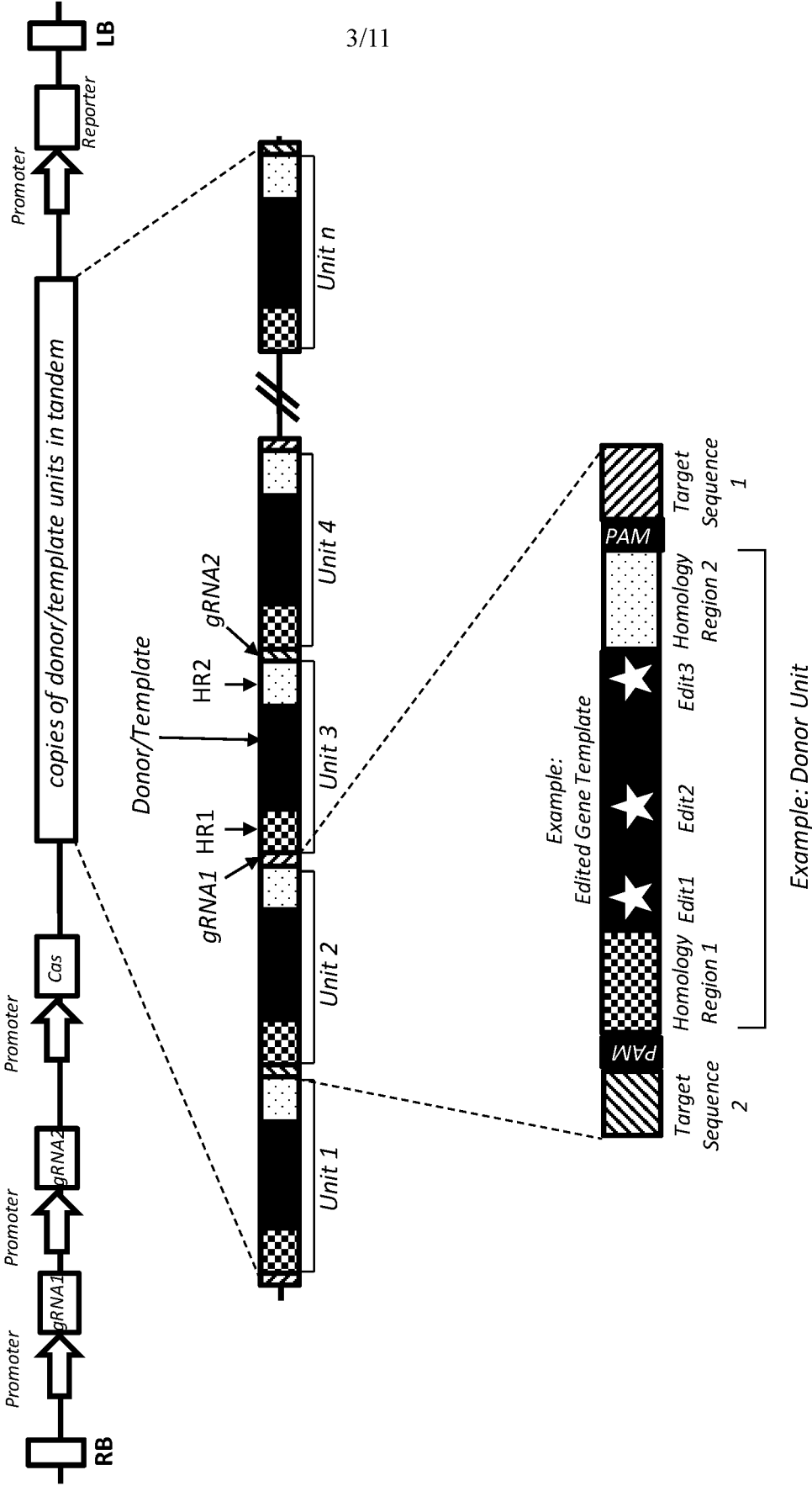


FIG. 3



4/11

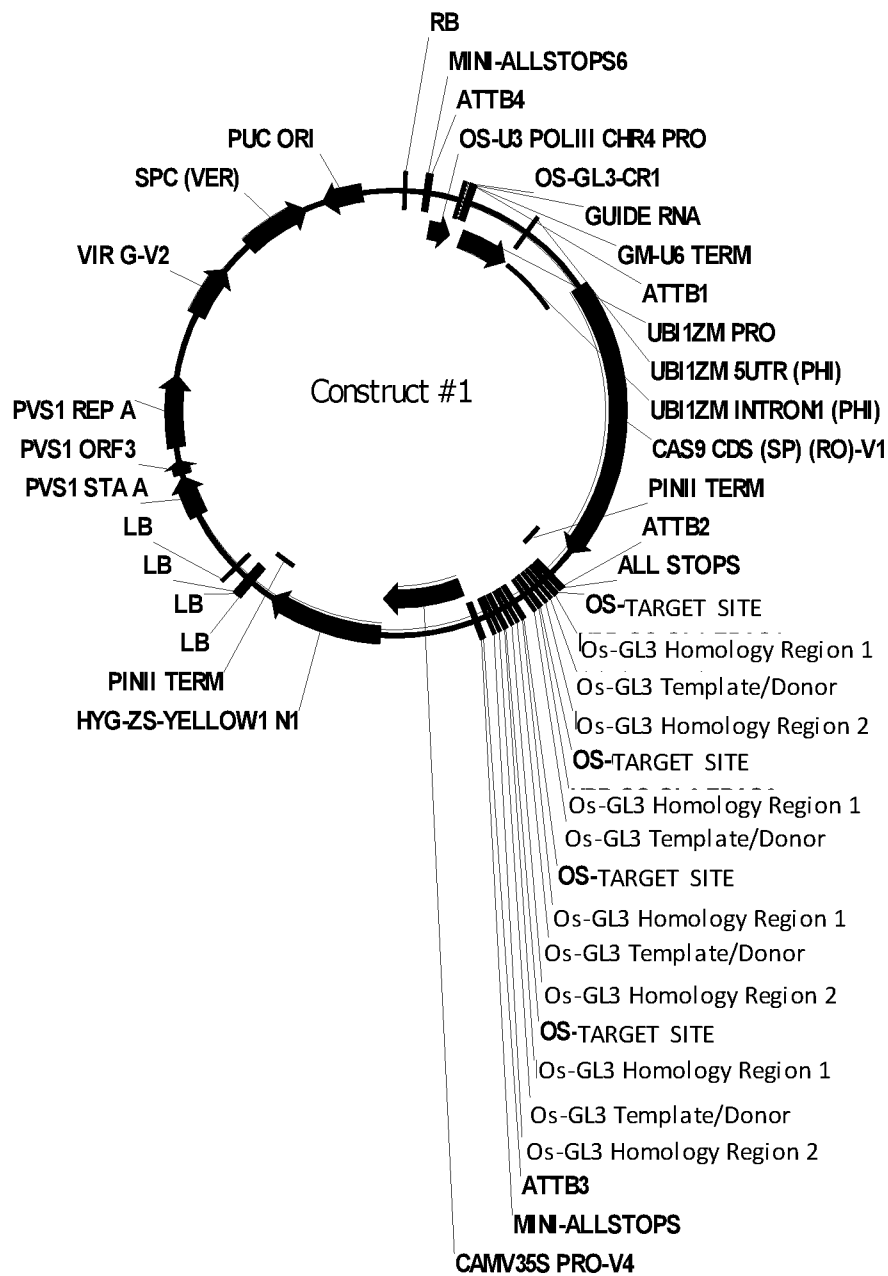
FIG. 4

FIG. 5

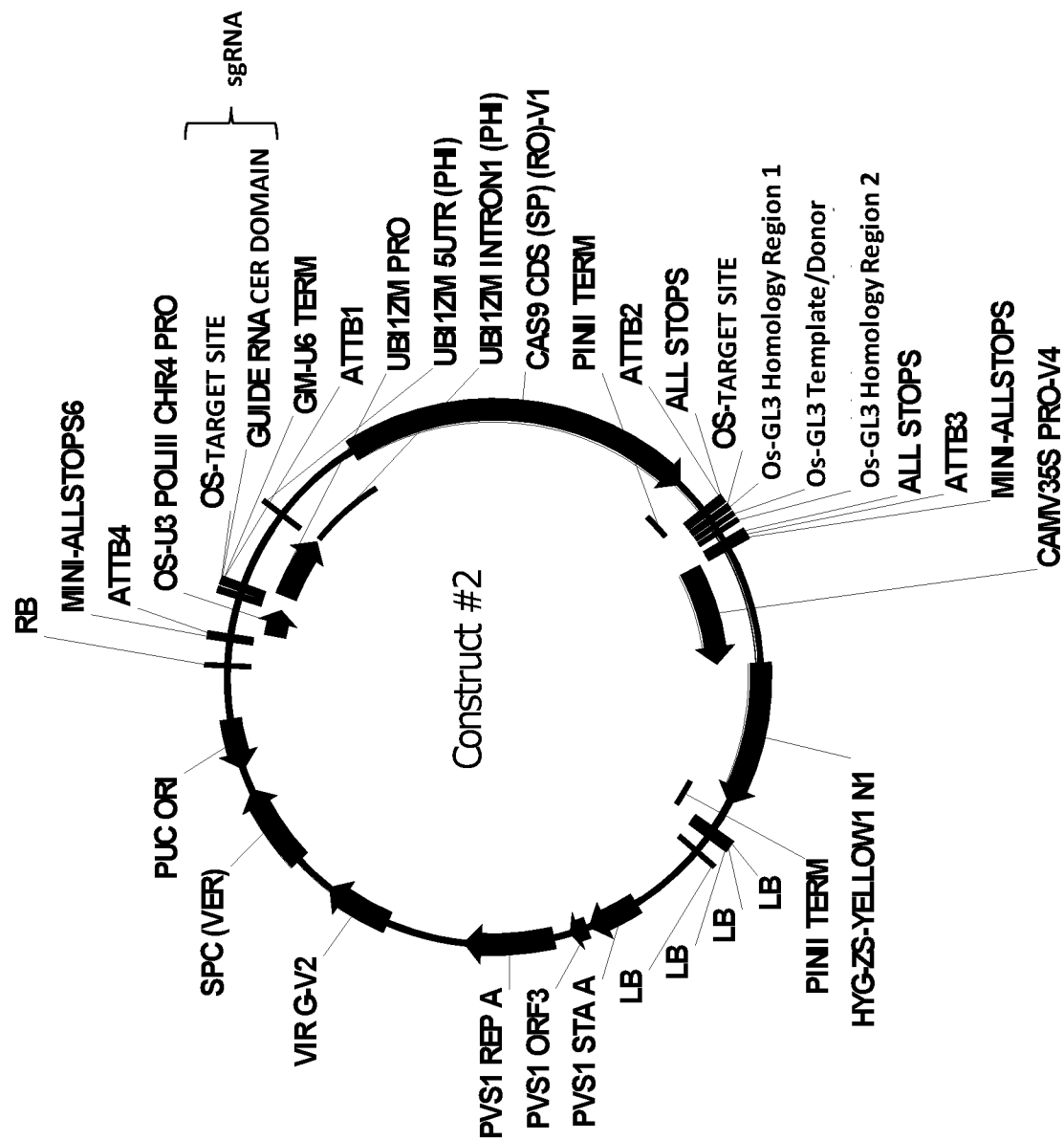


FIG. 6

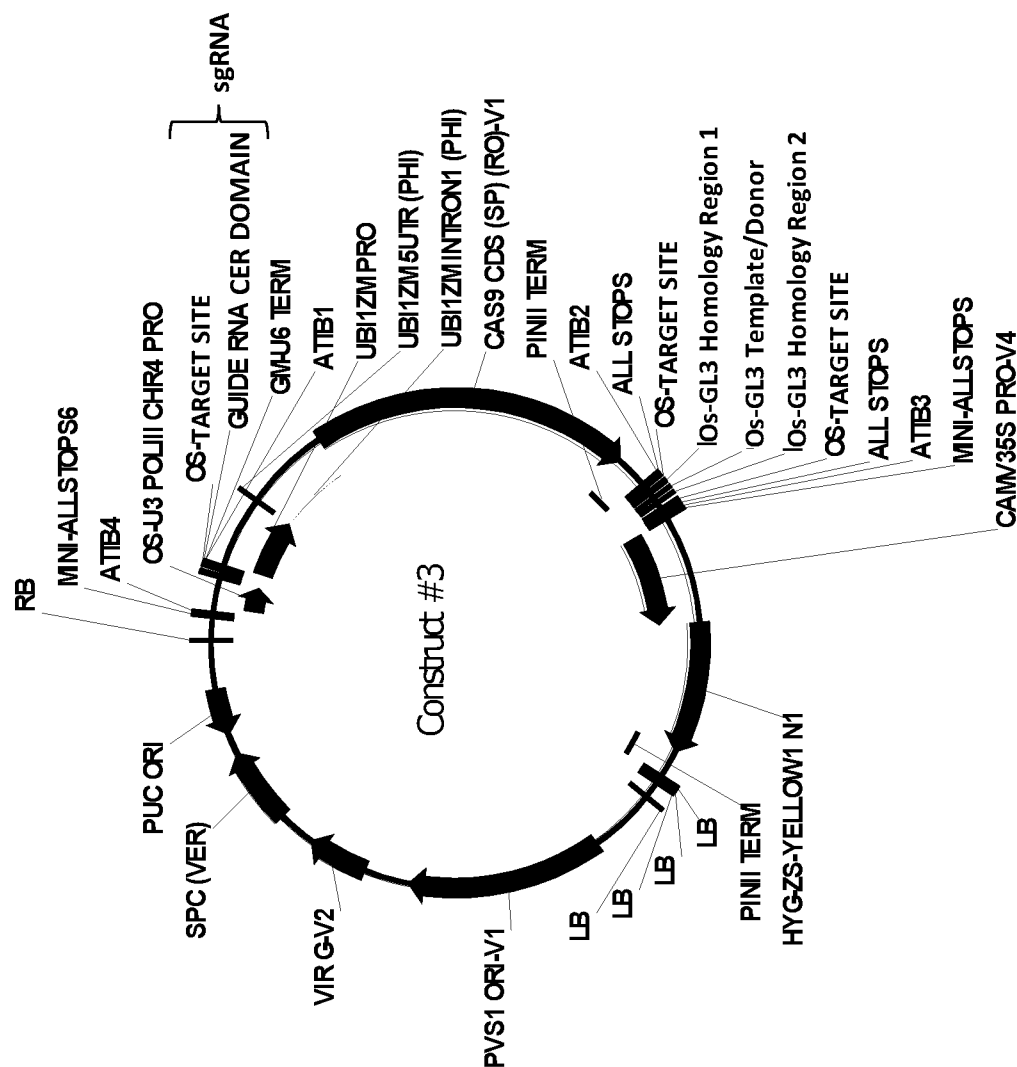


FIG. 7A

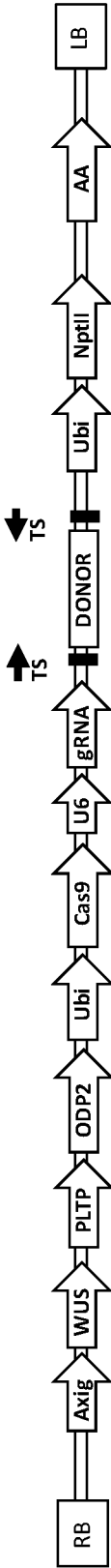


FIG. 7B

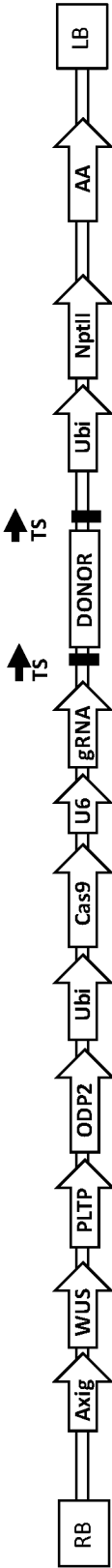
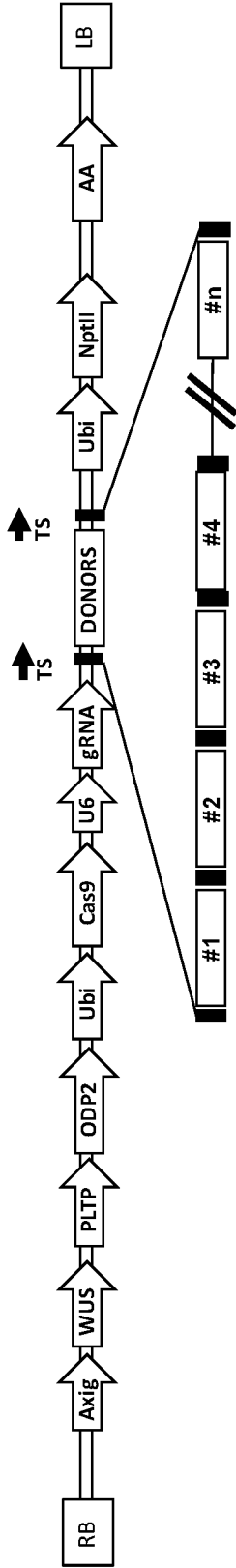
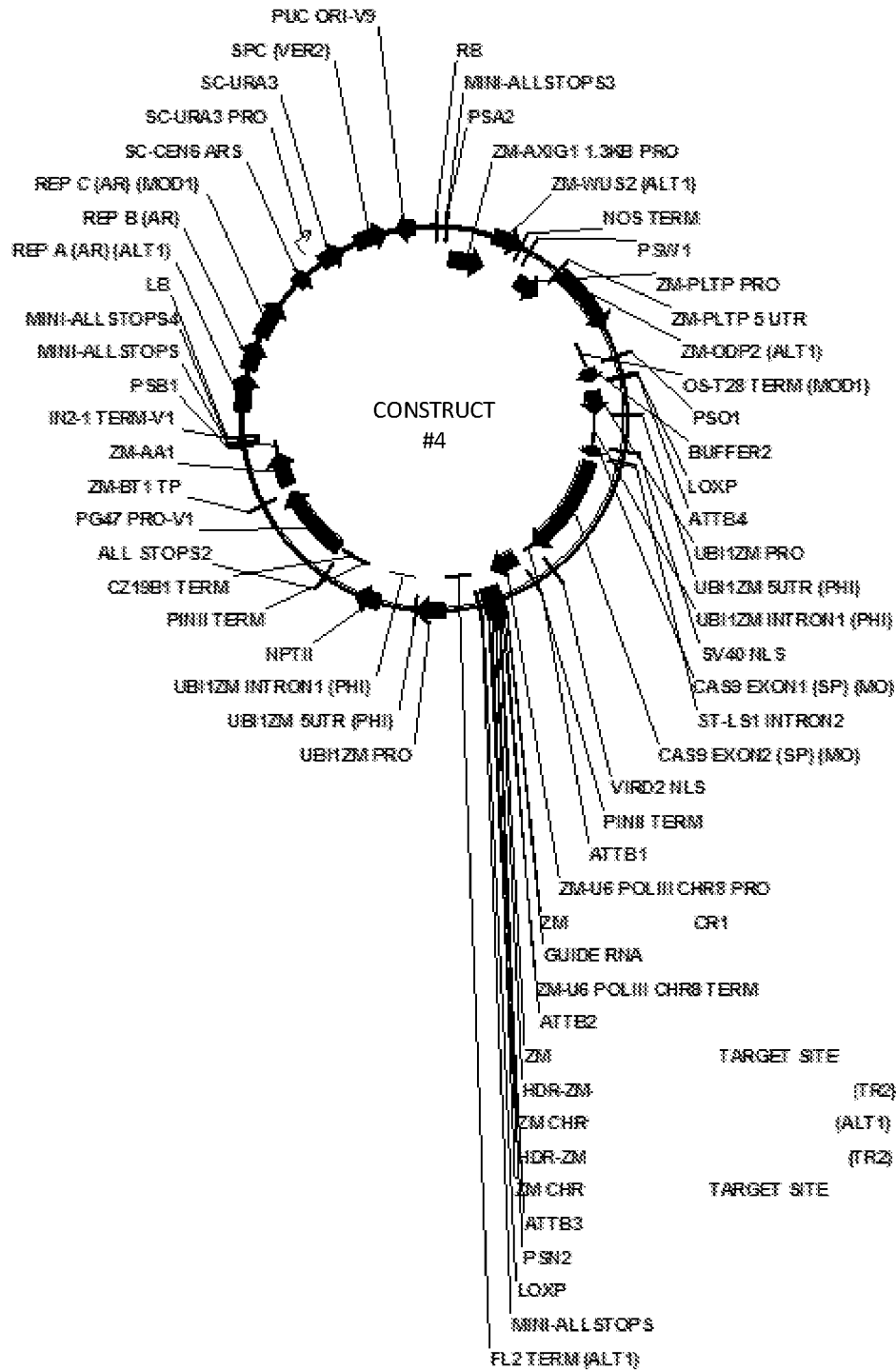


FIG. 7C

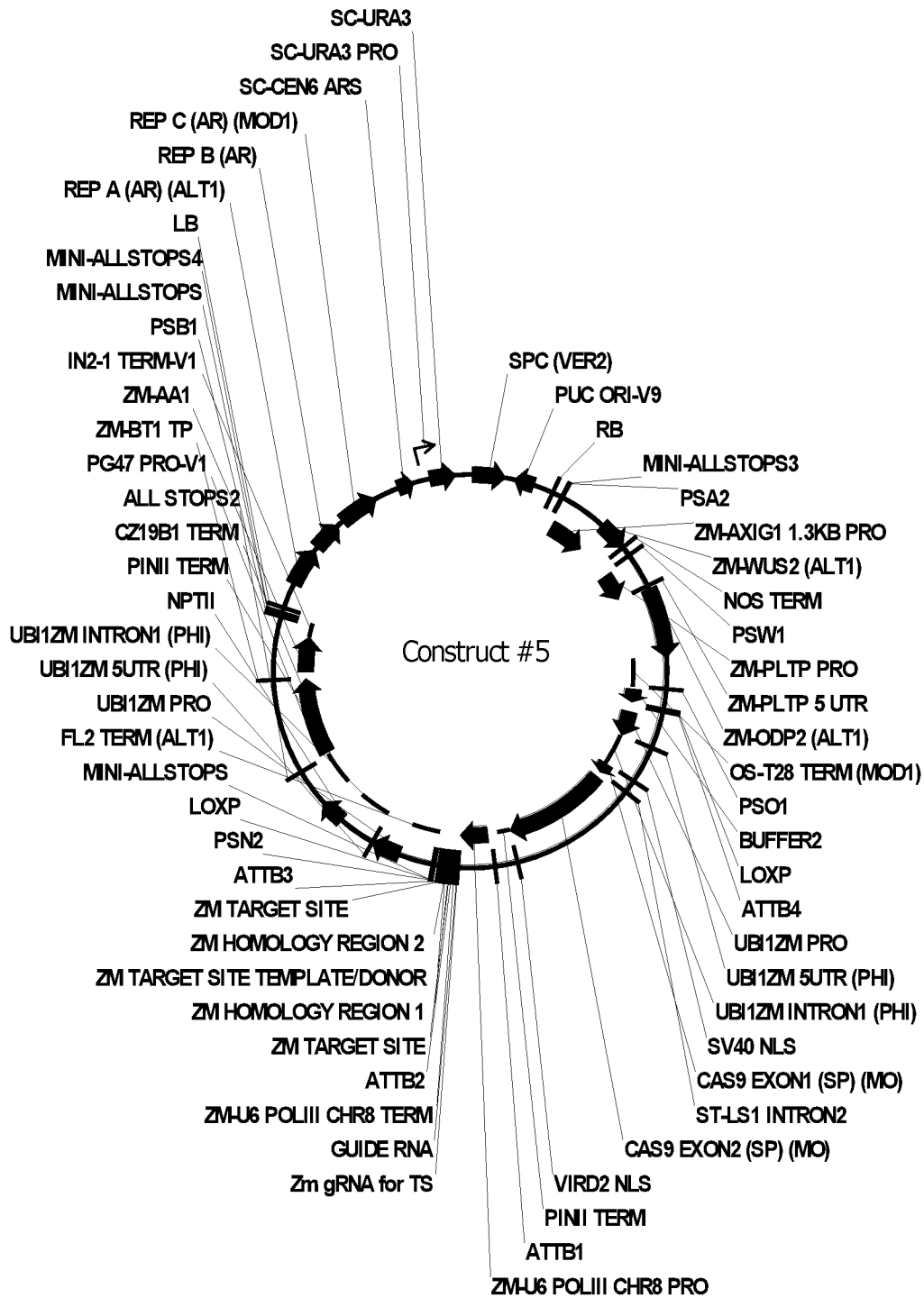


8/11

FIG. 8



9/11

FIG. 9

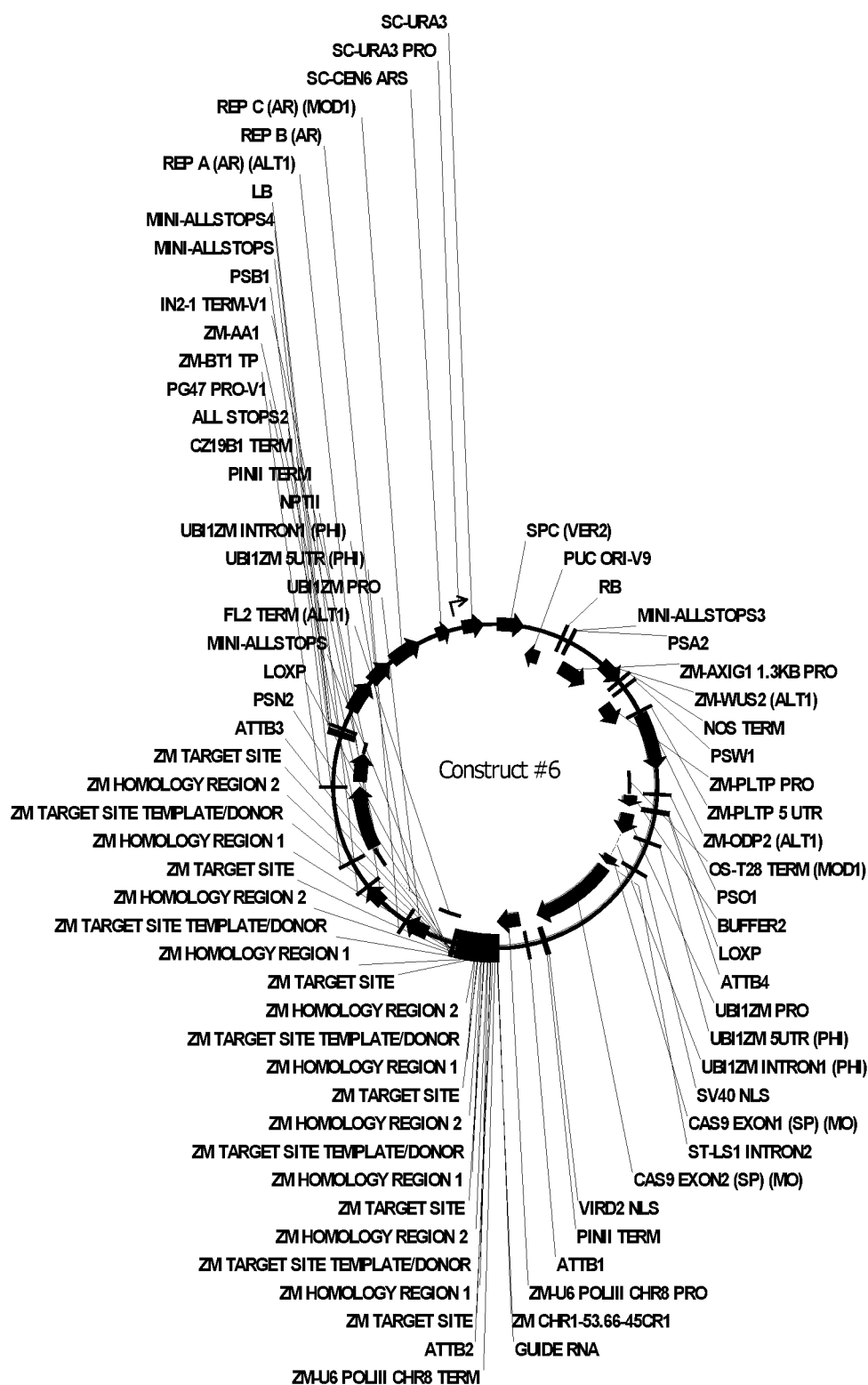


FIG. 11A

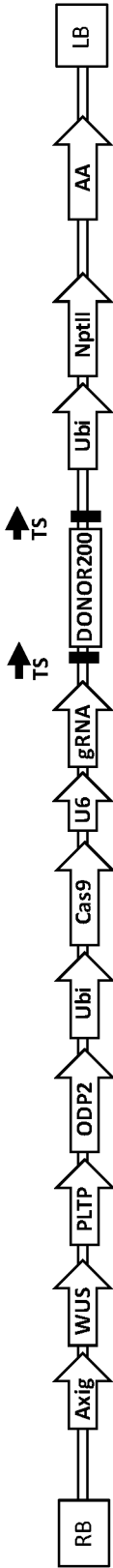


FIG. 11B

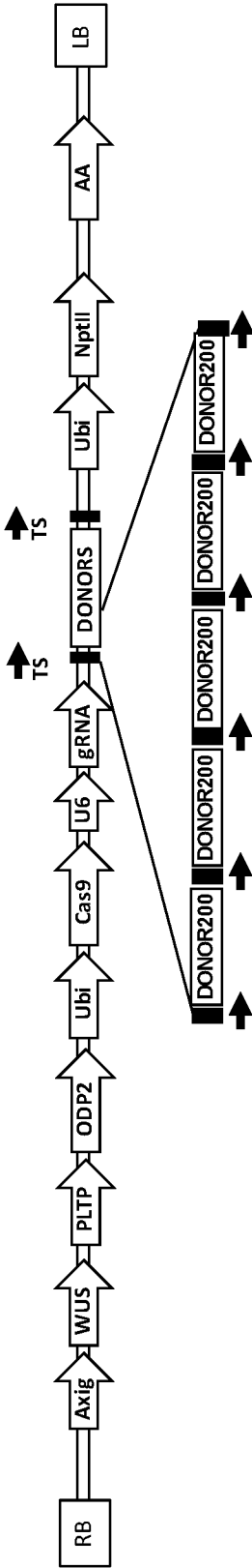


FIG. 11C

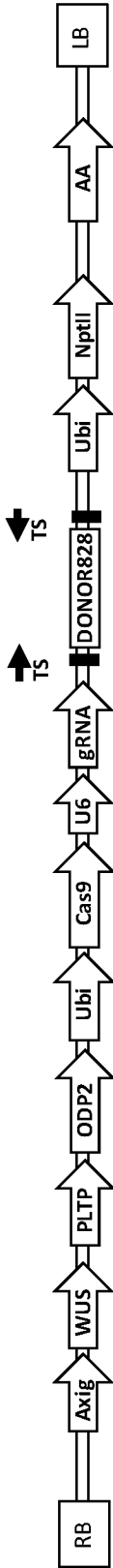
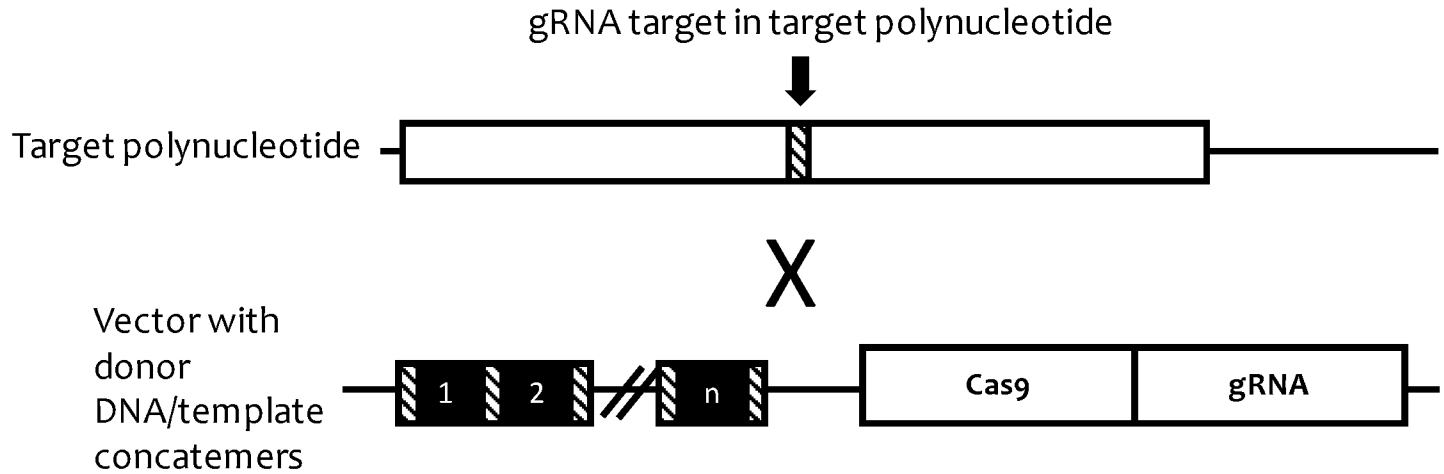


FIG. 1



“Target Sequence” (TS) of the target polynucleotide where the gRNA binds for Cas endonuclease cleavage



“Sequence Unit” comprising a heterologous polynucleotide Donor DNA molecule or Template for DSB repair