US 20150059010A1

(19) **United States**

(12) **Patent Application Publication**

Cigan et al.

(10) **Pub. No.: US 2015/0059010 A1**

(43) **Pub. Date: Feb. 26, 2015**

(54) **GENOME MODIFICATION USING GUIDE POLYNUCLEOTIDE/CAS ENDONUCLEASE SYSTEMS AND METHODS OF USE**

(71) Applicant: **Pioneer Hi-Bred International Inc**, Johnston, IA (US)

(72) Inventors: **Andrew Mark Cigan**, Johnston, IA (US); **Phillip A. Patten**, Menlo Park, CA (US); **Joshua K. Young**, Johnston, IA (US)

(21) Appl. No.: **14/463,691**

(22) Filed: **Aug. 20, 2014**

Related U.S. Application Data

(60) Provisional application No. 62/023,239, filed on Jul. 11, 2014, provisional application No. 61/953,090, filed on Mar. 14, 2014, provisional application No. 61/937,045, filed on Feb. 7, 2014, provisional application No. 61/882,532, filed on Sep. 25, 2013, provisional application No. 61/868,706, filed on Aug. 22, 2013.

Publication Classification

(51) **Int. Cl.**

C12N 15/82	(2006.01)
C12N 15/81	(2006.01)

(52) **U.S. Cl.**

CPC ***C12N 15/8241*** (2013.01); ***C12N 15/81*** (2013.01)

USPC **800/260**; 435/6.18; 435/462; 435/468; 435/471; 435/188; 536/23.1; 800/298; 800/312; 800/314; 800/317.2; 800/317.3; 800/320; 800/320.1; 800/320.2; 800/320.3; 800/322

(57) **ABSTRACT**

Compositions and methods are provided for genome modification of a target sequence in the genome of a cell. The methods and compositions employ a guide polynucleotide/Cas endonuclease system to provide an effective system for modifying or altering target sites within the genome of a cell or organism. Once a genomic target site is identified, a variety of methods can be employed to further modify the target sites such that they contain a variety of polynucleotides of interest. Compositions and methods are also provided for editing a nucleotide sequence in the genome of a cell. Breeding methods and methods for selecting plants utilizing a two component RNA polynucleotide and Cas endonuclease system are also disclosed.

FIGURE 1A

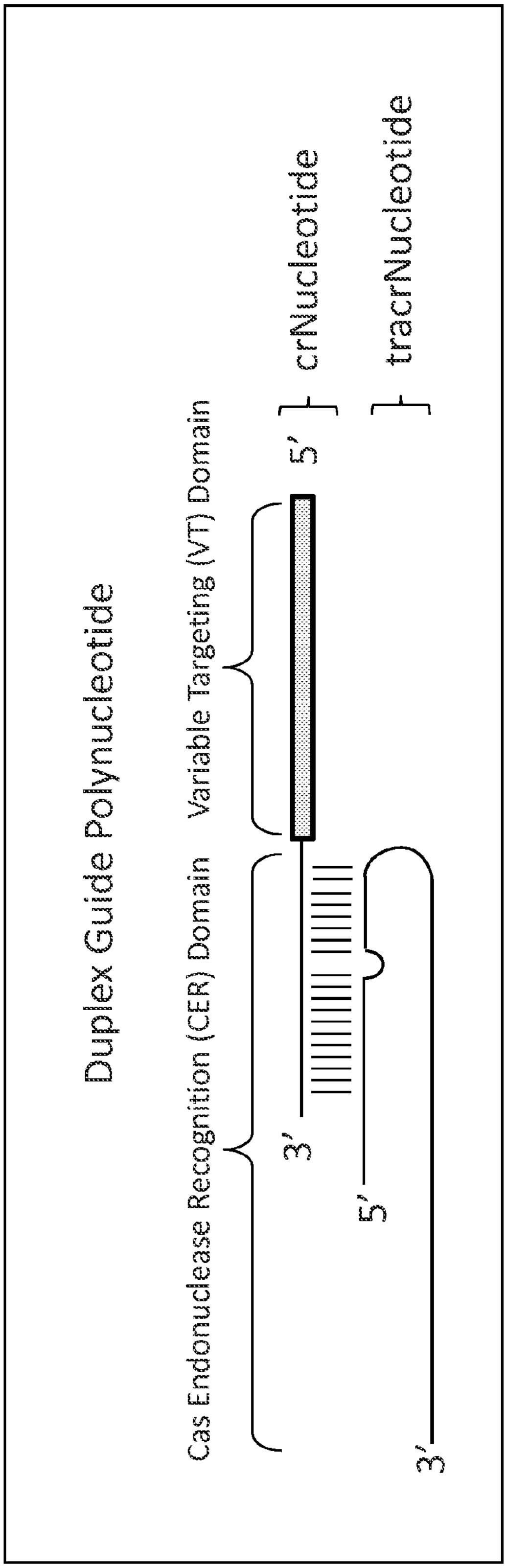

Duplex Guide Polynucleotide
Cas Endonuclease Recognition (CER) Domain
Variable Targeting (VT) Domain
5'
3'
5'
3'
crNucleotide
tracrNucleotide

FIGURE 1B

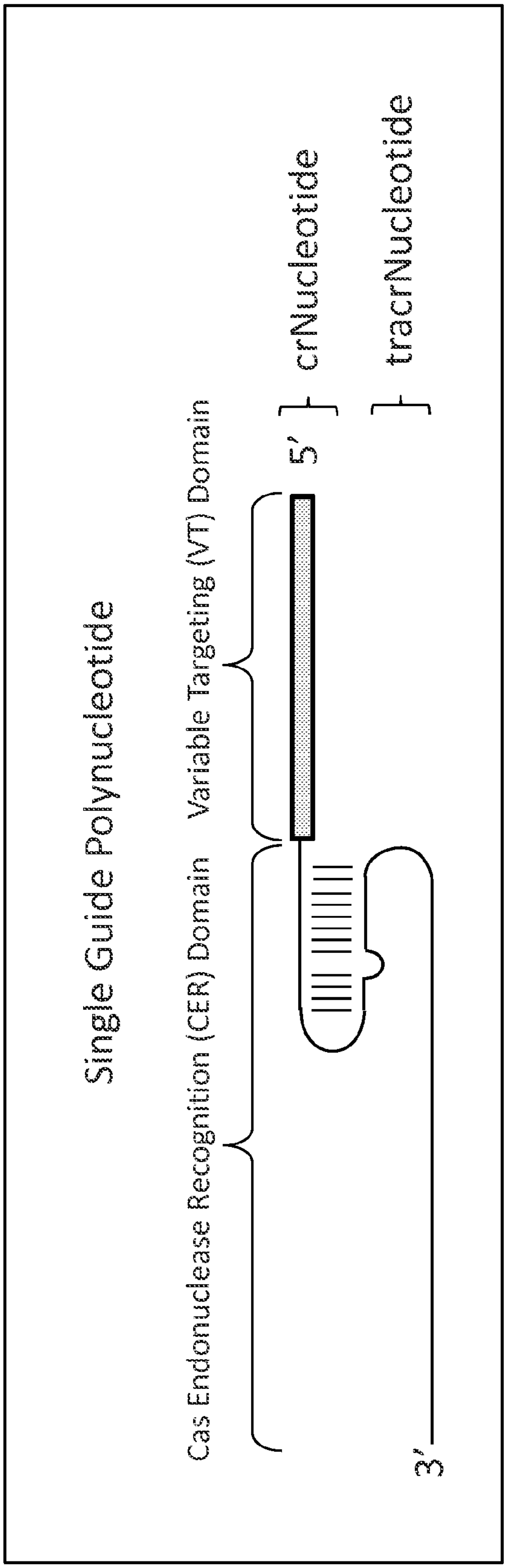

Single Guide Polynucleotide
Cas Endonuclease Recognition (CER) Domain
Variable Targeting (VT) Domain
5′
3′
crNucleotide
tracrNucleotide

FIGURE 2A

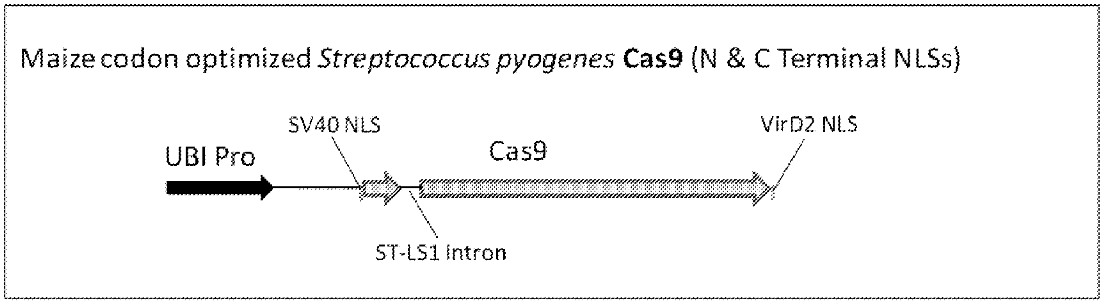

Maize codon optimized *Streptococcus pyogenes* **Cas9** (N & C Terminal NLSs)

FIGURE 2B

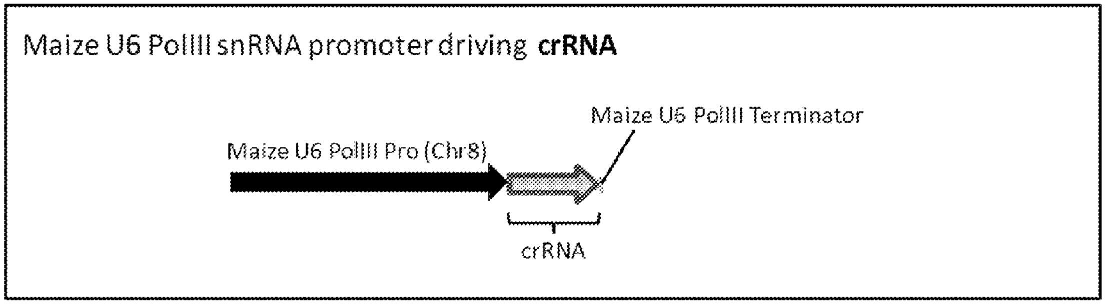

Maize U6 PolIII snRNA promoter driving **crRNA**

FIGURE 2C

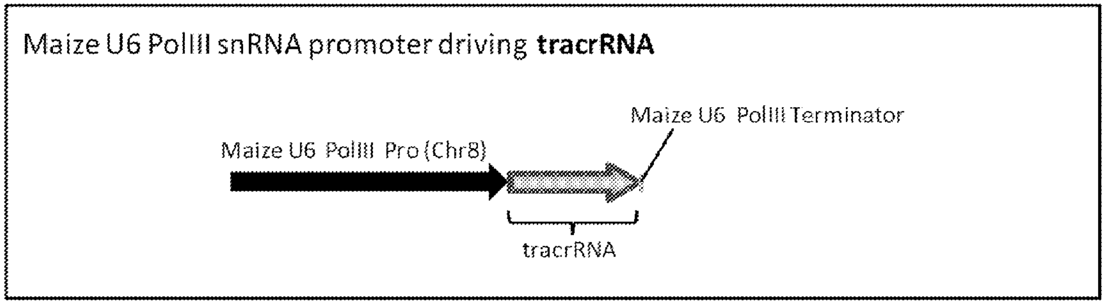

Maize U6 PolIII snRNA promoter driving **tracrRNA**

FIGURE 3A

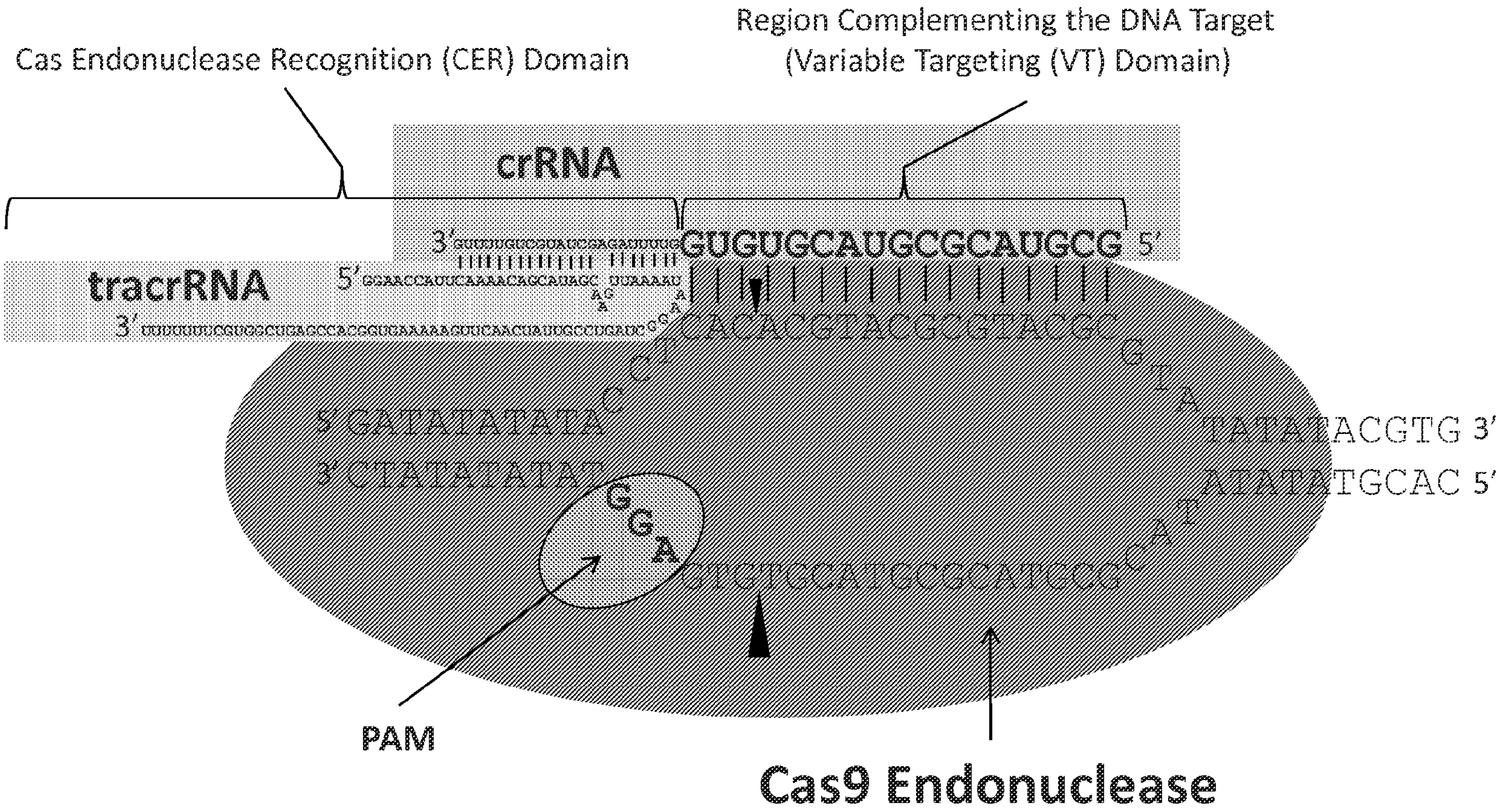

Cas Endonuclease Recognition (CER) Domain
Region Complementing the DNA Target
(Variable Targeting (VT) Domain)
crRNA
3′ GUUUUGUCGUAUCGAGAUUUUG GUGUGCAUGCGCAUGCG 5′
tracrRNA
5′ GGAACCAUUCAAAACAGCAUAGC AAG UUAAAAU AA
3′ UUUUUUUCGUGGCUGAGCCACGGUGAAAAAGUUCAACUAUUGCCUGAUC GG
5′ GATATATATA C C T CACACGTACGCGTACGC G T A TATATACGTG 3′
3′ CTATATATAT G G A GTGTGCATGCGCATGCG C A T ATATATGCAC 5′
PAM
Cas9 Endonuclease

FIGURE 3B

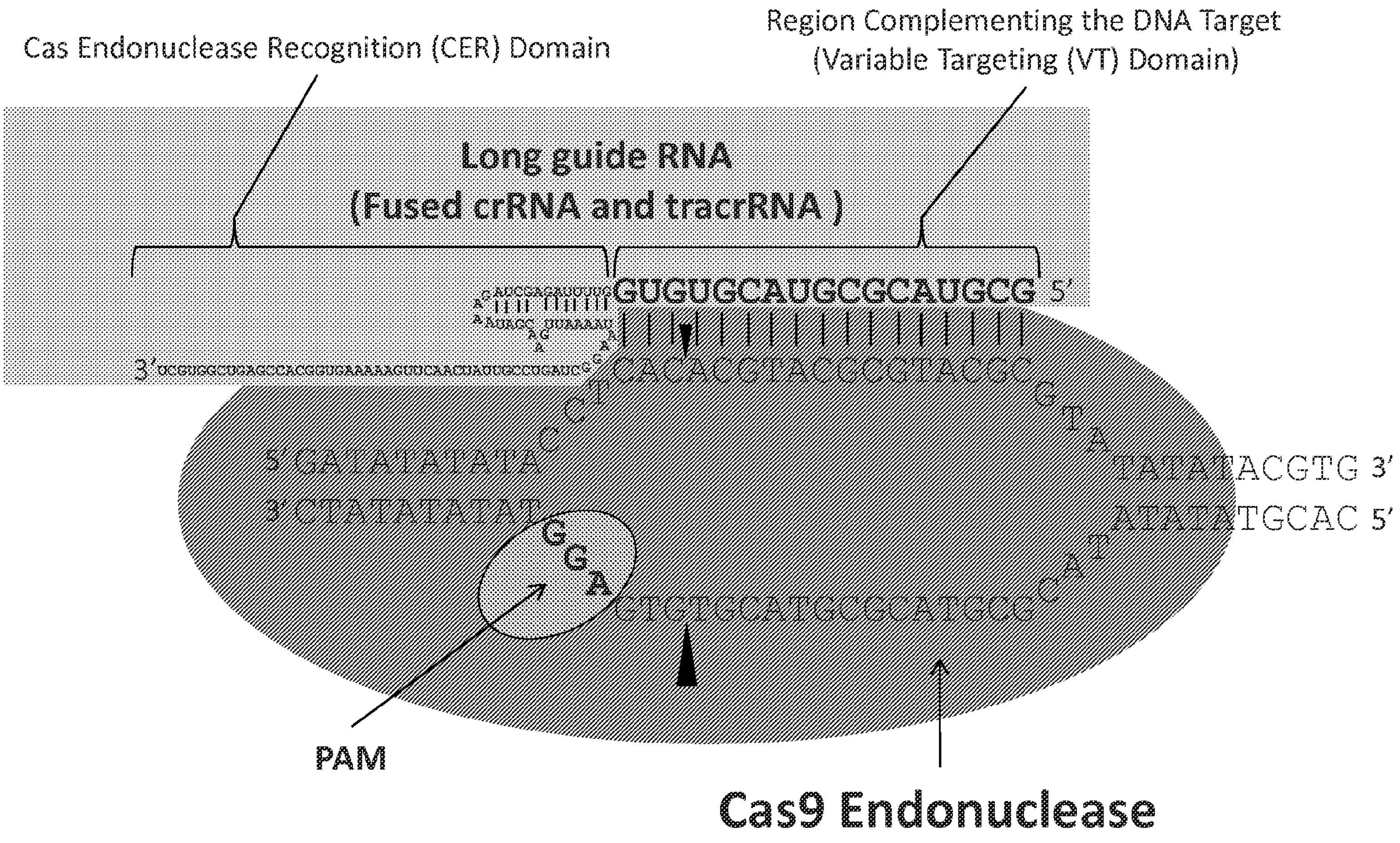
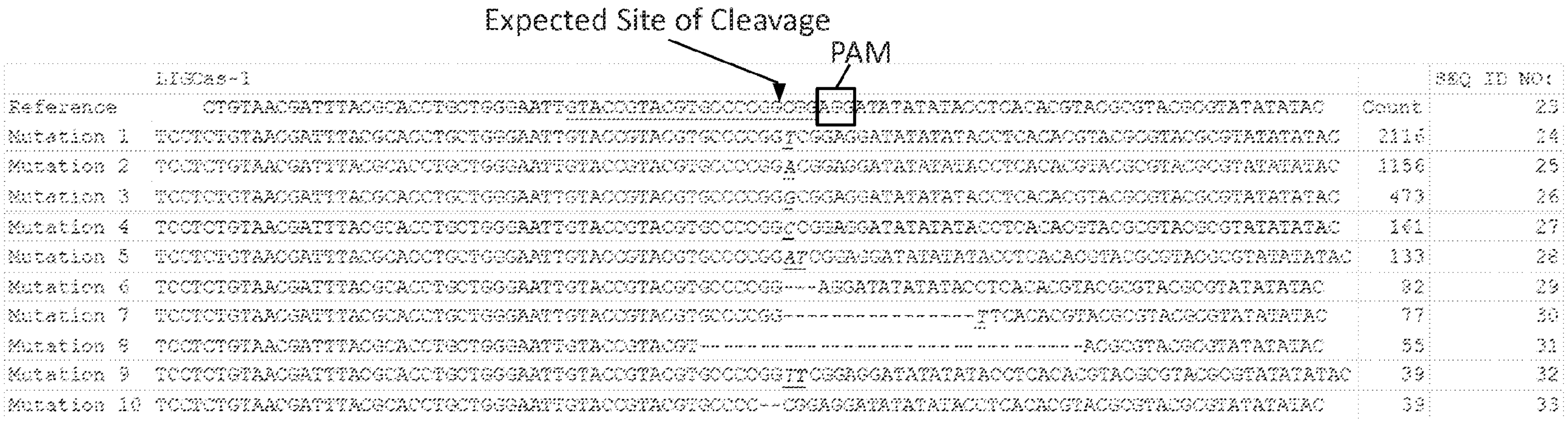

Cas Endonuclease Recognition (CER) Domain
Region Complementing the DNA Target
(Variable Targeting (VT) Domain)
Long guide RNA
(Fused crRNA and tracrRNA)
GAUCGAGAUUUUG
AAUAGC UUAAAAU
A G A
A A G
GUGUGCAUGCGCAUGCG 5'
3'UCGUGGCUGAGCCACGGUGAAAAAGUUCAACUAUUGCCUGAUCG
CACACGTACGCGTACGC
5'GATATATATA
CCT
GTA
TATATACGTG 3'
3'CTATATATAT
ATATATGCAC 5'
GGA
GTGTGCATGCGCATGCG
CAT
PAM
Cas9 Endonuclease

FIGURE 4A

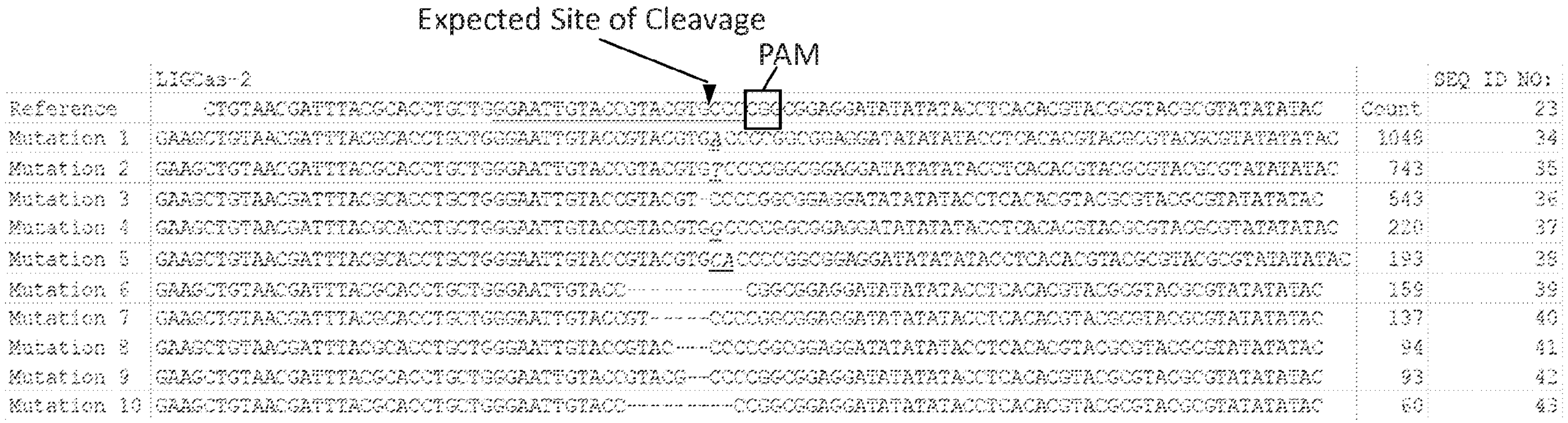


Expected Site of Cleavage

PAM

	LIGCas-1		SEQ ID NO:
Reference	CTGTAACGATTTACGCACCTGCTGGAATTGTACCGTACGTGCCCCGGCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	Count	23
Mutation 1	TCCTCTGTAACGATTTACGCACCTGCTGGAATTGTACCGTACGTGCCCCGGTCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	2116	24
Mutation 2	TCCTCTGTAACGATTTACGCACCTGCTGGAATTGTACCGTACGTGCCCCGGACGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	1156	25
Mutation 3	TCCTCTGTAACGATTTACGCACCTGCTGGAATTGTACCGTACGTGCCCCGGGCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	473	26
Mutation 4	TCCTCTGTAACGATTTACGCACCTGCTGGAATTGTACCGTACGTGCCCCGGCCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	161	27
Mutation 5	TCCTCTGTAACGATTTACGCACCTGCTGGAATTGTACCGTACGTGCCCCGGATCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	133	28
Mutation 6	TCCTCTGTAACGATTTACGCACCTGCTGGAATTGTACCGTACGTGCCCCGG---AGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	82	29
Mutation 7	TCCTCTGTAACGATTTACGCACCTGCTGGAATTGTACCGTACGTGCCCCGG-----------------TTCACACGTACGCGTACGCGTATATATAC	77	30
Mutation 8	TCCTCTGTAACGATTTACGCACCTGCTGGAATTGTACCGTACGT--------------------------------ACGCGTACGCGTATATATAC	55	31
Mutation 9	TCCTCTGTAACGATTTACGCACCTGCTGGAATTGTACCGTACGTGCCCCGGTTCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	39	32
Mutation 10	TCCTCTGTAACGATTTACGCACCTGCTGGAATTGTACCGTACGTGCCCC--CGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	39	33

FIGURE 4B



	LIGCas-2		SEQ ID NO:
Reference	CTGTAACGATTTACGCACCTGCTGGGAATTGTACCGTACGTGCCCCGGCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	Count	23
Mutation 1	GAAGCTGTAACGATTTACGCACCTGCTGGGAATTGTACCGTACGTGACCCCGGCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	1048	34
Mutation 2	GAAGCTGTAACGATTTACGCACCTGCTGGGAATTGTACCGTACGTGTCCCCGGCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	743	35
Mutation 3	GAAGCTGTAACGATTTACGCACCTGCTGGGAATTGTACCGTACGT-CCCCGGCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	543	36
Mutation 4	GAAGCTGTAACGATTTACGCACCTGCTGGGAATTGTACCGTACGTGGCCCCGGCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	220	37
Mutation 5	GAAGCTGTAACGATTTACGCACCTGCTGGGAATTGTACCGTACGTGCACCCCGGCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	193	38
Mutation 6	GAAGCTGTAACGATTTACGCACCTGCTGGGAATTGTACC----------CGGCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	159	39
Mutation 7	GAAGCTGTAACGATTTACGCACCTGCTGGGAATTGTACCGT-----CCCCGGCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	137	40
Mutation 8	GAAGCTGTAACGATTTACGCACCTGCTGGGAATTGTACCGTAC---CCCCGGCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	94	41
Mutation 9	GAAGCTGTAACGATTTACGCACCTGCTGGGAATTGTACCGTACG--CCCCGGCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	93	42
Mutation 10	GAAGCTGTAACGATTTACGCACCTGCTGGGAATTGTACC---------CCGGCGGAGGATATATATACCTCACACGTACGCGTACGCGTATATATAC	60	43

FIGURE 4C

Expected Site of Cleavage

PAM

	LIGCas-3		SEQ ID NO:
Reference	CGCAAATGAGTAGCAGCGCACGTATATATACGCGTACGCGTACGTGTGAGGTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	Count	44
Mutation 1	AAGGCGCAAATGAGTAGCAGCGCACGTATATATACGCGTACGCGTACGTTGTGAGGTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	1208	45
Mutation 2	AAGGCGCAAATGAGTAGCAGCGCACGTATATATACGCGTACGCGTACG-GTGAGGTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	453	46
Mutation 3	AAGGCGCAAATGAGTAGCAGCGCACGTATATATACGCGTACGCGTACGT--GAGGTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	338	47
Mutation 4	AAGGCGCAAATGAGTAGCAGCGCACGTATATATACGCGTACGCGTAC-TGTGAGGTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	145	48
Mutation 5	AAGGCGCAAATGAGTAGCAGCGCACGTATATATA-----------------------------TCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	94	49
Mutation 6	AAGGCGCAAATGAGTAGCAGCGCACGTATATATACGCGTACGCGT----GTGAGGTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	53	50
Mutation 7	AAGGCGCAAATGAGTAGCAGCGCACGTATATATACGCGTACG--------TGAGGTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	42	51
Mutation 8	AAGGCGCAAATGAGTAGCAGCGCACGTATATATACGCGTACGCGTAC--------TATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	37	52
Mutation 9	AAGGCGCAAATGAGTAGCAGCGCACGTATATATACGCGTACGCGTACG---GAGGTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	20	53
Mutation 10	AAGGCGCAAATGAGTAGCAGCGCACGTATATATACGCGTACGCGTACG-ATGAGGTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	19	54

FIGURE 5

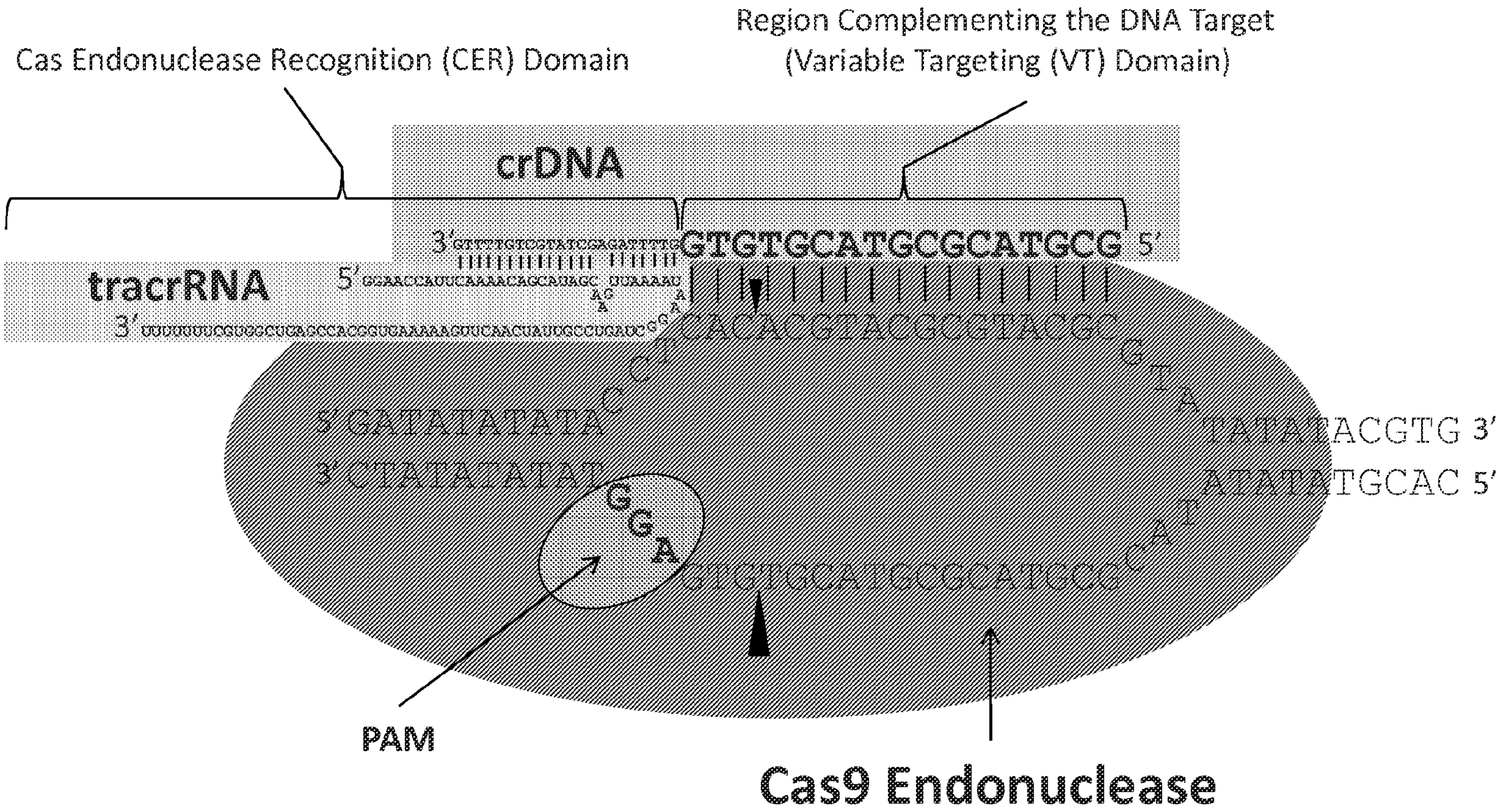
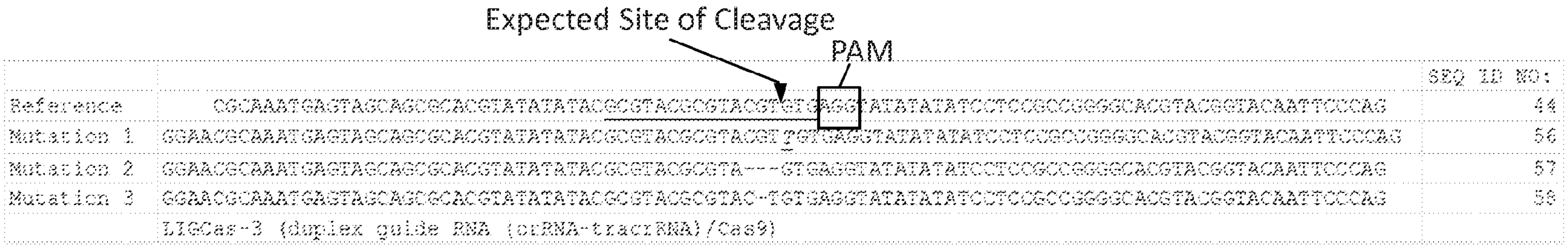
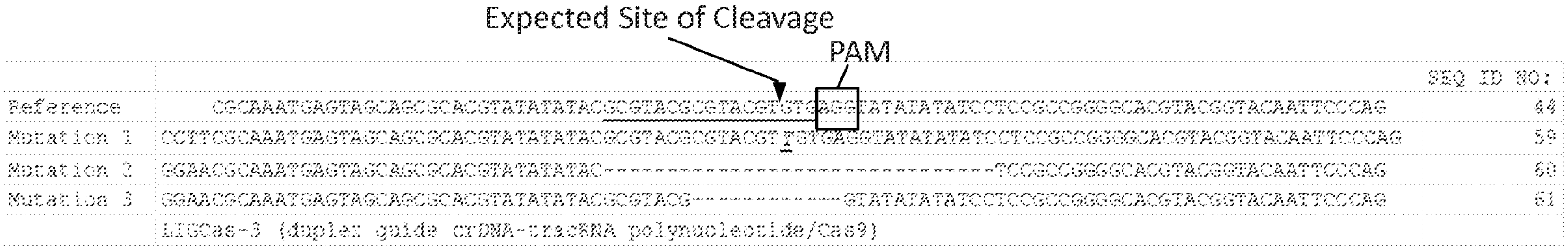

Cas Endonuclease Recognition (CER) Domain
Region Complementing the DNA Target (Variable Targeting (VT) Domain)
crDNA
3′GTTTTGTCGTATCGAGATTTTGGTGTGCATGCGCATGCG 5′
tracrRNA
5′GGAACCAUUCAAAACAGCAUAGC UUAAAAU
3′UUUUUUUCGUGGCUGAGCCACGGUGAAAAAGUUCAACUAUUGCCUGAUC
CACACGTACGCGTACGC
5′GATATATATA
TATATACGTG 3′
3′CTATATATAT
ATATATGCAC 5′
GGA
GTGTGCATGCGCATGCG
PAM
Cas9 Endonuclease

FIGURE 6A



		SEQ ID NO:
Reference	CGCAAATGAGTAGCAGCGCACGTATATATACGCGTACGCGTACGTGTGAGGTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	44
Mutation 1	GGAACGCAAATGAGTAGCAGCGCACGTATATATACGCGTACGCGTACGTTGTGAGGTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	56
Mutation 2	GGAACGCAAATGAGTAGCAGCGCACGTATATATACGCGTACGCGTA---GTGAGGTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	57
Mutation 3	GGAACGCAAATGAGTAGCAGCGCACGTATATATACGCGTACGCGTAC-TGTGAGGTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	58
	LIGCas-3 (duplex guide RNA (crRNA-tracrRNA)/Cas9)	

FIGURE 6B



		SEQ ID NO:
Reference	CGCAAATGAGTAGCAGCGCACGTATATATACGCGTACGCGTACGTGTGAGGTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	44
Mutation 1	CCTTCGCAAATGAGTAGCAGCGCACGTATATATACGCGTACGCGTACGTTGTGAGGTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	59
Mutation 2	GGAACGCAAATGAGTAGCAGCGCACGTATATATAC-------------------------------TCCGCCGGGGCACGTACGGTACAATTCCCAG	60
Mutation 3	GGAACGCAAATGAGTAGCAGCGCACGTATATATACGCGTACG------------GTATATATATCCTCCGCCGGGGCACGTACGGTACAATTCCCAG	61
	LIGCas-3 (duplex guide crDNA-tracRNA polynucleotide/Cas9)	

FIGURE 7

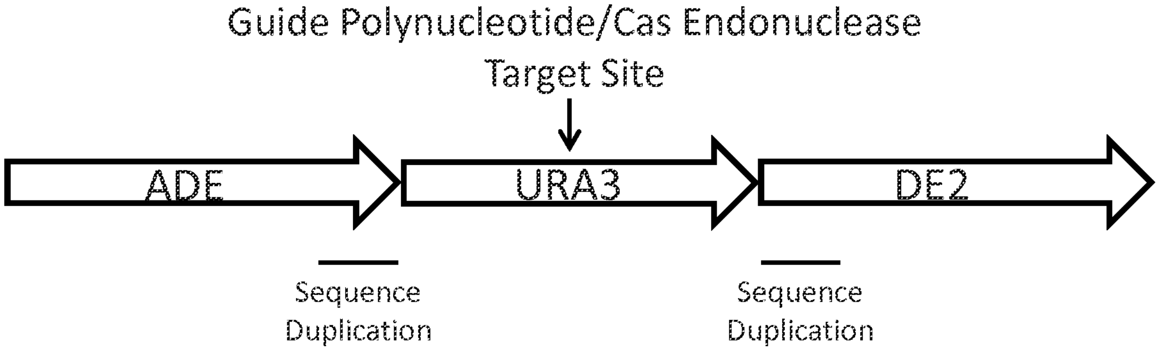

Guide Polynucleotide/Cas Endonuclease
Target Site
ADE
URA3
DE2
Sequence
Duplication
Sequence
Duplication

FIGURE 8

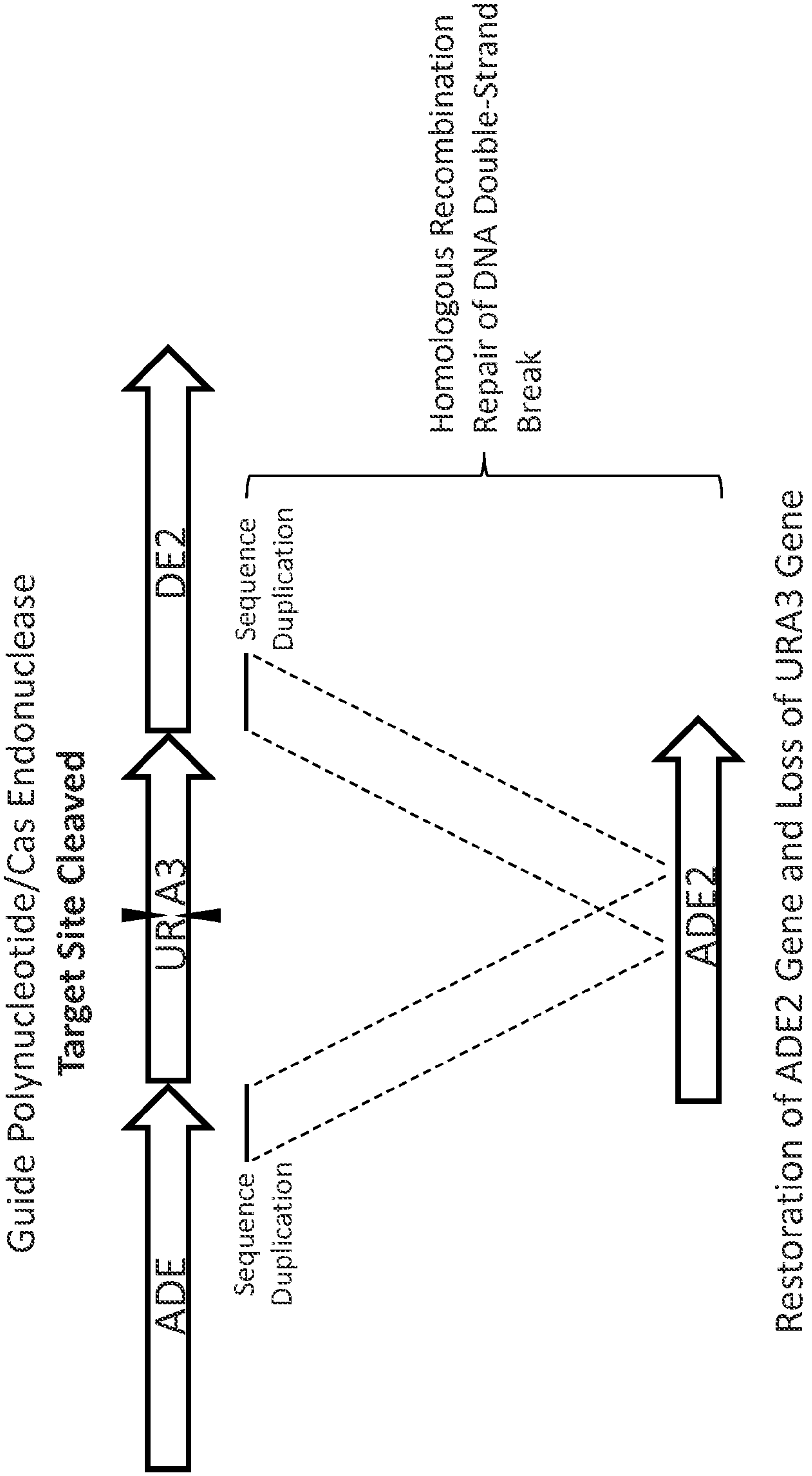
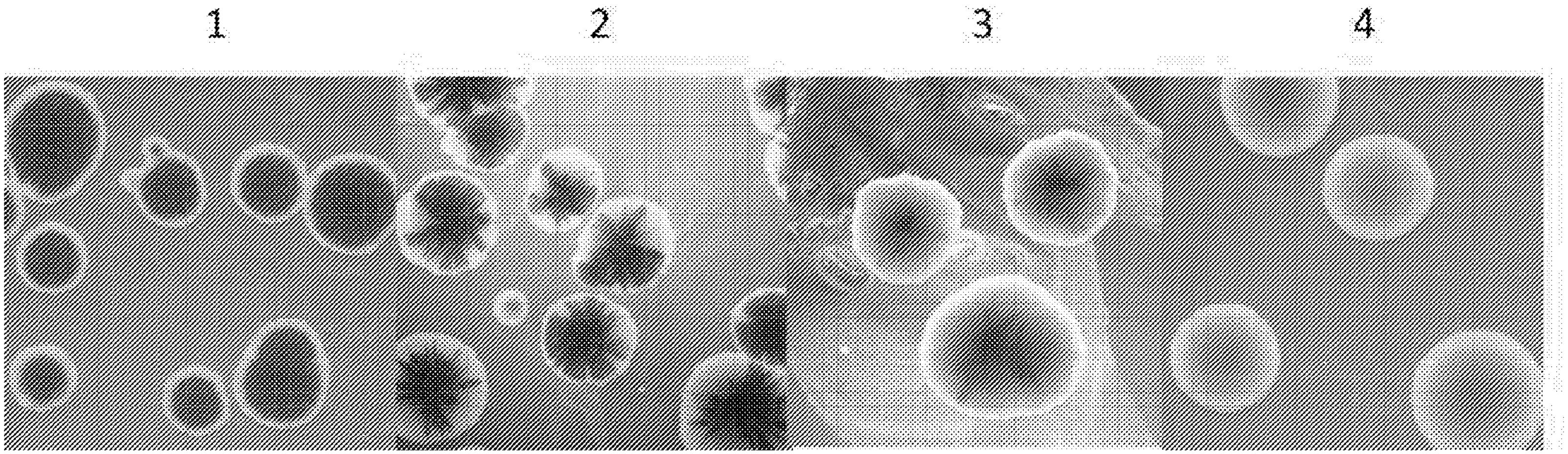

Guide Polynucleotide/Cas Endonuclease
Target Site Cleaved
ADE
URA3
DE2
Sequence Duplication
Sequence Duplication
Homologous Recombination Repair of DNA Double-Strand Break
ADE2
Restoration of ADE2 Gene and Loss of URA3 Gene

FIGURE 9

Yeast Red/White Colony Numerical Scoring System



GENOME MODIFICATION USING GUIDE POLYNUCLEOTIDE/CAS ENDONUCLEASE SYSTEMS AND METHODS OF USE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/868,706, filed Aug. 22, 2013, U.S. Provisional Application No. 61/882,532, filed Sep. 25, 2013, U.S. Provisional Application No. 61/937,045, filed Feb. 7, 2014, U.S. Provisional Application No. 61/953,090, filed Mar. 14, 2014, and U.S. Provisional Application No. 62/023239, filed Jul. 11, 2014; all of which are hereby incorporated herein in their entirety by reference.

FIELD

[0002] The disclosure relates to the field of molecular biology, in particular, to methods for altering the genome of a cell.

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

[0003] The official copy of the sequence listing is submitted electronically via EFS-Web as an ASCII formatted sequence listing with a file named 20140815_BB2344USNP_ST25_SequenceListing created on Aug. 15, 2014 and having a size of 82 kilobytes and is filed concurrently with the specification. The sequence listing contained in this ASCII formatted document is part of the specification and is herein incorporated by reference in its entirety.

BACKGROUND

[0004] Recombinant DNA technology has made it possible to insert foreign DNA sequences into the genome of an organism, thus, altering the organism's phenotype.

[0005] One method for inserting or modifying a DNA sequence involves homologous DNA recombination by introducing a transgenic DNA sequence flanked by sequences homologous to the genomic target. U.S. Pat. No. 5,527,695 describes transforming eukaryotic cells with DNA sequences that are targeted to a predetermined sequence of the eukaryote's DNA. Specifically, the use of site-specific recombination is discussed. Transformed cells are identified through use of a selectable marker included as a part of the introduced DNA sequences.

[0006] It was shown that artificially induced site-specific genomic double-stranded breaks in plant cells were repaired by homologous recombination with exogenously supplied DNA using two different pathways. (Puchta et al., (1996) *Proc. Natl. Acad. Sci. USA* 93:5055-5060; U.S. Patent Application Publication No. 2005/0172365A1 published Aug. 4, 2005; U.S. Patent Application Publication No. 2006/0282914 published Dec. 14, 2006; WO 2005/028942 published Jun. 2, 2005).

[0007] Since the isolation, cloning, transfer and recombination of DNA segments, including coding sequences and non-coding sequences, is most conveniently carried out using restriction endonuclease enzymes. Much research has focused on studying and designing endonucleases such as WO 2004/067736 published Aug. 12, 2004; U.S. Pat. No. 5,792,632 issued to Dujon et al., Aug. 11, 1998; U.S. Pat. No. 6,610,545 B2 issued to Dujon et al., Aug. 26, 2003; Chevalier et al., (2002) *Mol Cell* 10:895-905; Chevalier et al., (2001) *Nucleic Acids Res* 29:3757-3774; Seligman et al., (2002) *Nucleic Acids Res* 30:3870-3879.

[0008] Although several approaches have been developed to target a specific site for modification in the genome of a cell, there still remains a need for more efficient and effective methods for producing an organism, such as but not limited to yeast and fertile plants, having an altered genome comprising specific modifications in a defined region of the genome of the cell.

BRIEF SUMMARY

[0009] Compositions and methods are provided employing a guide polynucleotide/Cas endonuclease system for genome modification of a target sequence in the genome of a cell or organism, for gene editing, and for inserting a polynucleotide of interest into the genome of a cell or organism. The methods and compositions employ a guide polynucleotide/Cas endonuclease system to provide for an effective system for modifying or altering target sites and editing nucleotide sequences of interest within the genome of cell, wherein the guide polynucleotide is comprised of a DNA, RNA or a DNA-RNA combination sequence. Cells include, but are not limited to non-human, animal, bacterial, fungal, insect, yeast, and plant cells. Once a genomic target site is identified, a variety of methods can be employed to further modify the target sites such that they contain a variety of polynucleotides of interest. Breeding methods and methods for selecting plants utilizing a guide polynucleotide and Cas endonuclease system are also disclosed. Also provided are nucleic acid constructs, cells, yeast, plants, plant cells, explants, seeds and grain having the guide polynucleotide/Cas endonuclease system. Compositions and methods are also provided for editing a nucleotide sequence in the genome of a cell. The nucleotide sequence to be edited (the nucleotide sequence of interest) can be located within or outside a target site that is recognized by a Cas endonuclease.

[0010] Thus in a first embodiment of the disclosure, the composition comprises a guide polynucleotide comprising: (i) a first nucleotide sequence domain that is complementary to a nucleotide sequence in a target DNA; and, (ii) a second nucleotide sequence domain that interacts with a Cas endonuclease, wherein the first nucleotide sequence domain and the second nucleotide sequence domain are composed of deoxyribonucleic acids (DNA), ribonucleic acids (RNA), or a combination thereof, wherein the guide polynucleotide does not solely comprise ribonucleic acids. The % complementation between the first nucleotide sequence domain (Variable Targeting domain) and the target sequence can be 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%. The first nucleotide sequence domain (VT domain) comprises a contiguous stretch of 12 to 30 nucleotides.

[0011] In one embodiment, the first nucleotide sequence domain (VT domain) and the second nucleotide sequence domain of the guide polynucleotide are located on a single molecule. In another embodiment, the second nucleotide sequence domain (Cas Endonuclease Recognition domain) comprises two separate molecules that are capable of hybridizing along a region of complementarity.

[0012] In another embodiment, the composition comprises a guide polynucleotide, wherein the first nucleotide sequence domain is a DNA sequence and the second nucleotide sequence domain is selected from the group consisting of a DNA sequence, a RNA sequence, and a combination thereof.

[0013] In another embodiment, the composition comprises a guide polynucleotide, wherein the first nucleotide sequence domain and/or the second nucleotide sequence domain comprises at least one modification that optionally provides for an additional beneficial feature, wherein said at least one modification is selected from 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. The additional beneficial can be 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, or an increased cellular permeability.

[0014] In another embodiment, the composition comprises a guide polynucleotide/Cas endonuclease complex wherein the guide polynucleotide comprises (i) a first nucleotide sequence domain that is complementary to a nucleotide sequence in a target DNA; and (ii) a second nucleotide sequence domain that interacts with a Cas endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site.

[0015] In another embodiment, the composition comprises a guide polynucleotide/Cas endonuclease complex, wherein the first nucleotide sequence domain and/or the second nucleotide sequence domain of said guide polynucleotide comprises at least one modification that optionally provides for an additional beneficial feature, wherein said at least one modification is selected from 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.

[0016] In another embodiment, the composition comprises plant or seed comprising the guide polynucleotide or the guide polynucleotide/Cas endonuclease complex of the disclosure.

[0017] In another embodiment, the method comprises a method for modifying a target site in the genome of a cell, the method comprising introducing a guide polynucleotide into a cell having a Cas endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site.

[0018] In another embodiment, the method comprises a method for modifying a target site in the genome of a cell, the method comprising introducing a guide polynucleotide and a Cas endonuclease into a cell, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site.

[0019] In another embodiment, the method comprises a method for introducing a polynucleotide of interest into a target site in the genome of a cell, the method comprising: a) providing a guide polynucleotide, a donor DNA and a Cas endonuclease to a cell, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site; b) contacting the cell of (a) with a donor DNA comprising a polynucleotide of interest; and,

c) identifying at least one cell from (b) comprising in its genome the polynucleotide of interest integrated at said target site.

[0020] In another embodiment, the method comprises a method for modifying a target site in the genome of a cell, the method comprising: a) providing to a cell a crNucleotide, a first recombinant DNA construct capable of expressing a tracrRNA, and a second recombinant DNA capable of expressing a Cas endonuclease, wherein said crNucleotide is a deoxyribonucleotide sequence or a combination of a deoxyribonucleotide and ribonucleotide sequence, wherein said crNucleotide, said tracrRNA and said Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site; and, b) identifying at least one cell that has a modification at said target site, wherein the modification is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

[0021] In another embodiment, the method comprises a method for modifying a target site in the genome of a cell, the method comprising: a) providing to a cell a tracrNucleotide, a first recombinant DNA construct capable of expressing a crRNA and a second recombinant DNA capable of expressing a Cas endonuclease, wherein said tracrNucleotide is selected a deoxyribonucleotide sequence or a combination of a deoxyribonucleotide and ribonucleotide sequence, wherein said tracrNucleotide, said crRNA and said Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site; and, b) identifying at least one cell that has a modification at said target site, wherein the modification is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

[0022] In another embodiment, the method comprises a method for introducing a polynucleotide of interest into a target site in the genome of a cell, the method comprising: a) introducing into a cell a first recombinant DNA construct capable of expressing a guide polynucleotide, and a second recombinant DNA construct capable of expressing a Cas endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site; b) contacting the cell of (a) with a donor DNA comprising a polynucleotide of interest; and, c) identifying at least one cell from (b) comprising in its genome the polynucleotide of interest integrated at said target site.

[0023] In another embodiment, the method comprises a method for editing a nucleotide sequence in the genome of a cell, the method comprising introducing a guide polynucleotide, a polynucleotide modification template and at least one Cas endonuclease into a cell, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein the Cas endonuclease introduces a double-strand break at a target site in the genome of said cell, wherein said polynucleotide modification template comprises at least one nucleotide modification of said nucleotide sequence.

[0024] In another embodiment, the composition comprises a plant or seed comprising a guide polynucleotide and a Cas endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said Cas endonuclease and guide polynucleotide are capable of forming a complex and creating a double strand break in a genomic target site of said plant.

[0025] In another embodiment, the composition comprises a plant or seed comprising a recombinant DNA construct and a guide polynucleotide, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said recombinant DNA construct comprises a promoter operably linked to a nucleotide sequence encoding a plant optimized Cas endonuclease, wherein said plant optimized Cas endonuclease and guide polynucleotide are capable of forming a complex and creating a double strand break in a genomic target site of said plant.

[0026] In another embodiment, the method comprises a method for selecting a plant comprising an altered target site in its plant genome, the method comprising: a) obtaining a first plant comprising at least one Cas endonuclease capable of introducing a double strand break at a target site in the plant genome; b) obtaining a second plant comprising a guide polynucleotide that is capable of forming a complex with the Cas endonuclease of (a), wherein the guide polynucleotide does not solely comprise ribonucleic acids, c) crossing the first plant of (a) with the second plant of (b); d) evaluating the progeny of (c) for an alteration in the target site and e) selecting a progeny plant that possesses the desired alteration of said target site.

[0027] Additional embodiments of the methods and compositions of the present disclosure are disclosed below.

BRIEF DESCRIPTION OF THE DRAWINGS AND THE SEQUENCE LISTING

[0028] The disclosure can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing, which form a part of this application. The sequence descriptions and sequence listing attached hereto comply with the rules governing nucleotide and amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §§1.821-1.825. The sequence descriptions contain the three letter codes for amino acids as defined in 37 C.F.R. §§1.821-1.825, which are incorporated herein by reference.

FIGURES

[0029] FIG. **1**A shows a duplex guide polynucleotide containing a double molecule comprising a first nucleotide sequence domain (referred to as Variable Targeting domain or VT domain) that is complementary to a nucleotide sequence in a target DNA and a second nucleotide sequence domain (referred to as Cas endonuclease recognition domain or CER domain) that interacts with a Cas endonuclease polypeptide. The CER domain of the duplex guide polynucleotide comprises two separate molecules that are hybridized along a region of complementarity. The two separate molecules can be RNA, DNA, and/or RNA-DNA-combination sequences. The first molecule of the duplex guide polynucleotide comprising a VT domain linked to a CER domain (shown as crNucleotide) is referred to as "crDNA" (when composed of a contiguous stretch of DNA nucleotides) or "crRNA" (when composed of a contiguous stretch of RNA nucleotides), or "crDNA-RNA" (when composed of a combination of DNA and RNA nucleotides). The second molecule of the duplex guide polynucleotide comprising a CER domain (shown as tracrNucleotide) is referred to as "tracrRNA" (when composed of a contiguous stretch of RNA nucleotides) or "tracrDNA" (when composed of a contiguous stretch of DNA nucleotides) or "tracrDNA-RNA" (when composed of a combination of DNA and RNA nucleotides).

[0030] FIG. **1**B shows a single guide polynucleotide comprising a first nucleotide sequence domain (referred to as Variable Targeting domain or VT domain) that is complementary to a nucleotide sequence in a target DNA and a second nucleotide domain (referred to as Cas endonuclease recognition domain or CER domain) that interacts with a Cas endonuclease polypeptide. By "domain" it is meant a contiguous stretch of nucleotides that can be RNA, DNA, and/or RNA-DNA-combination sequences. The single guide polynucleotide comprises a crNucleotide (comprising a VT domain linked to a CER domain) linked to a tracrNucleotide (comprising a CER domain) with a linker nucleotide sequence (shown as a loop). The single guide polynucleotide being comprised of sequences from the crNucleotide and tracrNucleotide may be referred to as "single guide RNA" (when composed of a contiguous stretch of RNA nucleotides) or "single guide DNA" (when composed of a contiguous stretch of DNA nucleotides) or "single guide RNA-DNA" (when composed of a combination of RNA and DNA nucleotides).

[0031] FIG. **2**A-**2**C show expression cassettes for Cas9, crRNA and tracrRNA expression. FIG. **2**A shows a maize codon optimized Cas9 gene (encoding a Cas9 endonuclease) containing a potato ST-LS1 intron, a SV40 amino terminal nuclear localization sequence (SV40 NLS), and a VirD2 carboxyl terminal NLS (VirD2 NLS), operably linked to a plant ubiquitin promoter (UBI Pro) (SEQ ID NO: 5). The maize optimized Cas9 gene (just Cas9 coding sequence, no NLSs) corresponds to nucleotide positions 2037-2411 and 2601-6329 of SEQ ID NO: 5 with the potato intron residing at positions 2412-2600 of SEQ ID NO: 5.5V40 NLS is at positions 2010-2036 of SEQ ID NO: 5. VirD2 NLS is at positions 6330-6386 of SEQ ID NO: 5. FIG. **2** B shows a maize U6 polymerase III promoter operably linked to a nucleotide sequence encoding a crRNA molecule operably linked to a maize U6 terminator. The resulting maize optimized crRNA expression cassette is listed in SEQ ID NO: 8. FIG. **2** C shows a maize U6 polymerase III promoter operably linked to a nucleotide sequence encoding a tracrRNA molecule operably linked to a maize U6 PolIII terminator. The resulting maize optimized tracrRNA expression cassette is listed in SEQ ID NO: 9.

[0032] FIG. **3**A shows a duplex guide RNA/Cas9 endonuclease system and target DNA complex relative to the appropriately oriented PAM sequence (AGG) at the maize LIGCas-3 target sequence (SEQ ID NO: 14, Table 1). The duplex guide RNA (lighter gray backgrounds) comprises a crRNA molecule (SEQ ID NO: 10) containing a variable targeting domain (VT domain) base-pairing to the complementary strand of the LIGCas-3 target sequence, and a tracrRNA molecule (SEQ ID NO:11) comprising part of the CER domain. The Cas9 endonuclease is depicted in dark gray. Triangles point towards the expected site of DNA cleavage on both sense and anti-sense DNA strands.

[0033] FIG. **3**B shows a single guide RNA/Cas9 endonuclease complex interacting with the genomic LIGCas-3 target site relative to the appropriately oriented PAM sequence (AGG) at the maize genomic LIGCas-3 target site (SEQ ID NO: 14, Table 1). The single guide RNA (light gray background, SEQ ID NO: 96) is a fusion between a crRNA and tracrRNA and comprises a variable targeting domain that is complementary to one DNA strand of the double strand DNA genomic target site. The Cas9 endonuclease is shown in dark gray. Triangles point towards the expected site of DNA cleavage on both sense and anti-sense DNA strands.

[0034] FIGS. **4**A-**4**C show an alignment and count of the top 10 most frequent NHEJ mutations induced by the maize optimized guide RNA/Cas endonuclease system described herein at the maize genomic Liguleless 1 locus. The mutations were identified by deep sequencing. The PAM sequence and expected site of cleavage are also indicated. Deletions or insertions as a result of imperfect NHEJ are shown by a "–" or an italicized underlined nucleotide, respectively. In FIG. **4**A, the reference sequence (SEQ ID NO: 23) represents the unmodified LIGCas-1 locus with the target site underlined. The sequences comprising the mutations 1-10 of the LIGCas-1 target site correspond to SEQ ID NOs: 24-33, respectively. In FIG. **4**B, the reference sequence (SEQ ID NO: 23) represents the unmodified LIGCas-2 locus with the target site underlined. The sequences comprising the mutations 1-10 of the LIGCas-2 target site correspond to SEQ ID NOs: 34-43, respectively. In FIG. **4**C, the reference sequence (SEQ ID NO: 44) represents the unmodified LIGCas-3 locus with the target site underlined. The sequences comprising the mutations 1-10 of the LIGCas-3 target site correspond to SEQ ID NOs: 45-54, respectively.

[0035] FIG. **5** shows a duplex guide polynucleotide/Cas9 endonuclease system and target DNA complex relative to the appropriately oriented PAM sequence at the maize LIGCas-3 target sequence (SEQ ID NO: 14, Table 1). The duplex guide RNA (lighter gray backgrounds) comprises a crDNA molecule (SEQ ID NO: 55) containing a variable targeting domain (VT domain) base-pairing to the complementary strand of the LIGCas-3 target sequence and a tracrRNA molecule (SEQ ID NO: 11) comprising part of the CER domain. The Cas9 endonuclease is shown in dark gray. Triangles point towards the expected site of DNA cleavage on both sense and anti-sense DNA strands.

[0036] FIGS. **6**A-**6**B show alignments and counts of the top 3 most frequent NHEJ mutations induced by either a maize optimized duplex guide RNA/Cas endonuclease system (FIG. **6**A) or a maize optimized duplex guide polynucleotide/Cas endonuclease system (FIG. **6**B) described herein at the maize genomic Liguleless 1 locus. The mutations were identified by deep sequencing. The PAM sequence and expected site of cleavage are also indicated. Deletions or insertions as a result of imperfect NHEJ are shown by a "–" or an italicized underlined nucleotide, respectively. In FIG. **6**A, the NHEJ mutations originated from synthetic crRNA plus tracrRNA and Cas9 expression cassettes. The reference sequence (SEQ ID NO: 44) represents the unmodified LIGCas-3 locus with the target site underlined. The sequences comprising the mutations 1-3 of the LIGCas-3 target site correspond to SEQ ID NOs: 56-58, respectively. In FIG. **6**B, the NHEJ mutations originated from synthetic crDNA plus tracrRNA and Cas9 expression cassettes. The reference sequence (SEQ ID NO: 44) represents the unmodified LIGCas-3 locus with the target site underlined. The sequences comprising the mutations 1-3 of the LIGCas-3 target site correspond to SEQ ID NOs: 59-61, respectively.

[0037] FIG. **7** illustrates the disruption of the yeast ADE2 gene on chromosome 15 with URA3 coding sequence and 305 bp of duplicated ADE2 gene sequence resulting in the ADE:URA3:DE2 yeast screening strain.

[0038] FIG. **8** illustrates the scheme by which cleavage activity may be monitored in the yeast ADE:URA3:DE2 screening strain. If the URA3 target site is cleaved, the ADE2 sequence duplications flanking the URA3 coding sequence may be used as template for homologous recombination repair of the DNA double strand break. As depicted by dashed lines leading from the regions of ADE2 sequence duplication in the ADE:URA3:DE2 configuration to the ADE2 configuration, homologous recombination mediated repair of the double strand break results in the loss of the URA3 gene coding sequence and the gain of a functional ADE2 gene.

[0039] FIG. **9** shows the numerical scale and corresponding red/white sectoring of yeast colonies used to quantify cleavage activity. Since the sectoring phenotype is a qualitative measure of cleavage activity, a 0-4 numerical scoring system was implemented. A score of 0 indicates that no white sectors (no cutting) were observed; a score of 4 indicates completely white colonies (complete cutting of the recognition site); scores of 1-3 indicate intermediate white sectoring phenotypes (and intermediate degrees of recognition site cutting).

Sequences

[0040] SEQ ID NO: 1 is the nucleotide sequence of the Cas9 gene from *Streptococcus pyogenes* M1 GAS (SF370).

[0041] SEQ ID NO: 2 is the nucleotide sequence of the potato ST-LS1 intron.

[0042] SEQ ID NO: 3 is the amino acid sequence of SV40 amino N-terminal.

[0043] SEQ ID NO: 4 is the amino acid sequence of *Agrobacterium tumefaciens* bipartite VirD2 T-DNA border endonuclease carboxyl terminal.

[0044] SEQ ID NO: 5 is the nucleotide sequence of an expression cassette expressing the maize optimized Cas9.

[0045] SEQ ID NO: 6 is the nucleotide sequence of the maize U6 polymerase III promoter.

[0046] SEQ ID NO: 7 is the amino acid sequence a SV40 nuclear localization signal.

[0047] SEQ ID NO: 8 is the nucleotide sequence of a maize optimized crRNA expression cassette containing the variable targeting domain targeting the LIGCas-3 target sequence.

[0048] SEQ ID NO: 9 is the nucleotide sequence of a maize optimized tracrRNA expression cassette.

[0049] SEQ ID NO: 10 is the nucleotide sequence of a crRNA containing a variable targeting domain targeting the LIGCas-3 target sequence.

[0050] SEQ ID NO: 11 is the nucleotide sequence of the tracrRNA from *Streptococcus pyogenes* M1 GAS (SF370)>

[0051] SEQ ID NO: 12 is the nucleotide sequence of the maize genomic target site LIGCas-1 plus PAM sequence.

[0052] SEQ ID NO: 13 is the nucleotide sequence of the maize genomic target site LIGCas-2 plus PAM sequence.

[0053] SEQ ID NO: 14 is the nucleotide sequence of the maize genomic target site LIGCas-3 plus PAM sequence.

[0054] SEQ ID NOs: 15-22 are nucleotide sequences of PCR primers.

[0055] SEQ ID NO: 23 is the nucleotide sequence of the unmodified reference sequence for LIGCas-1 and LIGCas-2 locus (FIG. **4**A-**4**B)

[0056] SEQ ID NOs: 24-33 are the nucleotide sequences of mutations 1-10 for the LIGCas-1 locus (FIG. **4**A).

[0057] SEQ ID NOs: 34-43 are the nucleotide sequences of mutations 1-10 for the LIGCas-2 locus (FIG. **4**B).

[0058] SEQ ID NO: 44 is the nucleotide sequence of the unmodified reference sequence for LIGCas-3 (FIG. **4**C)

[0059] SEQ ID NOs: 45-54 are the nucleotide sequences of mutations 1-10 for the LIGCas-3 locus (FIG. **4**C).

[0060] SEQ ID NO: 55 is the nucleotide sequence of a crDNA (comprised of deoxyribonucleic acids) containing a variable targeting domain targeting the LIGCas-3 target sequence

[0061] SEQ ID NOs: 56-58 are the nucleotide sequences of mutations 1-3 for the LIGCas-3 locus (originating from synthetic crRNA plus tracrRNA and Cas9 expression cassettes) (FIG. **6**A).

[0062] SEQ ID NOs: 59-61 are the nucleotide sequences of mutations 1-3 for the LIGCas-3 locus (originating from synthetic crDNA plus tracrRNA and Cas9 expression cassettes) (FIG. **6**B).

[0063] SEQ ID NO: 62 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crRNA) that does not include any modification to its ribonucleotide sequence.

[0064] SEQ ID NO: 63 is the nucleotide sequence of a CER domain of a crNucleotide (crRNA) that does not include any modification to its ribonucleotide sequence.

[0065] SEQ ID NO: 64 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crRNA), that includes phosphorothioate bonds at the 5' end of its nucleotide sequence (G*C*G*). In the sequence listing, the first N at the 5' end represents a G ribonucleotide with a phosphorothioate bond, the second N represents a C ribonucleotide with a phosphorothioate bond and third N represents a G ribonucleotide with a Phosphorothioate bond.

[0066] SEQ ID NO: 65 is the nucleotide sequence of a CER domain of a crNucleotide (crRNA) that includes phosphorothioate bonds near the 3' end of its nucleotide sequence (U*U*U*). In the sequence listing, the Ns at the nineteenth, twentieth and twenty-first positions represent U ribonucleotides with phosphorothioate bonds.

[0067] SEQ ID NO: 66 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crRNA) that includes 2'-O-methyl RNA nucleotides at it 5' end (mGmCmG). In the sequence listing, the first N at the 5' end represents a G 2'-O-methyl ribonucleotide, the second N represents a C 2'-O-methyl ribonucleotide and the third N represents a G 2'-O-Methyl ribonucleotide.

[0068] SEQ ID NO: 67 is the nucleotide sequence of a CER domain of a crNucleotide (crRNA) that includes 2'-O-methyl RNA nucleotides near the 3' end of its nucleotide sequence (mUmUmG). In the sequence listing, the N at the twentieth position represents a U 2'-O-Methyl ribonucleotide, the N at the twenty-first position represents a U 2'-O-Methyl ribonucleotide and the N at the twenty-second position represents a G 2'-O-Methyl ribonucleotide.

[0069] SEQ ID NO: 68 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crRNA) that includes 2'-O-Methyl RNA nucleotides for each nucleotide. In the sequence listing, the first N at the 5' end represents a G 2'-O-Methyl ribonucleotide, a N at the second position represents a C 2'-O-Methyl ribonucleotide, a N at the third position represents a G 2'-O-Methyl ribonucleotide, a N at the fourth position represents an U 2'-O-Methyl ribonucleotide, a N at the fifth position represents an A 2'-O-Methyl ribonucleotide, a N at the sixth position represents a C 2'-O-Methyl ribonucleotide, a N at the seventh position represents a G 2'-O-Methyl ribonucleotide, a N at the eighth position represents a C 2'-O-Methyl ribonucleotide, a N at the ninth position represents a G 2'-O-Methyl ribonucleotide, a N at the tenth position represents an U 2'-O-Methyl ribonucleotide, a N at the eleventh position represents an A 2'-O-Methyl ribonucleotide, a N at the twelfth position represents C 2'-O-Methyl ribonucleotide, a N at the thirteenth position represents a G 2'-O-Methyl ribonucleotide, a N at the fourteenth position represents an U 2'-O-Methyl ribonucleotide, a N at the fifteenth position represents a G 2'-β-Methyl ribonucleotide, a N at the sixteenth position represents an U 2'-O-Methyl ribonucleotide and a N seventeenth position represents a G 2'-O-Methyl ribonucleotide.

[0070] SEQ ID NO: 69 is the nucleotide sequence of a CER domain of a crNucleotide (crRNA) that include 2'-O-Methyl RNA nucleotides for each nucleotide.

[0071] In the sequence listing, the first N at the 5' end represents a G 2'-O-Methyl ribonucleotide, a N at the second position represents an U 2'-O-Methyl ribonucleotide, a N at the third position represents an U 2'-O-Methyl ribonucleotide, a N at the fourth position represents an U 2'-O-Methyl ribonucleotide, a N at the fifth position represents an U 2'-O-Methyl ribonucleotide, a N at the sixth position represents an A 2'-O-Methyl ribonucleotide, a N at the seventh position represents a G 2'-O-Methyl ribonucleotide, a N at the eighth position represents an A 2'-O-Methyl ribonucleotide, a N at the ninth position represents a G 2'-O-Methyl ribonucleotide, a N at the tenth position represents a C 2'-O-Methyl ribonucleotide, a N at the eleventh position represents an U 2'-O-Methyl ribonucleotide, a N at the twelfth position represents an A 2'-O-Methyl ribonucleotide, a N at the thirteenth position represents an U 2'-O-Methyl ribonucleotide, a N at the fourteenth position represents a G 2'-O-Methyl ribonucleotide, a N at the fifteenth position represents a C 2'-O-Methyl ribonucleotide, a N at the sixteenth position represents an U 2'-O-Methyl ribonucleotide, a N at the seventeenth position represents a G 2'-O-Methyl ribonucleotide, a N at the eighteen position represents an U 2'-O-Methyl ribonucleotide, a N at the nineteenth position represents an U 2'-O-Methyl ribonucleotide, a N at the twentieth position represents an U 2'-O-Methyl ribonucleotide, a N at the twenty-first position represents an U 2'-O-Methyl ribonucleotide and a N at the twenty-second position represents a G 2'-O-Methyl ribonucleotide.

[0072] SEQ ID NO: 70 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crDNA) that does not include any modification to its deoxyribonucleotide sequence.

[0073] SEQ ID NO: 71 is the nucleotide sequence of a CER domain of a crNucleotide (crDNA) that does not include any modification to its deoxyribonucleotide sequence.

[0074] SEQ ID NO: 72 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crDNA), which includes one Locked Nucleic Acid nucleotide (+T) in its nucleotide sequence. In the sequence listing, an N at the sixteenth position represents a T Locked Nucleic Acid base.

[0075] SEQ ID NO: 73 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crDNA), which includes three Locked Nucleic Acid nucleotide (+C, +T, +T) in its nucleotide sequence. In the sequence listing, an N at the twelfth position represents a C Locked Nucleic Acid base, a N at the fourteenth position represents a T Locked Nucleic Acid base and a N at the sixteenth position represents a T Locked Nucleic Acid base.

[0076] SEQ ID NO: 74 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crDNA), that includes six Locked Nucleic Acid nucleotide (+C, +C, +T, +C, +T, +T) in its nucleotide sequence. In the sequence listing, a N at the sixth position represents a C Locked Nucleic Acid base, a N at the eighth position represents a C Locked Nucleic Acid base, a N at the tenth position represents a T Locked Nucleic Acid base, a N at the twelfth position represents a C Locked Nucleic Acid base, a N at the fourteenth position represents a T Locked Nucleic Acid base and a N at the sixteenth position represents a T Locked Nucleic Acid base.

[0077] SEQ ID NO: 75 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crDNA), that includes three Locked Nucleic Acid nucleotide (+C, +T, +T) in its nucleotide sequence and phosphorothioate bonds near the 5' end of its nucleotide sequence (G*C*G*). In the sequence listing, a first N at the 5' end represents a G deoxyribonucleotide with a phosphorothioate bond, a N at the second position represents a C deoxyribonucleotide with a phosphorothioate bond, a N at the third position represents a G deoxyribonucleotide with a phosphorothioate bond, a N at the twelfth position represents a C Locked Nucleic Acid base, a N at the fourteenth position represents a T Locked Nucleic Acid base and a N at the sixteenth position represents a T Locked Nucleic Acid base.

[0078] SEQ ID NO: 76 is the nucleotide sequence of a CER domain of a crNucleotide (crDNA) that includes three Locked Nucleic Acid nucleotide (T*T*T) near the 3' end of the nucleotide sequence. In the sequence listing, the Ns at the nineteenth, twentieth and twenty-first positions represent T deoxyribonucleotides with Phosphorothioate bonds.

[0079] SEQ ID NO: 77 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crDNA) that includes one 5-Methyl dC nucleotide in its nucleotide sequence. In the sequence listing, a N at the twelfth position represents a 5-Methyl dC deoxyribonucleotide.

[0080] SEQ ID NO: 78 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crDNA) that includes three 5-Methyl dC nucleotide in its nucleotide sequence. In the sequence listing, Ns at the sixth, eighth and twelfth positions represent 5-Methyl dC deoxyribonucleotides.

[0081] SEQ ID NO: 79 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crDNA) that includes one 2,6-diaminopurine nucleotide in its nucleotide sequence. In the sequence listing, a N at the eleventh position represents a 2,6-Diaminopurine deoxyribonucleotide.

[0082] SEQ ID NO: 80 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crDNA) that includes two 2,6-diaminopurine nucleotides in its nucleotide sequence. In the sequence listing, a Ns at the fifth and eleventh positions represent 2,6-diaminopurine deoxyribonucleotides.

[0083] SEQ ID NO: 81 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crDNA) that includes Locked Nucleic Acid nucleotides near the 5' end of its nucleotide sequence. In the sequence listing, the first N at the 5' end represent a G Locked Nucleic Acid base, second N represents a C Locked Nucleic Acid base and third N represents a G Locked Nucleic Acid base.

[0084] SEQ ID NO: 82 is the nucleotide sequence of a CER domain of a crNucleotide (crDNA) that includes Locked Nucleic Acid nucleotides near the 3' end of the nucleotide sequence. In the sequence listing, a N at the twentieth position represents a T Locked Nucleic Acid base, a N at the twenty-first position represents a T Locked Nucleic Acid base and a N at the twenty-second position represents a G Locked Nucleic Acid base.

[0085] SEQ ID NO: 83 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crDNA) that includes phosphorothioate bonds near the 5' end of its nucleotide sequence. In the sequence listing, a first N at the 5' end represents a G deoxyribonucleotide with a phosphorothioate bond, second N represents a C deoxyribonucleotide with a phosphorothioate bond and third N represents a G deoxyribonucleotide with a phosphorothioate bond.

[0086] SEQ ID NO: 84 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crDNA) that includes 2'-O-Methyl RNA nucleotides near the 5' end of its nucleotide sequence. In the sequence listing, a first N at the 5' end represents a G 2'-O-Methyl ribonucleotide, second N represents a C 2'-O-Methyl ribonucleotide and third N represents a G 2'-O-Methyl ribonucleotide.

[0087] SEQ ID NO: 85 is the nucleotide sequence of a CER domain of a crNucleotide (crDNA) that includes 2'-O-Methyl RNA nucleotides near the 3' end of the nucleotide sequence. In the sequence listing, a N at the twentieth position represents a U 2'-O-Methyl ribonucleotide, a N at the twenty-first position represents a U 2'-O-Methyl ribonucleotide and a N at the twenty-second position represent G 2'-O-Methyl ribonucleotide.

[0088] SEQ ID NO: 86 is the nucleotide sequence of a variable targeting domain of a crNucleotide (crDNA) that includes 2'-O-Methyl RNA nucleotides at each nucleotide except T of its nucleotide sequence. In the sequence listing, a first N at the 5' end represents a G 2'-O-Methyl ribonucleotide, a N at the second position represents a C 2'-O-Methyl ribonucleotide, a N at the third position represents a G 2'-O-Methyl ribonucleotide, a N at the fifth position represents an A 2'-O-Methyl ribonucleotide, a N at the sixth position represents a C 2'-O-Methyl ribonucleotide, a N at the seventh position represents a G 2'-O-Methyl ribonucleotide, a N at the eighth position represents a C 2'-O-Methyl ribonucleotide, a N at the ninth position represents a G 2'-O-Methyl ribonucleotide, a N at the eleventh position represents an A 2'-O-Methyl ribonucleotide, a N at the twelfth position represents C 2'-O-Methyl ribonucleotide, a N at the thirteenth position represents a G 2'-O-Methyl ribonucleotide, a N at the fifteenth position represents a G 2'-O-Methyl ribonucleotide and a N at the seventeenth position represents a G 2'-O-Methyl ribonucleotide.

[0089] SEQ ID NO: 87 is the nucleotide sequence of a CER domain of a crNucleotide (crDNA) that includes 2'-O-Methyl RNA nucleotides at each nucleotide except T the nucleotide sequence. In the sequence listing, a first N at the 5' end represents a G 2'-O-Methyl ribonucleotide, a N at the sixth position represents an A 2'-O-Methyl ribonucleotide, a N at the seventh position represents a G 2'-O-Methyl ribonucleotide, a N at the eighth position represents an A 2'-O-Methyl ribonucleotide, a N at the ninth position represents a G 2'-O-Methyl ribonucleotide, a N at the tenth position represents a C 2'-O-Methyl ribonucleotide, a N at the twelfth position represents an A 2'-O-Methyl ribonucleotide, a N at the fourteenth position represents a G 2'-O-Methyl ribonucleotide, a N at the fifteenth position represents a C 2'-O-Methyl ribonucleotide, a N at the seventeenth position represents a G 2'-O-Methyl ribonucleotide and a N at the twenty-second position represents a G 2'-O-Methyl ribonucleotide.

[0090] SEQ ID NO: 88 is the nucleotide sequence of the *Saccharomyces cerevisiae* codon optimized Cas9.

[0091] SEQ ID NO: 89 is the nucleotide sequence of the T7 promoter from bacteriophage T7.

[0092] SEQ ID NO: 90 is the nucleotide sequence of the ADE:URA3:DE2 target sequence (PAM sequence not included)

[0093] SEQ ID NO: 91-95 are the nucleotide sequences of Cas9 endonucleases.

[0094] SEQ ID NO: 96 is the nucleotide sequence of a single guide RNA targeting the LIGCas-3 target sequence (FIG. 3B).

DETAILED DESCRIPTION

[0095] The present disclosure includes compositions and methods for genome modification of a target sequence in the genome of a cell. The methods and compositions employ a guide polynucleotide/Cas endonuclease system to provide an effective system for modifying target sites within the genome of a cell. Cells include, but are not limited to, animal, bacterial, fungal, insect, yeast, and plant cells as well as plants and seeds produced by the methods described herein. Once a genomic target site is identified, a variety of methods can be employed to further modify the target sites such that they contain a variety of polynucleotides of interest. Breeding methods utilizing a guide polynucleotide/Cas endonuclease system are also disclosed. Compositions and methods are also provided for editing a nucleotide sequence in the genome of a cell. The nucleotide sequence to be edited (the nucleotide sequence of interest) can be located within or outside a target site that is recognized by a Cas endonuclease.

[0096] CRISPR loci (Clustered Regularly Interspaced Short Palindromic Repeats) (also known as SPIDRs—SPacer Interspersed Direct Repeats) constitute a family of recently described DNA loci. CRISPR loci consist of short and highly conserved DNA repeats (typically 24 to 40 bp, repeated from 1 to 140 times—also referred to as CRISPR-repeats) which are partially palindromic. The repeated sequences (usually specific to a species) are interspaced by variable sequences of constant length (typically 20 to 58 by depending on the CRISPR locus (WO2007/025097 published Mar. 1, 2007).

[0097] CRISPR loci were first recognized in *E. coli* (Ishino et al. (1987) J. Bacterial. 169:5429-5433; Nakata et al. (1989) J. Bacterial. 171:3553-3556). Similar interspersed short sequence repeats have been identified in Haloferax mediterranei, Streptococcus pyogenes, Anabaena, and *Mycobacterium tuberculosis* (Groenen et al. (1993) Mol. Microbiol. 10:1057-1065; Hoe et al. (1999) Emerg. Infect. Dis. 5:254-263; Masepohl et al. (1996) Biochim. Biophys. Acta 1307: 26-30; Mojica et al. (1995) Mol. Microbiol. 17:85-93). The CRISPR loci differ from other SSRs by the structure of the repeats, which have been termed short regularly spaced repeats (SRSRs) (Janssen et al. (2002) OMICS J. Integ. Biol. 6:23-33; Mojica et al. (2000) Mol. Microbiol. 36:244-246). The repeats are short elements that occur in clusters, that are always regularly spaced by variable sequences of constant length (Mojica et al. (2000) Mol. Microbiol. 36:244-246).

[0098] Cas gene includes a gene that is generally coupled, associated or close to or in the vicinity of flanking CRISPR loci. The terms "Cas gene", "CRISPR-associated (Cas) gene" are used interchangeably herein. A comprehensive review of the Cas protein family is presented in Haft et al. (2005) Computational Biology, PLoS Comput Biol 1(6): e60. doi: 10.1371/journal.pcbi.0010060. As described therein, 41 CRISPR-associated (Cas) gene families are described, in addition to the four previously known gene families. It shows that CRISPR systems belong to different classes, with different repeat patterns, sets of genes, and species ranges. The number of Cas genes at a given CRISPR locus can vary between species.

[0099] Cas endonuclease relates to a Cas protein encoded by a Cas gene, wherein said Cas protein is capable of introducing a double strand break into a DNA target sequence. The Cas endonuclease unwinds the DNA duplex in close proximity of the genomic target site and cleaves both DNA strands upon recognition of a target sequence by a guide polynucleotide, but only if the correct protospacer-adjacent motif (PAM) is approximately oriented at the 3' end of the target sequence (FIG. 3A, FIG. 3B).

[0100] In one embodiment, the Cas endonuclease is a Cas9 endonuclease that is capable of introducing a double strand break at a DNA target site, wherein the DNA cleavage at a specific location is enabled by a) base-pairing complementary between the DNA target site and the variable targeting domain of the guide polynucleotide, and b) the presence of a short protospacer adjacent motif (PAM) immediately adjacent to the DNA target site.

[0101] In one embodiment, the Cas endonuclease gene is a Cas9 endonuclease, such as but not limited to, Cas9 genes listed in SEQ ID NOs: 462, 474, 489, 494, 499, 505, and 518 of WO2007/025097 published Mar. 1, 2007, and incorporated herein by reference. In another embodiment, the Cas endonuclease gene is plant, maize or soybean optimized Cas9 endonuclease (FIG. 1 A). In another embodiment, the Cas endonuclease gene is operably linked to a SV40 nuclear targeting signal upstream of the Cas codon region and a bipartite VirD2 nuclear localization signal (Tinland et al. (1992) Proc. Natl. Acad. Sci. USA 89:7442-6) downstream of the Cas codon region.

[0102] In one embodiment, the Cas endonuclease gene is a Cas9 endonuclease gene of SEQ ID NO: 1, 91, 92, 93, 94, 95 or nucleotides 2037-6329 of SEQ ID NO:5, or any functional fragment or variant thereof.

[0103] The terms "functional fragment", "fragment that is functionally equivalent" and "functionally equivalent fragment" are used interchangeably herein. These terms refer to a portion or subsequence of a Cas endonuclease sequence in which the ability to create a double-strand break is retained.

[0104] The terms "functional variant", "Variant that is functionally equivalent" and "functionally equivalent variant" are used interchangeably herein. These terms refer to a variant of the Cas endonuclease in which the ability create a double-strand break is retained. Fragments and variants can be obtained via methods such as site-directed mutagenesis and synthetic construction.

[0105] In one embodiment, the Cas endonuclease gene is a plant codon optimized *streptococcus pyogenes* Cas9 gene that can recognize any genomic sequence of the form N(12-30)NGG can in principle be targeted.

[0106] Endonucleases are enzymes that cleave the phosphodiester bond within a polynucleotide chain, and include restriction endonucleases that cleave DNA at specific sites without damaging the bases. Restriction endonucleases include Type I, Type II, Type III, and Type IV endonucleases, which further include subtypes. In the Type I and Type III systems, both the methylase and restriction activities are contained in a single complex. Endonucleases also include meganucleases, also known as homing endonucleases (HEases), which like restriction endonucleases, bind and cut at a specific recognition site, however the recognition sites for meganucleases are typically longer, about 18 bp or more. (patent application WO-PCT PCT/US12/30061 filed on Mar. 22, 2012) Meganucleases have been classified into four families based on conserved sequence motifs (Belfort M, and Perlman P S J. Biol. Chem. 1995; 270:30237-30240). These motifs participate in the coordination of metal ions and hydrolysis of phosphodiester bonds. HEases are notable for their long recognition sites, and for tolerating some sequence polymorphisms in their DNA substrates. The naming convention for meganuclease is similar to the convention for other restriction endonuclease. Meganucleases are also characterized by prefix F—, I—, or PI— for enzymes encoded by free-standing ORFs, introns, and inteins, respectively. One step in the recombination process involves polynucleotide cleavage at or near the recognition site. This cleaving activity can be used to produce a double-strand break. For reviews of site-specific recombinases and their recognition sites, see, Sauer (1994) Curr Op Biotechnol 5:521-7; and Sadowski (1993) FASEB 7:760-7. In some examples the recombinase is from the Integrase or Resolvase families.

[0107] TAL effector nucleases are a new class of sequence-specific nucleases that can be used to make double-strand breaks at specific target sequences in the genome of a plant or other organism. (Miller et al. (2011) *Nature Biotechnology* 29:143-148). Zinc finger nucleases (ZFNs) include engineered double-strand break inducing agents comprised of a zinc finger DNA binding domain and a double-strand-break-inducing agent domain. Recognition site specificity is conferred by the zinc finger domain, which typically comprising two, three, or four zinc fingers, for example having a C2H2 structure, however other zinc finger structures are known and have been engineered. Zinc finger domains are amenable for designing polypeptides which specifically bind a selected polynucleotide recognition sequence. ZFNs consist of an engineered DNA-binding zinc finger domain linked to a non-specific endonuclease domain, for example nuclease domain from a Type IIs endonuclease such as FokI. Additional functionalities can be fused to the zinc-finger binding domain, including transcriptional activator domains, transcription repressor domains, and methylases. In some examples, dimerization of nuclease domain is required for cleavage activity. Each zinc finger recognizes three consecutive base pairs in the target DNA. For example, a 3 finger domain recognized a sequence of 9 contiguous nucleotides, with a dimerization requirement of the nuclease, two sets of zinc finger triplets are used to bind a 18 nucleotide recognition sequence.

[0108] In one embodiment of the disclosure, the composition comprises a plant or seed comprising a guide polynucleotide and a Cas9 endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said Cas9 endonuclease and guide polynucleotide are capable of forming a complex and creating a double strand break in a genomic target site of said plant.

[0109] Bacteria and archaea have evolved adaptive immune defenses termed Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated (Cas) systems that use short RNA to direct degradation of foreign nucleic acids (WO2007/025097 published Mar. 1, 2007, and incorporated herein by reference.) The type II CRISPR/Cas system from bacteria employs a crRNA and tracrRNA to guide the Cas endonuclease to its DNA target. The crRNA (CRISPR RNA) contains the region complementary to one strand of the double strand DNA target and base pairs with the tracrRNA (trans-activating CRISPR RNA) forming a RNA duplex that directs the Cas endonuclease to cleave the DNA target.

[0110] As used herein, the term "guide polynucleotide", relates to a polynucleotide sequence that can form a complex with a Cas endonuclease and enables the Cas endonuclease to recognize and optionally cleave a DNA target site. The guide polynucleotide can be a single molecule or a double molecule. The guide polynucleotide sequence can be a RNA sequence, a DNA sequence, or a combination thereof (a RNA-DNA combination sequence). Optionally, the guide polynucleotide can comprise at least one nucleotide, phosphodiester bond or linkage modification such as, but not limited, to Locked Nucleic Acid (LNA), 5-methyl dC, 2,6-Diaminopurine, 2'-Fluoro A, 2'-Fluoro U, 2'-O-Methyl RNA, Phosphorothioate bond, linkage to a cholesterol molecule, linkage to a polyethylene glycol molecule, linkage to a spacer 18 (hexaethylene glycol chain) molecule, or 5' to 3' covalent linkage resulting in circularization. In some embodiment of this disclosure, the guide polynucleotide does not solely comprise ribonucleic acids (RNAs). A guide polynucleotide that solely comprises ribonucleic acids is also referred to as a "guide RNA".

[0111] The guide polynucleotide can be a double molecule (also referred to as duplex guide polynucleotide) comprising a first nucleotide sequence domain (referred to as Variable Targeting domain or VT domain) that is complementary to a nucleotide sequence in a target DNA and a second nucleotide sequence domain (referred to as Cas endonuclease recognition domain or CER domain) that interacts with a Cas endonuclease polypeptide (FIG. **1**A). The CER domain of the double molecule guide polynucleotide comprises two separate molecules that are hybridized along a region of complementarity (FIG. **1**A). The two separate molecules can be RNA, DNA, and/or RNA-DNA—combination sequences. In one embodiment of this disclosure, the duplex guide polynucleotide does not solely comprise ribonucleic acids (RNAs) as show in, for example, but not limiting to, FIG. **3**A). In some embodiments, the first molecule of the duplex guide polynucleotide comprising a VT domain linked to a CER domain (shown as "crNucleotide" in FIG. **1**A) is referred to as "crDNA" (when composed of a contiguous stretch of DNA nucleotides) or "crRNA" (when composed of a contiguous stretch of RNA nucleotides), or "crDNA-RNA" (when composed of a combination of DNA and RNA nucleotides). In some embodiments the second molecule of the duplex guide polynucleotide comprising a CER domain (shown as tracrNucleotide in FIG. **1**A) is referred to as "tracrRNA" (when composed of a contiguous stretch of RNA nucleotides) or "tracrDNA" (when composed of a contiguous stretch of DNA nucleotides) or "tracrDNA-RNA" (when composed of a combination of DNA and RNA nucleotides).

[0112] The guide polynucleotide can also be a single molecule comprising a first nucleotide sequence domain (referred to as Variable Targeting domain or VT domain) that is complementary to a nucleotide sequence in a target DNA and a second nucleotide domain (referred to as endonuclease recognition domain or CER domain) that interacts with a Cas endonuclease polypeptide (FIG. **1**B). By "domain" it is meant a contiguous stretch of nucleotides that can be RNA, DNA, and/or RNA-DNA-combination sequence. The VT domain and/or the CER domain of a single guide polynucleotide can comprise a RNA sequence, a DNA sequence, or a RNA-DNA-combination sequence. In some embodiments the single guide polynucleotide comprises a crNucleotide (comprising a VT domain linked to a CER domain) linked to a tracrNucleotide (comprising a CER domain), wherein the linkage is a nucleotide sequence comprising a RNA sequence, a DNA sequence, or a RNA-DNA combination sequence (FIG. **1**B). The single guide polynucleotide being comprised of sequences from the crNucleotide and tracrNucleotide may be referred to as "single guide RNA" (when composed of a contiguous stretch of RNA nucleotides) or "single guide DNA" (when composed of a contiguous stretch of DNA nucleotides) or "single guide RNA-DNA" (when composed of a combination of RNA and DNA nucleotides).

[0113] One advantage of using a single guide polynucleotide versus a duplex guide polynucleotide is that only one expression cassette needs to be made to express the single guide polynucleotide.

[0114] The term "variable targeting domain" or "VT domain" is used interchangeably herein and refers to a nucleotide sequence that is complementary to one strand (nucleotide sequence) of a double strand DNA target site. The % 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 target 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 (see for example modifications described herein), or any combination thereof.

[0115] The term "Cas endonuclease recognition domain" or "CER domain" of a guide polynucleotide is used interchangeably herein and relates to a nucleotide sequence (such as a second nucleotide sequence domain of a guide polynucleotide), that interacts with a Cas endonuclease polypeptide. The CER domain can be composed of a DNA sequence, a RNA sequence, a modified DNA sequence, a modified RNA sequence (see for example modifications described herein), or any combination thereof.

[0116] The nucleotide sequence linking the crNucleotide and the tracrNucleotide of a single guide polynucleotide can comprise a RNA sequence, a DNA sequence, or a RNA-DNA combination sequence. In one embodiment, the nucleotide sequence linking the crNucleotide and the tracrNucleotide of a single guide polynucleotide can be at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100 nucleotides in length. In another embodiment, the nucleotide sequence linking the crNucleotide and the tracrNucleotide of a single guide polynucleotide can comprise a tetraloop sequence, such as, but not limiting to a GAAA tetraloop sequence.

[0117] 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.

[0118] In one embodiment of the disclosure, the composition comprises a guide polynucleotide comprising: (i) a first nucleotide sequence domain (VT domain) that is complementary to a nucleotide sequence in a target DNA; and, (ii) a second nucleotide sequence domain (CER domain) that interacts with a Cas endonuclease, wherein the first nucleotide sequence domain and the second nucleotide sequence domain are composed of deoxyribonucleic acids (DNA), ribonucleic acids (RNA), or a combination thereof, wherein the guide polynucleotide does not solely comprise ribonucleic acids. The % complementation between the first nucleotide sequence domain (Variable Targeting 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%.

[0119] In one embodiment of the disclosure, the first nucleotide sequence domain (VT domain) and the second nucleotide sequence domain (CER domain) of the guide polynucleotide are located on a single molecule. In another embodiment, the second nucleotide sequence domain (Cas Endonuclease Recognition domain) comprises two separate molecules that are capable of hybridizing along a region of complementarity.

[0120] In another embodiment, the composition comprises a guide polynucleotide, wherein the first nucleotide sequence domain is a DNA sequence and the second nucleotide sequence domain is selected from the group consisting of a DNA sequence, a RNA sequence, and a combination thereof.

[0121] In one embodiment, the composition comprises a guide polynucleotide, wherein the first nucleotide sequence domain (VT domain) is a DNA sequence and the second nucleotide sequence domain (CER domain) is selected from the group consisting of a DNA sequence, a RNA sequence, and a combination thereof.

[0122] The guide polynucleotide and Cas endonuclease are capable of forming a complex, referred to as the "guide polynucleotide/Cas endonuclease complex", that enables the Cas endonuclease to introduce a double strand break at a DNA target site.

[0123] In one embodiment, the composition comprises a guide polynucleotide/Cas endonuclease complex wherein the guide polynucleotide comprises (i) a first nucleotide sequence domain that is complementary to a nucleotide sequence in a target DNA; and (ii) a second nucleotide sequence domain that interacts with a Cas endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site.

[0124] In another embodiment, the composition comprises a guide polynucleotide/Cas endonuclease complex, wherein the first nucleotide sequence domain (VT domain) and the second nucleotide sequence domain (CER domain) of the guide polynucleotide are composed of deoxyribonucleic acids (DNA), ribonucleic acids (RNA), or a combination thereof, wherein the guide polynucleotide does not solely comprise ribonucleic acids.

[0125] In one embodiment the guide polynucleotide can be introduce into the plant cell directly using particle bombardment.

[0126] When the guide polynucleotide comprises solely of RNA sequences (also referred to as "guide RNA") it can be introduced indirectly by introducing a recombinant DNA molecule comprising the corresponding guide DNA sequence operably linked to a plant specific promoter that is capable of transcribing the guide polynucleotide in said plant cell. The term "corresponding guide DNA" refers to a DNA molecule that is identical to the RNA molecule but has a "T" substituted for each "U" of the RNA molecule.

[0127] In some embodiments, the guide polynucleotide is introduced via particle bombardment or *Agrobacterium* transformation of a recombinant DNA construct comprising the corresponding guide DNA operably linked to a plant U6 polymerase III promoter.

[0128] The terms "target site", "target sequence", "target DNA", "target locus", "genomic target site", "genomic target sequence", and "genomic target locus" are used interchangeably herein and refer to a polynucleotide sequence in the genome (including choloroplastic and mitochondrial DNA) of a cell at which a double-strand break is induced in the cell genome by a Cas endonuclease. The target site can be an endogenous site in the genome of an cell or organism, or alternatively, the target site can be heterologous to the cell or organism and thereby not be naturally occurring in the genome, 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 or organism and is at the endogenous or native position of that target sequence in the genome of a cell or organism. Cells include, but are not limited to animal, bacterial, fungal, insect, yeast, and plant cells as well as plants and seeds produced by the methods described herein.

[0129] In one embodiments, the target site can be similar to a DNA recognition site or target site that that is specifically recognized and/or bound by a double-strand break inducing agent such as a LIG3-4 endonuclease (US patent publication 2009-0133152 A1 (published May 21, 2009) or a MS26++ meganuclease (U.S. patent application Ser. No. 13/526,912 filed Jun. 19, 2012).

[0130] 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 or organism, such as but not limiting to a plant or yeast. 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 or organism.

[0131] 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).

[0132] Methods for modifying a genomic target site of an organism such as but not limiting to a plant or yeast are disclosed herein.

[0133] In one embodiment, a method for modifying a target site in the genome of a plant cell comprises introducing a guide polynucleotide into a cell having a Cas endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site. This method can further comprise further comprising identifying at least one cell that has a modification at said target, wherein the modification at said target site is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii). This method can also further comprise introducing a donor DNA to said cell, wherein said donor DNA comprises a polynucleotide of interest.

[0134] Further provided is a method for method for modifying a target site in the genome of a cell, the method comprising introducing a guide polynucleotide and a Cas endonuclease into a cell, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site. This method can further comprise further comprising identifying at least one cell that has a modification at said target, wherein the modification at said target site is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii). This method can also further comprise introducing a donor DNA to said cell, wherein said donor DNA comprises a polynucleotide of interest.

[0135] Further provided is a method for modifying a target site in the genome of a cell, the method comprising: a) introducing into a cell a crNucleotide, a first recombinant DNA construct capable of expressing a tracrRNA, and a second recombinant DNA capable of expressing a Cas endonuclease, wherein said crNucleotide is a deoxyribonucleotide sequence or a combination of a deoxyribonucleotide and ribonucleotide sequence, wherein said crNucleotide, said tracrRNA and said Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site; and, b) identifying at least one cell that has a modification at said target site, wherein the modification is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

[0136] Further provided is a method for method for modifying a target site in the genome of a cell, the method comprising: a) introducing into a cell a tracrNucleotide, a first recombinant DNA construct capable of expressing a crRNA and a second recombinant DNA capable of expressing a Cas endonuclease, wherein said tracrNucleotide is selected a deoxyribonucleotide sequence or a combination of a deoxyribonucleotide and ribonucleotide sequence, wherein said tracrNucleotide, said crRNA and said Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site; and, b) identifying at least one cell that has a modification at said target site, wherein the modification is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

[0137] The length of 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, or more 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.

[0138] Active variants of genomic target sites can also be used. Such active variants can comprise at least 50%, 55%, 60%, 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 a Cas endonuclease. Assays to measure the 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 containing recognition sites.

[0139] 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 to provide integration of the polynucleotide of Interest at the target site. In one method provided, a polynucleotide of interest is provided to the cell in 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. Optionally, the donor DNA construct can further comprise 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 plant genome. 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 plant genome. A region of homology can be of any length that is sufficient to promote homologous recombination at the cleaved target site. For example, the region of homology can comprise at least 5-10, 5-15, 5-20, 5-25, 5-30, 5-35, 5-40, 5-45, 5-50, 5-55, 5-60, 5-65, 5-70, 5-75, 5-80, 5-85, 5-90, 5-95, 5-100, 5-200, 5-300, 5-400, 5-500, 5-600, 5-700, 5-800, 5-900, 5-1000, 5-1100, 5-1200, 5-1300, 5-1400, 5-1500, 5-1600, 5-1700, 5-1800, 5-1900, 5-2000, 5-2100, 5-2200, 5-2300, 5-2400, 5-2500, 5-2600, 5-2700, 5-2800. 5-2900, 5-3000, 5-3100 or more bases in length such that the region of homology has sufficient homology to undergo homologous recombination with the corresponding genomic region. "Sufficient homology" indicates that two polynucleotide sequences have sufficient structural similarity to act as substrates for a homologous recombination reaction. The structural similarity includes overall length of each polynucleotide fragment, as well as the sequence similarity of the polynucleotides. Sequence similarity can be described by the percent sequence identity over the whole length of the sequences, and/or by conserved regions comprising localized similarities such as contiguous nucleotides having 100% sequence identity, and percent sequence identity over a portion of the length of the sequences.

[0140] 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 bp. The amount of homology can also 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).

[0141] As used herein, a "genomic region" is a segment of a chromosome in the genome of a plant 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.

[0142] The region of homology on the donor DNA can have homology to any sequence flanking the target site. While in some embodiments 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. In still other embodiments, the regions of homology can also have homology with a fragment of the target site along with downstream genomic regions. 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.

[0143] As used herein, "homologous recombination" refers to 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.

[0144] Alteration of the genome of a plant cell, for example, through homologous recombination (HR), is a powerful tool for genetic engineering. Despite the low frequency of homologous recombination in higher plants, there are a few examples of successful homologous recombination of plant endogenous genes. The parameters for homologous recombination in plants have primarily been investigated by rescuing introduced truncated selectable marker genes. In these experiments, the homologous DNA fragments were typically between 0.3 kb to 2 kb. Observed frequencies for homologous recombination were on the order of 10^{-4} to 10^{-5}. See, for example, Halfter et al., (1992) Mol Gen Genet. 231: 186-93; Offring a et al., (1990) EMBO J. 9:3077-84; Offring a et al., (1993) Proc. Natl. Acad. Sci. USA 90:7346-50; Paszkowski et al., (1988) EMBO J. 7:4021-6; Hourda and Paszkowski, (1994) Mol Gen Genet. 243:106-11; and Risseeuw et al., (1995) Plant J 7:109-19.

[0145] Homologous recombination has been demonstrated in insects. In Drosophila, Dray and Gloor found that as little as 3 kb of total template:target homology sufficed to copy a large non-homologous segment of DNA into the target with reasonable efficiency (Dray and Gloor, (1997) Genetics 147: 689-99). Using FLP-mediated DNA integration at a target FRT in Drosophila, Golic et al., showed integration was approximately 10-fold more efficient when the donor and target shared 4.1 kb of homology as compared to 1.1 kb of homology (Golic et al., (1997) Nucleic Acids Res 25:3665). Data from *Drosophila* indicates that 2-4 kb of homology is sufficient for efficient targeting, but there is some evidence that much less homology may suffice, on the order of about 30 bp to about 100 bp (Nassif and Engels, (1993) Proc. Natl. Acad. Sci. USA 90:1262-6; Keeler and Gloor, (1997) Mol Cell Biol 17:627-34).

[0146] 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. Embryos bearing inserted transgenic ES cells develop as genetically offspring. By interbreeding siblings, homozygous mice carrying the selected genes can be obtained. An overview of the process is provided in Watson et al., (1992) Recombinant DNA, 2nd Ed., (Scientific American Books distributed by WH Freeman & Co.); Capecchi, (1989) Trends Genet. 5:70-6; and Bronson, (1994) J Biol Chem 269:27155-8. Homologous recombination in mammals other than mouse has been limited by the lack of stem cells capable of being transplanted to oocytes or developing embryos. However, McCreath et al., Nature 405:1066-9 (2000) reported successful homologous recombination in sheep by transformation and selection in primary embryo fibroblast cells.

[0147] Once a double-strand break is induced in the DNA, the cell's DNA repair mechanism is activated to repair the break. Error-prone DNA repair mechanisms can produce mutations at double-strand break sites. The most common repair mechanism to bring the broken ends together is the nonhomologous end-joining (NHEJ) pathway (Bleuyard et al., (2006) DNA Repair 5:1-12). The structural integrity of chromosomes is typically preserved by the repair, but deletions, insertions, or other rearrangements are possible (Siebert and Puchta, (2002) Plant Cell 14:1121-31; Pacher et al., (2007) Genetics 175:21-9). The two ends of one double-strand break are the most prevalent substrates of NHEJ (Kirik et al., (2000) EMBO J. 19:5562-6), however if two different double-strand breaks occur, the free ends from different breaks can be ligated and result in chromosomal deletions (Siebert and Puchta, (2002) Plant Cell 14:1121-31), or chromosomal translocations between different chromosomes (Pacher et al., (2007) Genetics 175:21-9).

[0148] 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).

[0149] Alternatively, the double-strand break can be repaired by homologous recombination between homologous DNA sequences. Once the sequence around the double-strand break is altered, for example, by exonuclease activities involved in the maturation of double-strand breaks, gene conversion pathways can restore the original structure if a homologous sequence is available, such as a homologous chromosome in non-dividing somatic cells, or a sister chromatid after DNA replication (Molinier et al., (2004) Plant Cell 16:342-52). Ectopic and/or epigenic DNA sequences may also serve as a DNA repair template for homologous recombination (Puchta, (1999) Genetics 152:1173-81).

[0150] DNA double-strand breaks appear to 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).

[0151] In some embodiments, the methods provided herein comprise contacting a cell with a donor DNA and a Cas endonuclease. Once a double-strand break is introduced in the target site by the Cas endonuclease, the first and second regions of homology of the donor DNA can undergo homologous recombination with their corresponding genomic regions of homology resulting in exchange of DNA between the donor and the genome.

[0152] As such, the provided methods result in the integration of the polynucleotide of interest of the donor DNA into the double-strand break in the target site in the genome of a cell or organism, thereby altering the original target site and producing an altered genomic target site.

[0153] In one embodiment of the disclosure, the method comprises a method for introducing a polynucleotide of interest into a target site in the genome of a cell, the method comprising: a) introducing a guide polynucleotide, a donor DNA and a Cas endonuclease into a cell, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site; b) contacting the cell of (a) with a donor DNA comprising a polynucleotide of interest; and, c) identifying at least one cell from (b) comprising in its genome the polynucleotide of interest integrated at said target. The guide polynucleotide, Cas endonuclease and donor DNA can be introduced by any means known in the art. These means include, but are not limited to direct delivery of each component via particle bombardment, delivery through one or more recombinant DNA expression cassettes, or any combination thereof.

[0154] In some embodiment of the disclosure, the method comprises a method for introducing a polynucleotide of interest into a target site in the genome of a cell, wherein the donor DNA and Cas endonuclease are introduced into said cell using at least one recombinant DNA construct capable of expressing the donor DNA and/or the Cas endonuclease; and/or, wherein the guide polynucleotide is introduced directly by particle bombardment.

[0155] In another embodiment of the disclosure, the method comprises method for introducing a polynucleotide of interest into a target site in the genome of a cell, the method comprising: a) introducing into a cell a first recombinant DNA construct capable of expressing a guide polynucleotide, and a second recombinant DNA construct capable of expressing a Cas endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site; b) contacting the cell of (a) with a donor DNA comprising a polynucleotide of interest; and, c) identifying at least one cell from (b) comprising in its genome the polynucleotide of interest integrated at said target site.

[0156] The donor DNA may be introduced by any means known in the art. For example, a cell or organism, such as but not limiting to a plant or yeast having a target site is provided. The donor DNA may be provided by any transformation method known in the art including, for example, *Agrobacterium*-mediated transformation or biolistic particle bombardment. The donor DNA may be present transiently in the cell or it could be introduced via a viral replicon. In the presence of the Cas endonuclease and the target site, the donor DNA is inserted into the transformed genome.

[0157] Another approach uses protein engineering of existing homing endonucleases to alter their target specificities. Homing endonucleases, such as I-SceI or I-CreI, bind to and cleave relatively long DNA recognition sequences (18 bp and 22 bp, respectively). These sequences are predicted to naturally occur infrequently in a genome, typically only 1 or 2 sites/genome. The cleavage specificity of a homing endonuclease can be changed by rational design of amino acid substitutions at the DNA binding domain and/or combinatorial assembly and selection of mutated monomers (see, for example, Arnould et al., (2006) J Mol Biol 355:443-58; Ashworth et al., (2006) Nature 441:656-9; Doyon et al., (2006) J Am Chem Soc 128:2477-84; Rosen et al., (2006) Nucleic Acids Res 34:4791-800; and Smith et al., (2006) Nucleic Acids Res 34:e149; Lyznik et al., (2009) U.S. Patent Application Publication No. 20090133152A1; Smith et al., (2007) U.S. Patent Application Publication No. 20070117128A1). Engineered meganucleases have been demonstrated that can cleave cognate mutant sites without broadening their specificity. An artificial recognition site specific to the wild type yeast I-SceI homing nuclease was introduced in maize genome and mutations of the recognition sequence were detected in 1% of analyzed F1 plants when a transgenic I-SceI was introduced by crossing and activated by gene excision (Yang et al., (2009) Plant Mol Biol 70:669-79). More practically, the maize liguleless locus was targeted using an engineered single-chain endonuclease designed based on the I-CreI meganuclease sequence. Mutations of the selected liguleless locus recognition sequence were detected in 3% of the TO transgenic plants when the designed homing nuclease was introduced by *Agrobacterium*-mediated transformation of immature embryos (Gao et al., (2010) Plant J 61:176-87).

[0158] Polynucleotides of interest are further described herein and are 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 genetic engineering will change accordingly.

Genome Editing Using the Guide Polynucleotide/Cas Endonuclease System.

[0159] As described herein, the guide polynucleotide/Cas endonuclease system can be used in combination with a co-delivered polynucleotide modification template to allow for editing of a genomic nucleotide sequence of interest. While numerous double-strand break-making systems exist, their practical applications for gene editing may be restricted due to the relatively low frequency of induced double-strand breaks (DSBs). To date, many genome modification methods rely on the homologous recombination system. Homologous recombination (HR) can provide molecular means for finding genomic DNA sequences of interest and modifying them according to the experimental specifications. Homologous recombination takes place in plant somatic cells at low frequency. The process can be enhanced to a practical level for genome engineering by introducing double-strand breaks (DSBs) at selected endonuclease target sites. The challenge has been to efficiently make DSBs at genomic sites of interest since there is a bias in the directionality of information transfer between two interacting DNA molecules (the broken one acts as an acceptor of genetic information). Described herein is the use of a guide polynucleotide/Cas system which provides flexible genome cleavage specificity and results in a high frequency of double-strand breaks at a DNA target site, thereby enabling efficient gene editing in a nucleotide sequence of interest, wherein the nucleotide sequence of interest to be edited can be located within or outside the target site recognized and cleaved by a Cas endonuclease.

[0160] 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).

[0161] The term “polynucleotide modification template” refers to 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.

[0162] In one embodiment, the disclosure describes a method for editing a nucleotide sequence in the genome of a cell, the method comprising introducing a guide polynucleotide, a polynucleotide modification template and at least one Cas endonuclease into a cell, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein the Cas endonuclease introduces a double-strand break at a target site in the genome of said cell, wherein said polynucleotide modification template comprises at least one nucleotide modification of said nucleotide sequence. Cells include, but are not limited to, animal, bacterial, fungal, insect, yeast, and plant cells as well as plants and seeds produced by the methods described herein. 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.

[0163] The nucleotide sequence to be edited can be a sequence that is endogenous, artificial, pre-existing, or transgenic to the cell that is being edited. For example, the nucleotide sequence in the genome of a cell can be a transgene that is stably incorporated into the genome of a cell. Editing of such transgene may result in a further desired phenotype or genotype. The nucleotide sequence in the genome of a cell can also be a mutated or pre-existing sequence that was either endogenous or artificial from origin such as an endogenous gene or a mutated gene of interest.

[0164] In one embodiment the nucleotide sequence can be a promoter wherein the editing of the promoter results in any one of the following or any one combination of the following: an increased promoter activity, an increased promoter tissue specificity, a decreased promoter activity, a decreased promoter tissue specificity, a mutation of DNA binding elements and/or a deletion or addition of DNA binding elements.

[0165] In one embodiment the nucleotide sequence can be a regulatory sequence in the genome of a cell. A regulatory sequence is a segment of a nucleic acid molecule which is capable of increasing or decreasing the expression of specific genes within an organism. Examples of regulatory sequences include, but are not limited to, transcription activators, transcriptions repressors, and translational repressors, splicing factors, miRNAs, siRNA, artificial miRNAs, a CAAT box, a CCAAT box, a Pribnow box, a TATA box, SECIS elements and polyadenylation signals. In some embodiments the editing of a regulatory element results in altered protein translation, RNA cleavage, RNA splicing or transcriptional termination.

Regulatory Sequence Modifications Using the Guide Polynucleotide/Cas Endonuclease System

[0166] In one embodiment the nucleotide sequence to be modified can be a regulatory sequence such as a promoter wherein the editing of the promoter comprises replacing the promoter (also referred to as a "promoter swap" or "promoter replacement") or promoter fragment with a different promoter (also referred to as replacement promoter) or promoter fragment (also referred to as replacement promoter fragment), wherein the promoter replacement results in any one of the following or any one combination of the following: an increased promoter activity, an increased promoter tissue specificity, a decreased promoter activity, a decreased promoter tissue specificity, a new promoter activity, an inducible promoter activity, an extended window of gene expression, a modification of the timing or developmental progress of gene expression in the same cell layer or other cell layer (such as but not limiting to extending the timing of gene expression in the tapetum of maize anthers (U.S. Pat. No. 5,837,850 issued Nov. 17, 1998), a mutation of DNA binding elements and/or a deletion or addition of DNA binding elements. The promoter (or promoter fragment) to be modified can be a promoter (or promoter fragment) that is endogenous, artificial, pre-existing, or transgenic to the cell that is being edited. The replacement promoter (or replacement promoter fragment) can be a promoter (or promoter fragment) that is endogenous, artificial, pre-existing, or transgenic to the cell that is being edited.

[0167] In one embodiment the nucleotide sequence can be a promoter wherein the editing of the promoter comprises replacing an ARGOS 8 promoter with a *Zea mays* GOS2 PRO:GOS2-intron promoter.

[0168] In one embodiment the nucleotide sequence can be a promoter wherein the editing of the promoter comprises replacing a native EPSPS1 promoter from with a plant ubiquitin promoter.

[0169] In one embodiment the nucleotide sequence can be a promoter wherein the editing of the promoter comprises replacing an endogenous maize NPK1 promoter with a stress inducible maize RAB17 promoter.

[0170] In one embodiment the nucleotide sequence can be a promoter wherein the promoter to be edited is selected from the group comprising *Zea mays*-PEPC1 promoter (Kausch et al, Plant Molecular Biology, 45: 1-15, 2001), *Zea mays* Ubiquitin promoter (UBI1ZM PRO, Christensen et al, plant Molecular Biology 18: 675-689, 1992), *Zea mays*-Rootmet2 promoter (U.S. Pat. No. 7,214,855), Rice actin promoter (OS-ACTIN PRO, U.S. Pat. No. 5,641,876; McElroy et al, The Plant Cell, Vol 2, 163-171, February 1990), Sorghum RCC3 promoter (US 2012/0210463 filed on 13 Feb. 2012), *Zea mays*-GOS2 promoter (U.S. Pat. No. 6,504,083), *Zea mays*-ACO2 promoter (U.S. application Ser. No. 14/210,711 filed 14 Mar. 2014) or *Zea mays*-oleosin promoter (U.S. Pat. No. 8,466,341 B2).

[0171] In another embodiment, the guide polynucleotide/Cas endonuclease system can be used in combination with a co-delivered polynucleotide modification template or donor DNA sequence to allow for the insertion of a promoter or promoter element into a genomic nucleotide sequence of interest, wherein the promoter insertion (or promoter element insertion) results in any one of the following or any one combination of the following: an increased promoter activity (increased promoter strength), an increased promoter tissue specificity, a decreased promoter activity, a decreased promoter tissue specificity, a new promoter activity, an inducible promoter activity, an extended window of gene expression, a modification of the timing or developmental progress of gene expression a mutation of DNA binding elements and/or an addition of DNA binding elements. Promoter elements to be inserted can be, but are not limited to, promoter core elements (such as, but not limited to, a CAAT box, a CCAAT box, a Pribnow box, a and/or TATA box, translational regulation sequences and/or a repressor system for inducible expression (such as TET operator repressor/operator/inducer elements, or Sulphonylurea (Su) repressor/operator/inducer elements. The dehydration-responsive element (DRE) was first identified as a cis-acting promoter element in the promoter of the drought-responsive gene rd29A, which contains a 9 bp conserved core sequence, TACCGACAT (Yamaguchi-Shinozaki, K., and Shinozaki, K. (1994) *Plant Cell* 6, 251-264). Insertion of DRE into an endogenous promoter may confer a drought inducible expression of the downstream gene. Another example is ABA-responsive elements (ABREs) that contain a (C/T)ACGTGGC consensus sequence found to be present in numerous ABA and/or stress-regulated genes (Busk P. K., Pages M. (1998) Plant Mol. Biol. 37:425-435). Insertion of 35S enhancer or MMV enhancer into an endogenous promoter region will increase gene expression (U.S. Pat. No. 5,196,525). The promoter (or promoter element) to be inserted can be a promoter (or promoter element) that is endogenous, artificial, pre-existing, or transgenic to the cell that is being edited.

[0172] In one embodiment, the guide polynucleotide/Cas endonuclease system can be used to insert an enhancer element, such as but not limited to a Cauliflower Mosaic Virus 35 S enhancer, in front of an endogenous FMT1 promoter to enhance expression of the FTM1.

[0173] In one embodiment, the guide polynucleotide/Cas endonuclease system can be used to insert a component of the TET operator repressor/operator/inducer system, or a component of the sulphonylurea (Su) repressor/operator/inducer system into plant genomes to generate or control inducible expression systems.

[0174] In another embodiment, the guide polynucleotide/Cas endonuclease system can be used to allow for the deletion of a promoter or promoter element, wherein the promoter deletion (or promoter element deletion) results in any one of the following or any one combination of the following: a permanently inactivated gene locus, an increased promoter activity (increased promoter strength), an increased promoter tissue specificity, a decreased promoter activity, a decreased promoter tissue specificity, a new promoter activity, an inducible promoter activity, an extended window of gene expression, a modification of the timing or developmental progress of gene expression, a mutation of DNA binding elements and/or an addition of DNA binding elements. Promoter elements to be deleted can be, but are not limited to, promoter core elements, promoter enhancer elements or 35 S enhancer elements (as described in Example 32) The promoter or promoter fragment to be deleted can be endogenous, artificial, pre-existing, or transgenic to the cell that is being edited.

[0175] In one embodiment, the guide polynucleotide/Cas endonuclease system can be used to delete the ARGOS 8 promoter present in a maize genome as described herein.

[0176] In one embodiment, the guide polynucleotide/Cas endonuclease system can be used to delete a 35S enhancer element present in a plant genome as described herein.

Terminator Modifications Using the Guide Polynucleotide/Cas Endonuclease System

[0177] In one embodiment the nucleotide sequence to be modified can be a terminator wherein the editing of the terminator comprises replacing the terminator (also referred to as a "terminator swap" or "terminator replacement") or terminator fragment with a different terminator (also referred to as replacement terminator) or terminator fragment (also referred to as replacement terminator fragment), wherein the terminator replacement results in any one of the following or any one combination of the following: an increased terminator activity, an increased terminator tissue specificity, a decreased terminator activity, a decreased terminator tissue specificity, a mutation of DNA binding elements and/or a deletion or addition of DNA binding elements." The terminator (or terminator fragment) to be modified can be a terminator (or terminator fragment) that is endogenous, artificial, pre-existing, or transgenic to the cell that is being edited. The replacement terminator (or replacement terminator fragment) can be a terminator (or terminator fragment) that is endogenous, artificial, pre-existing, or transgenic to the cell that is being edited.

[0178] In one embodiment the nucleotide sequence to be modified can be a terminator wherein the terminator to be edited is selected from the group comprising terminators from maize Argos 8 or SRTF18 genes, or other terminators, such as potato PinII terminator, sorghum actin terminator (SB-ACTIN TERM, WO 2013/184537 A1 published December 2013), sorghum SB-GKAF TERM (WO2013019461), rice T28 terminator (OS-T28 TERM, WO 2013/012729 A2), AT-T9 TERM (WO 2013/012729 A2) or GZ-W64A TERM (U.S. Pat. No. 7,053,282).

[0179] In one embodiment, the guide polynucleotide/Cas endonuclease system can be used in combination with a co-delivered polynucleotide modification template or donor DNA sequence to allow for the insertion of a terminator or terminator element into a genomic nucleotide sequence of interest, wherein the terminator insertion (or terminator element insertion) results in any one of the following or any one combination of the following: an increased terminator activity (increased terminator strength), an increased terminator tissue specificity, a decreased terminator activity, a decreased terminator tissue specificity, a mutation of DNA binding elements and/or an addition of DNA binding elements. The terminator (or terminator element) to be inserted can be a terminator (or terminator element) that is endogenous, artificial, pre-existing, or transgenic to the cell that is being edited.

[0180] In another embodiment, the guide polynucleotide/Cas endonuclease system can be used to allow for the deletion of a terminator or terminator element, wherein the terminator deletion (or terminator element deletion) results in any one of the following or any one combination of the following: an increased terminator activity (increased terminator strength), an increased terminator tissue specificity, a decreased terminator activity, a decreased terminator tissue specificity, a mutation of DNA binding elements and/or an addition of DNA binding elements. The terminator or terminator fragment to be deleted can be endogenous, artificial, pre-existing, or transgenic to the cell that is being edited.

Additional Regulatory Sequence Modifications Using the Guide Polynucleotide/Cas Endonuclease System

[0181] In one embodiment, the guide polynucleotide/Cas endonuclease system can be used to modify or replace a regulatory sequence in the genome of a cell. A regulatory sequence is a segment of a nucleic acid molecule which is capable of increasing or decreasing the expression of specific genes within an organism and/or is capable of altering tissue specific expression of genes within an organism. Examples of regulatory sequences include, but are not limited to, 3' UTR (untranslated region) region, 5' UTR region, transcription activators, transcriptional enhancers transcriptions repressors, translational repressors, splicing factors, miRNAs, siRNA, artificial miRNAs, promoter elements, CAMV 35 S enhancer, MMV enhancer elements (PCT/US14/23451 filed Mar. 11, 2013), SECIS elements, polyadenylation signals, and polyubiquitination sites. In some embodiments the editing (modification) or replacement of a regulatory element results in altered protein translation, RNA cleavage, RNA splicing, transcriptional termination or post translational modification. In one embodiment, regulatory elements can be identified within a promoter and these regulatory elements can be edited or modified do to optimize these regulatory elements for up or down regulation of the promoter.

[0182] In one embodiment, the genomic sequence of interest to be modified is a polyubiquitination site, wherein the modification of the polyubiquitination sites results in a modified rate of protein degradation. The ubiquitin tag condemns proteins to be degraded by proteasomes or autophagy. Proteasome inhibitors are known to cause a protein overproduction. Modifications made to a DNA sequence encoding a protein of interest can result in at least one amino acid modification of the protein of interest, wherein said modification allows for the polyubiquitination of the protein (a post translational modification) resulting in a modification of the protein degradation

[0183] In one embodiment, the genomic sequence of interest to be modified is a polyubiquitination site on a maize EPSPS gene, wherein the polyubiquitination site modified resulting in an increased protein content due to a slower rate of EPSPS protein degradation.

[0184] In one embodiment, the genomic sequence of interest to be modified is a an intron site, wherein the modification consist of inserting an intron enhancing motif into the intron which results in modulation of the transcriptional activity of the gene comprising said intron.

[0185] In one embodiment, the genomic sequence of interest to be modified is a an intron site, wherein the modification consist of replacing a soybean EPSP1 intron with a soybean ubiquitin intron 1 as described herein (Example 25)

[0186] In one embodiment, the genomic sequence of interest to be modified is a an intron or UTR site, wherein the modification consist of inserting at least one microRNA into said intron or UTR site, wherein expression of the gene comprising the intron or UTR site also results in expression of said microRNA, which in turn can silence any gene targeted by the microRNA without disrupting the gene expression of the native/transgene comprising said intron.

[0187] In one embodiment, the guide polynucleotide/Cas endonuclease system can be used to allow for the deletion or mutation of a Zinc Finger transcription factor, wherein the deletion or mutation of the Zinc Finger transcription factor results in or allows for the creation of a dominant negative Zinc Finger transcription factor mutant (Li et al 2013 Rice zinc finger protein DST enhances grain production through controlling Gn1a/OsCKX2 expression PNAS 110:3167-3172). Insertion of a single base pair downstream zinc finger domain will result in a frame shift and produces a new protein which still can bind to DNA without transcription activity. The mutant protein will compete to bind to cytokinin oxidase gene promoters and block the expression of cytokinin oxidase gene. Reduction of cytokinin oxidase gene expression will increase cytokinin level and promote panicle growth in rice and ear growth in maize, and increase yield under normal and stress conditions.

Modifications of Splicing Sites and/or Introducing Alternate Splicing Sites Using the Guide Polynucleotide/Cas Endonuclease System

[0188] Protein synthesis utilizes mRNA molecules that emerge from pre-mRNA molecules subjected to the maturation process. The pre-mRNA molecules are capped, spliced and stabilized by addition of polyA tails. Eukaryotic cells developed a complex process of splicing that result in alternative variants of the original pre-mRNA molecules. Some of them may not produce functional templates for protein synthesis. In maize cells, the splicing process is affected by splicing sites at the exon-intron junction sites. An example of a canonical splice site is AGGT. Gene coding sequences can contains a number of alternate splicing sites that may affect the overall efficiency of the pre-mRNA maturation process and as such may limit the protein accumulation in cells. The guide polynucleotide/Cas endonuclease system can be used in combination with a co-delivered polynucleotide modification template to edit a gene of interest to introduce a canonical splice site at a described junction or any variant of a splicing site that changes the splicing pattern of pre-mRNA molecules.

[0189] In one embodiment, the nucleotide sequence of interest to be modified is a maize EPSPS gene, wherein the modification of the gene consists of modifying alternative splicing sites resulting in enhanced production of the functional gene transcripts and gene products (proteins).

[0190] In one embodiment, the nucleotide sequence of interest to be modified is a gene, wherein the modification of the gene consists of editing the intron borders of alternatively spliced genes to alter the accumulation of splice variants.

Modifications of Nucleotide Sequences Encoding a Protein of Interest Using the Guide Polynucleotide/Cas Endonuclease System

[0191] In one embodiment, the guide polynucleotide/Cas endonuclease system can be used to modify or replace a coding sequence in the genome of a cell, wherein the modification or replacement results in any one of the following, or any one combination of the following: an increased protein (enzyme) activity, an increased protein functionality, a decreased protein activity, a decreased protein functionality, a site specific mutation, a protein domain swap, a protein knock-out, a new protein functionality, a modified protein functionality.

[0192] In one embodiment the protein knockout is due to the introduction of a stop codon into the coding sequence of interest.

[0193] In one embodiment the protein knockout is due to the deletion of a start codon into the coding sequence of interest.

Amino Acid and/or Protein Fusions Using the Guide Polynucleotide/Cas Endonuclease System

[0194] In one embodiment, the guide polynucleotide/Cas endonuclease system can be used with or without a co-delivered polynucleotide sequence to fuse a first coding sequence encoding a first protein to a second coding sequence encoding a second protein in the genome of a cell, wherein the protein fusion results in any one of the following or any one combination of the following: an increased protein (enzyme) activity, an increased protein functionality, a decreased protein activity, a decreased protein functionality, a new protein functionality, a modified protein functionality, a new protein localization, a new timing of protein expression, a modified protein expression pattern, a chimeric protein, or a modified protein with dominant phenotype functionality.

[0195] In one embodiment, the guide polynucleotide/Cas endonuclease system can be used with or without a co-delivered polynucleotide sequence to fuse a first coding sequence encoding a chloroplast localization signal to a second coding sequence encoding a protein of interest, wherein the protein fusion results in targeting the protein of interest to the chloroplast.

[0196] In one embodiment, the guide polynucleotide/Cas endonuclease system can be used with or without a co-delivered polynucleotide sequence to fuse a first coding sequence encoding a chloroplast localization signal to a second coding sequence encoding a protein of interest, wherein the protein fusion results in targeting the protein of interest to the chloroplast.

[0197] In one embodiment, the guide polynucleotide/Cas endonuclease system can be used with or without a co-delivered polynucleotide sequence to fuse a first coding sequence encoding a chloroplast localization signal (e.g., a chloroplast transit peptide) to a second coding sequence, wherein the protein fusion results in a modified protein with dominant phenotype functionality

Gene Silencing by Expressing an Inverted Repeat into a Gene of Interest Using the Guide Polynucleotide/Cas Endonuclease System

[0198] In one embodiment, the guide polynucleotide/Cas endonuclease system can be used in combination with a co-delivered polynucleotide sequence to insert an inverted gene fragment into a gene of interest in the genome of an organism, wherein the insertion of the inverted gene fragment can allow for an in-vivo creation of an inverted repeat (hairpin) and results in the silencing of said endogenous gene.

[0199] In one embodiment the insertion of the inverted gene fragment can result in the formation of an in-vivo created inverted repeat (hairpin) in a native (or modified) promoter of a gene and/or in a native 5' end of the native gene. The inverted gene fragment can further comprise an intron which can result in an enhanced silencing of the targeted gene.

Genome deletion for Trait Locus Characterization

[0200] Trait mapping in plant breeding often results in the detection of chromosomal regions housing one or more genes controlling expression of a trait of interest. For a qualitative trait, the guide polynucleotide/Cas endonuclease system can be used to eliminate candidate genes in the identified chromosomal regions to determine if deletion of the gene affects expression of the trait. For quantitative traits, expression of a trait of interest is governed by multiple quantitative trait loci (QTL) of varying effect-size, complexity, and statistical significance across one or more chromosomes. In cases of negative effect or deleterious QTL regions affecting a complex trait, the guide polynucleotide/Cas endonuclease system can be used to eliminate whole regions delimited by marker-assisted fine mapping, and to target specific regions for their selective elimination or rearrangement. Similarly, presence/absence variation (PAV) or copy number variation (CNV) can be manipulated with selective genome deletion using the guide polynucleotide/Cas endonuclease system.

In one embodiment, the region of interest can be flanked by two independent guide polynucleotide/CAS endonuclease target sequences. Cutting would be done concurrently. The deletion event would be the repair of the two chromosomal ends without the region of interest. Alternative results would include inversions of the region of interest, mutations at the cut sites and duplication of the region of interest

[0201] Methods for Identifying at Least One Plant Cell Comprising in its Genome a Polynucleotide of Interest Integrated at the Target Site.

[0202] 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 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. See, for example, U.S. patent application Ser. No. 12/147,834, herein incorporated to the extent necessary for the methods described herein.

[0203] 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.

[0204] Polynucleotides/polypeptides of interest include, but are not limited to, herbicide-tolerance coding sequences, insecticidal coding sequences, nematicidal coding sequences, antimicrobial coding sequences, antifungal coding sequences, antiviral coding sequences, abiotic and biotic stress tolerance coding sequences, or sequences modifying plant traits such as yield, grain quality, nutrient content, starch quality and quantity, nitrogen fixation and/or utilization, fatty acids, and oil content and/or composition. More specific polynucleotides of interest include, but are not limited to, genes that improve crop yield, polypeptides that improve desirability of crops, 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, or to biotic stress, such as attacks by fungi, viruses, bacteria, insects, and nematodes, and development of diseases associated with these organisms. General categories of genes of interest include, for example, those genes 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 categories of transgenes, for example, include genes encoding important traits for agronomics, insect resistance, disease resistance, herbicide resistance, fertility or sterility, grain characteristics, and commercial products. Genes of interest include, generally, those involved in oil, starch, carbohydrate, or nutrient metabolism as well as those affecting kernel size, sucrose loading, and the like that can be stacked or used in combination with other traits, such as but not limited to herbicide resistance, described herein.

[0205] 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. Pat. Nos. 5,703,049, 5,885,801, 5,885,802, and 5,990,389, herein incorporated by reference. Another example is lysine and/or sulfur rich seed protein encoded by the soybean 2S albumin described in U.S. Pat. No. 5,850,016, and the chymotrypsin inhibitor from barley, described in Williamson et al. (1987) Eur. J. Biochem. 165:99-106, the disclosures of which are herein incorporated by reference.

[0206] Commercial traits can also be encoded on a polynucleotide of interest that could increase for example, starch for ethanol production, or provide expression of proteins. Another important commercial use of transformed plants is the production of polymers and bioplastics such as described in U.S. Pat. No. 5,602,321. Genes such as β-Ketothiolase, PHBase (polyhydroxybutyrate synthase), and acetoacetyl-CoA reductase (see Schubert et al. (1988) J. Bacteriol. 170: 5837-5847) facilitate expression of polyhydroxyalkanoates (PHAs).

[0207] Derivatives of the coding sequences can be made by site-directed mutagenesis to increase the level of preselected amino acids in the encoded polypeptide. For example, the gene encoding the barley high lysine polypeptide (BHL) is derived from barley chymotrypsin inhibitor, U.S. application Ser. No. 08/740,682, filed Nov. 1, 1996, and WO 98/20133, the disclosures of which are herein incorporated by reference. Other proteins include methionine-rich plant proteins such as from sunflower seed (Lilley et al. (1989) *Proceedings of the World Congress on Vegetable Protein Utilization in Human Foods and Animal Feedstuffs*, ed. Applewhite (American Oil Chemists Society, Champaign, Ill.), pp. 497-502; herein incorporated by reference); corn (Pedersen et al. (1986) *J. Biol. Chem.* 261:6279; Kirihara et al. (1988) *Gene* 71:359; both of which are herein incorporated by reference); and rice (Musumura et al. (1989) *Plant Mol. Biol.* 12:123, herein incorporated by reference). Other agronomically important genes encode latex, Floury 2, growth factors, seed storage factors, and transcription factors.

[0208] Polynucleotides that improve crop yield include dwarfing genes, such as Rht1 and Rht2 (Peng et al. (1999) Nature 400:256-261), and those that increase plant growth, such as ammonium-inducible glutamate dehydrogenase. Polynucleotides that improve desirability of crops include, for example, those that allow plants to have reduced saturated fat content, those that boost the nutritional value of plants, and those that increase grain protein. Polynucleotides that improve salt tolerance are those that increase or allow plant growth in an environment of higher salinity than the native environment of the plant into which the salt-tolerant gene(s) has been introduced.

[0209] Polynucleotides/polypeptides that influence amino acid biosynthesis include, for example, anthranilate synthase (AS; EC 4.1.3.27) which catalyzes the first reaction branching from the aromatic amino acid pathway to the biosynthesis of tryptophan in plants, fungi, and bacteria. In plants, the chemical processes for the biosynthesis of tryptophan are compartmentalized in the chloroplast. See, for example, US Pub. 20080050506, herein incorporated by reference. Additional sequences of interest include Chorismate Pyruvate Lyase (CPL) which refers to a gene encoding an enzyme which catalyzes the conversion of chorismate to pyruvate and pHBA. The most well characterized CPL gene has been isolated from *E. coli* and bears the GenBank accession number M96268. See, U.S. Pat. No. 7,361,811, herein incorporated by reference.

[0210] These 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. Pat. 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. Pat. 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.

[0211] 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), in particular the sulfonylurea-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, U.S. Pat. Nos. 7,626,077, 5,310,667, 5,866,775, 6,225,114, 6,248,876, 7,169,970, 6,867,293, and U.S. Provisional Application No. 61/401,456, each of which is herein incorporated by reference. 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.

[0212] Sterility genes can also be encoded in an expression cassette and provide an alternative to physical detasseling. Examples of genes used in such ways include male fertility genes such as MS26 (see for example U.S. Pat. Nos. 7,098,388, 7,517,975, 7,612,251), MS45 (see for example U.S. Pat. Nos. 5,478,369, 6,265,640) or MSCA1 (see for example U.S. Pat. No. 7,919,676). Maize plants (*Zea mays* L.) can be bred by both self-pollination and cross-pollination techniques. Maize has male flowers, located on the tassel, and female flowers, located on the ear, on the same plant. It can self-pollinate ("selfing") or cross pollinate. Natural pollination occurs in maize when wind blows pollen from the tassels to the silks that protrude from the tops of the incipient ears. Pollination may be readily controlled by techniques known to those of skill in the art. The development of maize hybrids requires the development of homozygous inbred lines, the crossing of these lines, and the evaluation of the crosses. Pedigree breeding and recurrent selection are two of the breeding methods used to develop inbred lines from populations. Breeding programs combine desirable traits from two or more inbred lines or various broad-based sources into breeding pools from which new inbred lines are developed by selfing and selection of desired phenotypes. A hybrid maize variety is the cross of two such inbred lines, each of which may have one or more desirable characteristics lacked by the other or which complement the other. The new inbreds are crossed with other inbred lines and the hybrids from these crosses are evaluated to determine which have commercial potential. The hybrid progeny of the first generation is designated F1. The F1 hybrid is more vigorous than its inbred parents. This hybrid vigor, or heterosis, can be manifested in many ways, including increased vegetative growth and increased yield.

[0213] Hybrid maize seed can be produced by a male sterility system incorporating manual detasseling. To produce hybrid seed, the male tassel is removed from the growing female inbred parent, which can be planted in various alternating row patterns with the male inbred parent. Consequently, providing that there is sufficient isolation from sources of foreign maize pollen, the ears of the female inbred will be fertilized only with pollen from the male inbred. The resulting seed is therefore hybrid (F1) and will form hybrid plants.

[0214] Field variation impacting plant development can result in plants tasseling after manual detasseling of the female parent is completed. Or, a female inbred plant tassel may not be completely removed during the detasseling process. In any event, the result is that the female plant will successfully shed pollen and some female plants will be self-pollinated. This will result in seed of the female inbred being harvested along with the hybrid seed which is normally produced. Female inbred seed does not exhibit heterosis and therefore is not as productive as F1 seed. In addition, the presence of female inbred seed can represent a germplasm security risk for the company producing the hybrid.

[0215] Alternatively, the female inbred can be mechanically detasseled by machine. Mechanical detasseling is approximately as reliable as hand detasseling, but is faster and less costly. However, most detasseling machines produce more damage to the plants than hand detasseling. Thus, no form of detasseling is presently entirely satisfactory, and a need continues to exist for alternatives which further reduce production costs and to eliminate self-pollination of the female parent in the production of hybrid seed.

[0216] Mutations that cause male sterility in plants have the potential to be useful in methods for hybrid seed production for crop plants such as maize and can lower production costs by eliminating the need for the labor-intensive removal of male flowers (also known as de-tasseling) from the maternal parent plants used as a hybrid parent. Mutations that cause male sterility in maize have been produced by a variety of methods such as X-rays or UV-irradiations, chemical treatments, or transposable element insertions (ms23, ms25, ms26, ms32) (Chaubal et al. (2000) Am J Bot 87:1193-1201). Conditional regulation of fertility genes through fertility/sterility "molecular switches" could enhance the options for designing new male-sterility systems for crop improvement (Unger et al. (2002) Transgenic Res 11:455-465).

[0217] Besides identification of novel genes impacting male fertility, there remains a need to provide a reliable system of producing genetic male sterility.

[0218] In U.S. Pat. No. 5,478,369, a method is described by which the Ms45 male fertility gene was tagged and cloned on maize chromosome 9. Previously, there had been described a male fertility gene on chromosome 9, ms2, which had never been cloned and sequenced. It is not allelic to the gene referred to in the '369 patent. See Albertsen, M. and Phillips, R. L., "Developmental Cytology of 13 Genetic Male Sterile Loci in Maize" Canadian Journal of Genetics & Cytology 23:195-208 (January 1981). The only fertility gene cloned before that had been the *Arabidopsis* gene described at Aarts, et al., supra.

[0219] Examples of genes that have been discovered subsequently that are important to male fertility are numerous and include the *Arabidopsis* ABORTED MICROSPORES (AMS) gene, Sorensen et al., The Plant Journal (2003) 33(2): 413-423); the *Arabidopsis* MS1 gene (Wilson et al., The Plant Journal (2001) 39(2):170-181); the NEF1 gene (Ariizumi et al., The Plant Journal (2004) 39(2):170-181); *Arabidopsis* AtGPAT1 gene (Zheng et al., The Plant Cell (2003) 15:1872-1887); the *Arabidopsis* dde2-2 mutation was shown to be defective in the allene oxide syntase gene (Malek et al., Planta (2002) 216:187-192); the *Arabidopsis* faceless pollen-1 gene (flp1) (Ariizumi et al, Plant Mol. Biol. (2003) 53:107-116); the *Arabidopsis* MALE MEIOCYTE DEATH1 gene (Yang et al., The Plant Cell (2003) 15: 1281-1295); the tapetum-specific zinc finger gene, TAZ1 (Kapoor et al., The Plant Cell (2002) 14:2353-2367); and the TAPETUM DETERMINANT1 gene (Lan et al, The Plant Cell (2003) 15:2792-2804).

[0220] Other known male fertility mutants or genes from *Zea mays* are listed in U.S. Pat. No. 7,919,676 incorporated herein by reference.

[0221] Other genes include kinases and those encoding compounds toxic to either male or female gametophytic development.

[0222] 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.

[0223] 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. Pat. Nos. 5,283,184 and 5,034,323; herein incorporated by reference.

[0224] 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 contains 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.

[0225] 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 (3-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.

[0226] Additional selectable markers include genes that confer resistance to herbicidal compounds, such as glufosinate ammonium, bromoxynil, imidazolinones, and 2,4-dichlorophenoxyacetate (2,4-D). See for example, Yarranton, (1992) Curr Opin Biotech 3:506-11; Christopherson et al., (1992) Proc. Natl. Acad. Sci. USA 89:6314-8; Yao et al., (1992) Cell 71:63-72; Reznikoff, (1992) Mol Microbiol 6:2419-22; Hu et al., (1987) Cell 48:555-66; Brown et al., (1987) Cell 49:603-12; Figge et al., (1988) Cell 52:713-22; Deuschle et al., (1989) Proc. Natl. Acad. Sci. USA 86:5400-4; Fuerst et al., (1989) Proc. Natl. Acad. Sci. USA 86:2549-53; Deuschle et al., (1990) Science 248:480-3; Gossen, (1993) Ph.D. Thesis, University of Heidelberg; Reines et al., (1993) Proc. Natl. Acad. Sci. USA 90:1917-21; Labow et al., (1990) Mol Cell Biol 10:3343-56; Zambretti et al., (1992) Proc. Natl. Acad. Sci. USA 89:3952-6; Baim et al., (1991) Proc. Natl. Acad. Sci. USA 88:5072-6; Wyborski et al., (1991) Nucleic Acids Res 19:4647-53; Hillen and Wissman, (1989) Topics Mol Struc Biol 10:143-62; Degenkolb et al., (1991) Antimicrob Agents Chemother 35:1591-5; Kleinschnidt et al., (1988) Biochemistry 27:1094-104; Bonin, (1993) Ph.D. Thesis, University of Heidelberg; Gossen et al., (1992) Proc. Natl. Acad. Sci. USA 89:5547-51; Oliva et al., (1992) Antimicrob Agents Chemother 36:913-9; Hlavka et al., (1985) Handbook of Experimental Pharmacology, Vol. 78 (Springer-Verlag, Berlin); Gill et al., (1988) Nature 334:721-4. Commercial traits can also be encoded on a gene or genes that could increase for example, starch for ethanol production, or provide expression of proteins. Another important commercial use of transformed plants is the production of polymers and bioplastics such as described in U.S. Pat. No. 5,602,321. Genes such as β-Ketothiolase, PHBase (polyhydroxyburyrate synthase), and acetoacetyl-CoA reductase (see Schubert et al. (1988) *J. Bacteriol.* 170:5837-5847) facilitate expression of polyhyroxyalkanoates (PHAs).

[0227] Exogenous products include plant enzymes and products as well as those from other sources including prokaryotes and other eukaryotes. Such products include enzymes, cofactors, hormones, and the like. The level of proteins, particularly modified proteins having improved amino acid distribution to improve the nutrient value of the plant, can be increased. This is achieved by the expression of such proteins having enhanced amino acid content.

[0228] The transgenes, recombinant DNA molecules, DNA sequences of interest, and polynucleotides of interest can be comprise one or more DNA sequences for gene silencing. Methods for gene silencing involving the expression of DNA sequences in plant are known in the art include, but are not limited to, cosuppression, antisense suppression, double-stranded RNA (dsRNA) interference, hairpin RNA (hpRNA) interference, intron-containing hairpin RNA (ihpRNA) interference, transcriptional gene silencing, and micro RNA (miRNA) interference

[0229] 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 that is single- or double-stranded, optionally containing 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.

[0230] "Open reading frame" is abbreviated ORF.

[0231] The terms "subfragment that is functionally equivalent" and "functionally equivalent subfragment" are used interchangeably herein. These terms refer to a portion or subsequence of an isolated nucleic acid fragment in which the ability to alter gene expression or produce a certain phenotype is retained whether or not the fragment or subfragment encodes an active enzyme. For example, the fragment or subfragment can be used in the design of genes to produce the desired phenotype in a transformed plant. Genes can be designed for use in suppression by linking a nucleic acid fragment or subfragment thereof, whether or not it encodes an active enzyme, in the sense or antisense orientation relative to a plant promoter sequence.

[0232] The term "conserved domain" or "motif" means a set of 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.

[0233] 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 fragments 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 fragments that do not substantially alter the functional properties of the resulting nucleic acid fragment relative to the initial, unmodified fragment. These modifications include deletion, substitution, and/or insertion of one or more nucleotides in the nucleic acid fragment.

[0234] Substantially similar nucleic acid sequences encompassed may be defined by their ability to hybridize (under moderately stringent conditions, e.g., 0.5×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.

[0235] 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.

[0236] 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.

[0237] 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° A) SDS (sodium dodecyl sulphate) at 37° C., and a wash in 1× to 2×SSC (20×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 0.5× to 1×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 0.1×SSC at 60 to 65° C.

[0238] "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.

[0239] 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 integer percentage from 50% to 100%. These identities can be determined using any of the programs described herein.

[0240] 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, Wis.). 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.

[0241] 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, Wis.). 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.

[0242] 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, Wis.). 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.

[0243] Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using GAP Version 10 (GCG, Accelrys, San Diego, Calif.) 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.

[0244] "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.

[0245] 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 integer percentage from 50% to 100%. Indeed, any integer 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%.

[0246] "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 nature with its own regulatory sequences.

[0247] 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.

[0248] As used herein, a "targeted mutation" is a mutation in a native gene that was made by altering a target sequence within the native gene using a method involving a double-strand-break-inducing agent that is capable of inducing a double-strand break in the DNA of the target sequence as disclosed herein or known in the art.

[0249] In one embodiment, the targeted mutation is the result of a guide polynucleotide/Cas endonuclease induced gene editing as described herein. The 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 a Cas endonuclease.

[0250] The term "genome" as it applies to a plant 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.

[0251] 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.

[0252] 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.

[0253] "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 may 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.

[0254] "A plant-optimized nucleotide sequence" is nucleotide sequence that has been optimized for increased expression in plants, particularly for increased expression in plants or in one or more plants of interest. For example, a plant-optimized nucleotide sequence can be synthesized by modifying a nucleotide sequence encoding a protein such as, for example, double-strand-break-inducing agent (e.g., an 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.

[0255] Methods are available in the art for synthesizing plant-preferred genes. See, for example, U.S. Pat. Nos. 5,380,831, and 5,436,391, and Murray et al. (1989) *Nucleic Acids Res.* 17:477-498, herein incorporated by reference. Additional sequence modifications are known to enhance gene expression in a plant host. These include, for example, elimination of: one or more sequences encoding spurious polyadenylation signals, one or more exon-intron splice site signals, one or more transposon-like repeats, and other such well-characterized sequences that may be deleterious to gene expression. The G-C content of the sequence may be adjusted to levels average for a given plant host, as calculated by reference to known genes expressed in the host plant cell. When possible, the sequence is modified to avoid one or more predicted hairpin secondary mRNA structures. Thus, "a plant-optimized nucleotide sequence" of the present disclosure comprises one or more of such sequence modifications.

[0256] "Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. 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. Promoters that cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters".

[0257] 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. Since patterns of expression of a chimeric gene (or genes) introduced into a plant are controlled using promoters, there is an ongoing interest in the isolation of novel promoters which are capable of controlling the expression of a chimeric gene or (genes) at certain levels in specific tissue types or at specific plant developmental stages.

[0258] 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, N.Y.: Academic Press), pp. 1-82.

[0259] "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 fully processed 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).

[0260] "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.

[0261] "RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complimentary copy of the DNA sequence, it is referred to as the primary transcript. A RNA transcript is referred to as the mature RNA when it is a RNA sequence derived from post-transcriptional processing of the primary transcript. "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, a 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. Pat. 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.

[0262] 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.

[0263] 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.

[0264] "PCR" or "polymerase chain reaction" is a technique for the synthesis of specific DNA segments and consists of a series of repetitive denaturation, annealing, and extension cycles. Typically, a double-stranded DNA is heat denatured, and two primers complementary to the 3' boundaries of the target segment are annealed to the DNA at low temperature, and then extended at an intermediate temperature. One set of these three consecutive steps is referred to as a "cycle".

[0265] 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.

[0266] The terms "plasmid", "vector" and "cassette" refer to an 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 containing a foreign gene and having elements in addition to the foreign gene that facilitates transformation of a particular host cell. "Expression cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that allow for expression of that gene in a foreign host.

[0267] The terms "recombinant DNA molecule", "recombinant construct", "expression construct", "construct", "construct", and "recombinant DNA construct" are used interchangeably herein. A recombinant construct comprises an artificial combination of nucleic acid fragments, e.g., regulatory and coding sequences that are not all found together in nature. For example, a 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 transform 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.

[0268] The term "expression", as used herein, refers to the production of a functional end-product (e.g., an mRNA, guide polynucleotide, or a protein) in either precursor or mature form.

[0269] The term "introduced" means providing a nucleic acid (e.g., expression construct) or protein into a cell. Introduced 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 provision of a nucleic acid or protein to the cell. Introduced includes reference to stable or transient transformation methods, as well as sexually crossing. Thus, "introduced" in the context of inserting a nucleic acid fragment (e.g., a recombinant DNA construct/expression construct) into a cell, means "transfection" or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid fragment into a eukaryotic or prokaryotic cell where the nucleic acid fragment may be incorporated into the genome of the cell (e.g., chromosome, plasmid, plastid, or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (e.g., transfected mRNA).

[0270] "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). "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.

[0271] "Stable transformation" refers to the transfer of a nucleic acid fragment into a genome of a host organism, including both nuclear and organellar genomes, resulting in genetically stable inheritance. In contrast, "transient transformation" refers to the transfer of a nucleic acid fragment into the nucleus, or other DNA-containing organelle, of a host organism resulting in gene expression without integration or stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms.

[0272] The commercial development of genetically improved germplasm has also advanced to the stage of introducing multiple traits into crop plants, often referred to as a gene stacking approach. In this approach, multiple genes conferring different characteristics of interest can be introduced into a plant. Gene stacking can be accomplished by many means including but not limited to co-transformation, retransformation, and crossing lines with different genes of interest.

[0273] The term "plant" refers to whole plants, plant organs, plant tissues, seeds, plant cells, seeds and progeny of the same. Plant cells include, without limitation, cells from seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen and microspores. Plant parts include differentiated and undifferentiated tissues including, but not limited to roots, stems, shoots, leaves, pollens, seeds, tumor tissue and various forms of cells and culture (e.g., single cells, protoplasts, embryos, and callus tissue). The plant tissue may be in plant or in a plant organ, tissue or cell culture. The term "plant organ" refers to plant tissue or a group of tissues that constitute a morphologically and functionally distinct part of a plant. The term "genome" refers to the entire complement of genetic material (genes and non-coding sequences) that is present in each cell of an organism, or virus or organelle; and/or a complete set of chromosomes inherited as a (haploid) unit from one parent. "Progeny" comprises any subsequent generation of a plant.

[0274] A transgenic plant includes, for example, a plant which comprises within its genome a heterologous polynucleotide introduced by a transformation step. The heterologous polynucleotide can be stably integrated within the genome such that the polynucleotide is passed on to successive generations. The heterologous polynucleotide may be integrated into the genome alone or as part of a recombinant DNA construct. A transgenic plant can also comprise more than one heterologous polynucleotide within its genome. Each heterologous polynucleotide may confer a different trait to the transgenic plant. A heterologous polynucleotide can include a sequence that originates from a foreign species, or, if from the same species, can be substantially modified from its native form. Transgenic can include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic. The alterations of the genome (chromosomal or extra-chromosomal) by conventional plant breeding methods, by the genome editing procedure described herein that does not result in an insertion of a foreign polynucleotide, or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation are not intended to be regarded as transgenic.

[0275] In one embodiment of the disclosure, the composition comprises a plant or seed comprising a recombinant DNA construct and a guide polynucleotide, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said recombinant DNA construct comprises a promoter operably linked to a nucleotide sequence encoding a plant optimized Cas endonuclease, wherein said plant optimized Cas endonuclease and guide polynucleotide are capable of forming a complex and creating a double strand break in a genomic target site of said plant.

[0276] In another embodiment of the disclosure, the composition further comprising a polynucleotide of interest integrated into a genomic target site of said plant.

[0277] In another embodiment of the disclosure, the composition further comprising a modification at a genomic target site, wherein the modification is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

[0278] In another embodiment of the disclosure, the composition comprises a plant or seed comprising at least one altered target sequence, wherein the at least one altered target sequence originated from a corresponding target sequence that was recognized and cleaved by a guide polynucleotide/Cas endonuclease complex, wherein the Cas endonuclease is capable of introducing a double-strand break at said target site in the plant genome, wherein said guide polynucleotide does not solely comprise ribonucleic acids.

[0279] In another embodiment of the disclosure, the composition comprises a plant or seed comprising a modified nucleotide sequence, wherein the modified nucleotide sequence was produced by providing a guide polynucleotide, a polynucleotide modification template and at least one Cas endonuclease to a cell, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein the Cas endonuclease is capable of introducing a double-strand break at a target site in the plant genome, wherein said polynucleotide modification template comprises at least one nucleotide modification of said nucleotide sequence.

[0280] 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 contained 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. 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.

[0281] A “centimorgan” (cM) or “map unit” is the distance between two 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.

[0282] Breeding Methods and Methods for Selecting Plants Utilizing a Two Component RNA Guide and Cas Endonuclease System.

[0283] The present disclosure finds use in the breeding of plants comprising one or more transgenic traits. Most commonly, transgenic traits are randomly inserted throughout the plant genome as a consequence of transformation systems based on *Agrobacterium*, biolistics, or other commonly used procedures. More recently, gene targeting protocols have been developed that enable directed transgene insertion. One important technology, site-specific integration (SSI) enables the targeting of a transgene to the same chromosomal location as a previously inserted transgene. Custom-designed meganucleases and custom-designed zinc finger meganucleases allow researchers to design nucleases to target specific chromosomal locations, and these reagents allow the targeting of transgenes at the chromosomal site cleaved by these nucleases.

[0284] The currently used systems for precision genetic engineering of eukaryotic genomes, e.g. plant genomes, rely upon homing endonucleases, meganucleases, zinc finger nucleases, and transcription activator—like effector nucleases (TALENs), which require de novo protein engineering for every new target locus.

[0285] The highly specific, guide polynucleotide/Cas9 endonuclease system described herein, is more easily customizable and therefore more useful when modification of many different target sequences is the goal. In one embodiment, the disclosure takes further advantage of the multiple component nature of the guide polynucleotide/Cas system, with its constant protein component, the Cas endonuclease, and its variable and easily reprogrammable targeting component, the guide polynucleotide. As described herein, the guide polynucleotide can comprise a DNA, RNA or DNA-RNA combination sequence making it very customizable and therefore more useful for when modification of one or many different target sequences is the goal.

[0286] The guide polynucleotide/Cas system described herein is especially useful for genome engineering, especially plant genome engineering, in circumstances where nuclease off-target cutting can be toxic to the targeted cells. In one embodiment of the guide polynucleotide/Cas system described herein, the constant component, in the form of an expression-optimized Cas9 gene, is stably integrated into the target genome, e.g. plant genome. Expression of the Cas9 gene is under control of a promoter, e.g. plant promoter, which can be a constitutive promoter, tissue-specific promoter or inducible promoter, e.g. temperature-inducible, stress-inducible, developmental stage inducible, or chemically inducible promoter. In the absence of the variable targeting domain, of the guide polynucleotide, the Cas protein is not able to recognize and cut DNA and therefore its presence in the plant cell should have little or no consequence. Hence a key advantage of the guide polynucleotide/Cas system described herein is the ability to create and maintain a cell line or transgenic organism capable of efficient expression of the Cas protein with little or no consequence to cell viability. In order to induce cutting at desired genomic sites to achieve targeted genetic modifications, guide polynucleotides can be introduced by a variety of methods into cells containing the stably-integrated and expressed Cas gene. For example, guide polynucleotides can be chemically or enzymatically synthesized, and introduced into the Cas expressing cells via direct delivery methods such a particle bombardment or electroporation.

[0287] Alternatively, genes capable of efficiently expressing guide polynucleotides in the target cells can be synthesized chemically, enzymatically or in a biological system, and these genes can be introduced into the Cas expressing cells via direct delivery methods such a particle bombardment, electroporation or biological delivery methods such as *Agrobacterium* mediated DNA delivery.

[0288] One embodiment of the disclosure is a method for selecting a plant comprising an altered target site in its plant genome, the method comprising: a) obtaining a first plant comprising at least one Cas endonuclease capable of introducing a double strand break at a target site in the plant genome; b) obtaining a second plant comprising a guide polynucleotide that is capable of forming a complex with the Cas endonuclease of (a), wherein the guide polynucleotide does not solely comprise ribonucleic acids, c) crossing the first plant of (a) with the second plant of (b); d) evaluating the progeny of (c) for an alteration in the target site and e) selecting a progeny plant that possesses the desired alteration of said target site.

[0289] Another embodiment of the disclosure is a method for selecting a plant comprising an altered target site in its plant genome, the method comprising: a) obtaining a first plant comprising at least one Cas endonuclease capable of introducing a double strand break at a target site in the plant genome; b) obtaining a second plant comprising a guide polynucleotide and a donor DNA, wherein the guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide is capable of forming a complex with the Cas endonuclease of (a), wherein said donor DNA comprises a polynucleotide of interest; c) crossing the first plant of (a) with the second plant of (b); d) evaluating the progeny of (c) for an alteration in the target site and e) selecting a progeny plant that comprises the polynucleotide of interest inserted at said target site.

[0290] Another embodiment of the disclosure is a method for selecting a plant comprising an altered target site in its plant genome, the method comprising selecting at least one progeny plant that comprises an alteration at a target site in its plant genome, wherein said progeny plant was obtained by crossing a first plant expressing at least one Cas endonuclease to a second plant comprising a guide polynucleotide and a donor DNA, wherein the guide polynucleotide does not solely comprise ribonucleic acids, wherein said Cas endonuclease is capable of introducing a double strand break at said target site, wherein said donor DNA comprises a polynucleotide of interest.

[0291] As disclosed herein, a guide polynucleotide/Cas system mediating gene targeting can be used in methods for directing transgene insertion and/or for producing complex transgenic trait loci comprising multiple transgenes in a fashion similar as disclosed in WO2013/0198888 (published Aug. 1, 2013) 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. In one embodiment, a complex transgenic trait locus is a genomic locus that has multiple transgenes genetically linked to each other. By inserting independent transgenes within 0.1, 0.2, 0.3, 04, 0.5, 1, 2, or even 5 centimorgans (cM) from each other, the transgenes can be bred as a single genetic locus (see, for example, U.S. patent application Ser. No. 13/427,138) or PCT application PCT/US2012/030061. After selecting a plant comprising a transgene, plants containing (at least) one transgenes can be crossed to form an F1 that contains 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.

[0292] 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. 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 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 containing target sites.

[0293] A variety of methods are known for the introduction of nucleotide sequences and polypeptides into an organism, including, for example, transformation, sexual crossing, and the introduction of the polypeptide, DNA, or mRNA into the cell.

[0294] Methods for contacting, providing, and/or introducing a composition into various organisms are known and include but are not limited to, stable transformation methods, transient transformation methods, virus-mediated methods, and sexual breeding. Stable transformation indicates that the introduced polynucleotide integrates into the genome of the organism and is capable of being inherited by progeny thereof. Transient transformation indicates that the introduced composition is only temporarily expressed or present in the organism.

[0295] Protocols for introducing polynucleotides and polypeptides into plants may vary depending on the type of plant or plant cell targeted for transformation, such as monocot or dicot. Suitable methods of introducing polynucleotides and polypeptides into plant cells and subsequent insertion into the plant genome include microinjection (Crossway et al., (1986) *Biotechniques* 4:320-34 and U.S. Pat. No. 6,300, 543), meristem transformation (U.S. Pat. No. 5,736,369), electroporation (Riggs et al., (1986) *Proc. Natl. Acad. Sci. USA* 83:5602-6, *Agrobacterium*-mediated transformation (U.S. Pat. Nos. 5,563,055 and 5,981,840), direct gene transfer (Paszkowski et al., (1984) *EMBO J.* 3:2717-22), and ballistic particle acceleration (U.S. Pat. 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. Pat. 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. Pat. 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*).

[0296] Alternatively, polynucleotides may be introduced into plants by contacting plants 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. Pat. Nos. 5,889,191, 5,889,190, 5,866,785, 5,589,367 and 5,316,931. Transient transformation methods include, but are not limited to, the introduction of polypeptides, such as a double-strand break inducing agent, directly into the organism, the introduction of polynucleotides such as DNA and/or RNA polynucleotides, and the introduction of the RNA transcript, such as an mRNA encoding a double-strand break inducing agent, into the organism. Such methods include, for example, microinjection or particle bombardment. See, for example Crossway et al., (1986) *Mol Gen Genet.* 202:179-85; Nomura et al., (1986) *Plant Sci* 44:53-8; Hepler et al., (1994) *Proc. Natl. Acad. Sci. USA* 91:2176-80; and, Hush et al., (1994) *J Cell Sci* 107:775-84.

[0297] The term "dicot" refers to the subclass of angiosperm plants also knows as "dicotyledoneae" and includes reference to whole plants, plant organs (e.g., leaves, stems, roots, etc.), seeds, plant cells, and progeny of the same. Plant cell, as used herein includes, without limitation, seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores.

[0298] 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 are from the same plant or genetically identical plants).

[0299] 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 or a selected allele of a marker or QTL.

[0300] Standard DNA isolation, purification, molecular cloning, vector construction, and verification/characterization methods are well established, see, for example Sambrook et al., (1989) *Molecular Cloning: A Laboratory Manual*, (Cold Spring Harbor Laboratory Press, NY). Vectors and constructs include circular plasmids, and linear polynucleotides, comprising a polynucleotide of interest and optionally other components including linkers, adapters, regulatory regions, introns, restriction sites, enhancers, insulators, selectable markers, nucleotide sequences of interest, promoters, and/or other sites that aid in vector construction or analysis. In some examples a recognition site and/or target site can be contained within an intron, coding sequence, 5' UTRs, 3' UTRs, and/or regulatory regions.

[0301] The present disclosure further provides expression constructs for expressing in a yeast or plant, plant cell, or plant part a guide polynucleotide/Cas system that is capable of binding to and creating a double strand break in a target site. In one embodiment, the expression constructs of the disclosure comprise a promoter operably linked to a nucleotide sequence encoding a Cas gene and a promoter operably linked to a guide polynucleotide of the present disclosure. The promoter is capable of driving expression of an operably linked nucleotide sequence in a plant cell.

[0302] A promoter is a region of DNA involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. A plant promoter is 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. Constitutive promoters include, for example, the core promoter of the Rsyn7 promoter and other constitutive promoters disclosed in WO99/43838 and U.S. Pat. No. 6,072,050; 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; Christensen et al., (1992) *Plant Mol Biol* 18:675-89); pEMU (Last et al., (1991) *Theor Appl Genet.* 81:581-8); MAS (Velten et al., (1984) *EMBO J.* 3:2723-30); ALS promoter (U.S. Pat. No. 5,659,026), and the like. Other constitutive promoters are described in, for example, U.S. Pat. Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; 5,608,142 and 6,177,611. In some examples an inducible promoter may be used. 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.

[0303] Chemical-regulated promoters can be used to modulate the expression of a gene in a plant through the application of an exogenous chemical regulator. The promoter may be a chemical-inducible promoter, where application of the chemical induces gene expression, or a chemical-repressible promoter, where application of the chemical represses gene expression. Chemical-inducible promoters include, but are not limited to, the maize In2-2 promoter, activated by benzene sulfonamide herbicide safeners (De Veylder et al., (1997) *Plant Cell Physiol* 38:568-77), the maize GST promoter (GST-II-27, WO93/01294), activated by hydrophobic electrophilic compounds used as pre-emergent herbicides, and the tobacco PR-1a promoter (Ono et al., (2004) *Biosci Biotechnol Biochem* 68:803-7) activated by salicylic acid. Other chemical-regulated promoters include steroid-responsive promoters (see, for example, the glucocorticoid-inducible promoter (Schena et al., (1991) *Proc. Natl. Acad. Sci. USA* 88:10421-5; McNellis et al., (1998) *Plant J* 14:247-257); tetracycline-inducible and tetracycline-repressible promoters (Gatz et al., (1991) *Mol Gen Genet.* 227:229-37; U.S. Pat. Nos. 5,814,618 and 5,789,156).

[0304] Tissue-preferred promoters can be utilized to target enhanced expression within a particular plant tissue. Tissue-preferred promoters include, for example, Kawamata et al., (1997) *Plant Cell Physiol* 38:792-803; Hansen et al., (1997) *Mol Gen Genet.* 254:337-43; Russell et al., (1997) *Transgenic Res* 6:157-68; Rinehart et al., (1996) *Plant Physiol* 112:1331-41; Van Camp et al., (1996) *Plant Physiol* 112:525-35; Canevascini et al., (1996) *Plant Physiol* 112:513-524; Lam, (1994) *Results Probl Cell Differ* 20:181-96; and Guevara-Garcia et al., (1993) *Plant J* 4:495-505. Leaf-preferred promoters include, for example, Yamamoto et al., (1997) *Plant J* 12:255-65; Kwon et al., (1994) *Plant Physiol* 105:357-67; Yamamoto et al., (1994) *Plant Cell Physiol* 35:773-8; Gotor et al., (1993) *Plant J* 3:509-18; Orozco et al., (1993) *Plant Mol Biol* 23:1129-38; Matsuoka et al., (1993) *Proc. Natl. Acad. Sci. USA* 90:9586-90; Simpson et al., (1958) *EMBO J.* 4:2723-9; Timko et al., (1988) *Nature* 318:57-8. Root-preferred promoters include, for example, Hire et al., (1992) *Plant Mol Biol* 20:207-18 (soybean root-specific glutamine synthase gene); Miao et al., (1991) *Plant Cell* 3:11-22 (cytosolic glutamine synthase (GS)); Keller and Baumgartner, (1991) *Plant Cell* 3:1051-61 (root-specific control element in the GRP 1.8 gene of French bean); Sanger et al., (1990) *Plant Mol Biol* 14:433-43 (root-specific promoter of *A. tumefaciens* mannopine synthase (MAS)); Bogusz et al., (1990) *Plant Cell* 2:633-41 (root-specific promoters isolated from *Parasponia andersonii* and *Trema tomentosa*); Leach and Aoyagi, (1991) *Plant Sci* 79:69-76 (*A. rhizogenes* rolC and rolD root-inducing genes); Teeri et al., (1989) *EMBO J.* 8:343-50 (Agrobacterium wound-induced TR1' and TR2' genes); VfENOD-GRP3 gene promoter (Kuster et al., (1995) *Plant Mol Biol* 29:759-72); and rolB promoter (Capana et al., (1994) *Plant Mol Biol* 25:681-91; phaseolin gene (Murai et al., (1983) *Science* 23:476-82; Sengopta-Gopalen et al., (1988) *Proc. Natl. Acad. Sci. USA* 82:3320-4). See also, U.S. Pat. Nos. 5,837,876; 5,750,386; 5,633,363; 5,459,252; 5,401,836; 5,110,732 and 5,023,179.

[0305] Seed-preferred promoters include both seed-specific promoters active during seed development, as well as seed-germinating promoters active during seed germination. See, Thompson et al., (1989) *BioEssays* 10:108. Seed-preferred promoters include, but are not limited to, Cim1 (cytokinin-induced message); cZ19B1 (maize 19 kDa zein); and milps (myo-inositol-1-phosphate synthase); (WO00/11177; and U.S. Pat. No. 6,225,529). For dicots, seed-preferred promoters include, but are not limited to, bean β-phaseolin, napin, β-conglycinin, soybean lectin, cruciferin, and the like. For monocots, seed-preferred promoters include, but are not limited to, maize 15 kDa zein, 22 kDa zein, 27 kDa gamma zein, waxy, shrunken 1, shrunken 2, globulin 1, oleosin, and nuc1. See also, WO00/12733, where seed-preferred promoters from END1 and END2 genes are disclosed.

[0306] A phenotypic marker is a screenable or 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 contains 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.

[0307] 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.

[0308] Additional selectable markers include genes that confer resistance to herbicidal compounds, such as glufosinate ammonium, bromoxynil, imidazolinones, and 2,4-dichlorophenoxyacetate (2,4-D). See for example, Yarranton, (1992) *Curr Opin Biotech* 3:506-11; Christopherson et al., (1992) *Proc. Natl. Acad. Sci. USA* 89:6314-8; Yao et al., (1992) *Cell* 71:63-72; Reznikoff, (1992) *Mol Microbiol* 6:2419-22; Hu et al., (1987) *Cell* 48:555-66; Brown et al., (1987) *Cell* 49:603-12; Figge et al., (1988) *Cell* 52:713-22; Deuschle et al., (1989) *Proc. Natl. Acad. Sci. USA* 86:5400-4; Fuerst et al., (1989) *Proc. Natl. Acad. Sci. USA* 86:2549-53; Deuschle et al., (1990) *Science* 248:480-3; Gossen, (1993) *Ph.D. Thesis*, University of Heidelberg; Reines et al., (1993) *Proc. Natl. Acad. Sci. USA* 90:1917-21; Labow et al., (1990) *Mol Cell Biol* 10:3343-56; Zambretti et al., (1992) *Proc. Natl. Acad. Sci. USA* 89:3952-6; Baim et al., (1991) *Proc. Natl. Acad. Sci. USA* 88:5072-6; Wyborski et al., (1991) *Nucleic Acids Res* 19:4647-53; Hillen and Wissman, (1989) *Topics Mol Struc Biol* 10:143-62; Degenkolb et al., (1991) *Antimicrob Agents Chemother* 35:1591-5; Kleinschnidt et al., (1988) *Biochemistry* 27:1094-104; Bonin, (1993) Ph.D. Thesis, University of Heidelberg; Gossen et al., (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-51; Oliva et al., (1992) *Antimicrob Agents Chemother* 36:913-9; Hlavka et al., (1985) *Handbook of Experimental Pharmacology*, Vol. 78 (Springer-Verlag, Berlin); Gill et al., (1988) *Nature* 334:721-4.

[0309] The cells having the introduced sequence may be grown or regenerated into plants using conventional conditions, see for example, McCormick et al., (1986) *Plant Cell Rep* 5:81-4. These plants may then be grown, and either pollinated with the same transformed strain or with a different transformed or untransformed strain, and the resulting progeny having the desired characteristic and/or comprising the introduced polynucleotide or polypeptide identified. Two or more generations may be grown to ensure that the polynucleotide is stably maintained and inherited, and seeds harvested.

[0310] Any plant can be used, including monocot and dicot plants. 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 aestivum*), sugarcane (*Saccharum* spp.), oats (Avena), barley (Hordeum), switchgrass (*Panicum virgatum*), pineapple (*Ananas comosus*), banana (*Musa* spp.), palm, ornamentals, turfgrasses, and other grasses. Examples of dicot plants that can be used include, but are not limited to, soybean (*Glycine max*), canola (*Brassica napus* and *B. campestris*), alfalfa (*Medicago sativa*), tobacco (*Nicotiana tabacum*), *Arabidopsis* (*Arabidopsis thaliana*), sunflower (*Helianthus annuus*), cotton (*Gossypium arboreum*), and peanut (*Arachis hypogaea*), tomato (*Solanum lycopersicum*), potato (*Solanum tuberosum*) etc.

[0311] The transgenes, recombinant DNA molecules, DNA sequences of interest, and polynucleotides of interest can comprise one or more genes of interest. Such genes of interest can encode, for example, a protein that provides agronomic advantage to the plant.

[0312] Marker Assisted Selection and Breeding of Plants

[0313] A primary motivation for development of molecular markers in crop species is the potential for increased efficiency in plant breeding through marker assisted selection (MAS). Genetic marker alleles, or alternatively, quantitative trait loci (QTL alleles, are used to identify plants that contain a desired genotype at one or more loci, and that are expected to transfer the desired genotype, along with a desired phenotype to their progeny. Genetic marker alleles (or QTL alleles) can be used to identify plants that contain a desired genotype at one locus, or at several unlinked or linked loci (e.g., a haplotype), and that would be expected to transfer the desired genotype, along with a desired phenotype to their progeny. It will be appreciated that for the purposes of MAS, the term marker can encompass both marker and QTL loci.

[0314] After a desired phenotype and a polymorphic chromosomal locus, e.g., a marker locus or QTL, are determined to segregate together, it is possible to use those polymorphic loci to select for alleles corresponding to the desired phenotype—a process called marker-assisted selection (MAS). In brief, a nucleic acid corresponding to the marker nucleic acid is detected in a biological sample from a plant to be selected.

This detection can take the form of hybridization of a probe nucleic acid to a marker, e.g., using allele-specific hybridization, southern blot analysis, northern blot analysis, in situ hybridization, hybridization of primers followed by PCR amplification of a region of the marker or the like. A variety of procedures for detecting markers are well known in the art. After the presence (or absence) of a particular marker in the biological sample is verified, the plant is selected, i.e., used to make progeny plants by selective breeding.

[0315] Plant breeders need to combine traits of interest with genes for high yield and other desirable traits to develop improved plant varieties. Screening for large numbers of samples can be expensive, time consuming, and unreliable. Use of markers, and/or genetically-linked nucleic acids is an effective method for selecting plant having the desired traits in breeding programs. For example, one advantage of marker-assisted selection over field evaluations is that MAS can be done at any time of year regardless of the growing season. Moreover, environmental effects are irrelevant to marker-assisted selection.

[0316] When a population is segregating for multiple loci affecting one or multiple traits, the efficiency of MAS compared to phenotypic screening becomes even greater because all the loci can be processed in the lab together from a single sample of DNA.

[0317] The DNA repair mechanisms of cells are the basis to introduce extraneous DNA or induce mutations on endogenous genes. DNA homologous recombination is a specialized way of DNA repair that the cells repair DNA damages using a homologous sequence. In plants, DNA homologous recombination happens at frequencies too low to be routinely used in gene targeting or gene editing until it has been found that the process can be stimulated by DNA double-strand breaks (Bibikova et al., (2001) Mol. Cell. Biol. 21:289-297; Puchta and Baltimore, (2003) Science 300:763; Wright et al., (2005) Plant J. 44:693-705).

[0318] 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" mean micromole(s), "g" means gram (s), "μg" means microgram(s), "ng" means nanogram(s), "U" means unit(s), "bp" means base pair(s) and "kb" means kilobase(s).

Non-limiting examples of compositions and methods disclosed herein are as follows:

[0319] 1. A guide polynucleotide comprising:

[0320] (i) a first nucleotide sequence domain that is complementary to a nucleotide sequence in a target DNA; and,

[0321] (ii) a second nucleotide sequence domain that interacts with a Cas endonuclease, wherein the first nucleotide sequence domain and the second nucleotide sequence domain are composed of deoxyribonucleic acids (DNA), ribonucleic acids (RNA), or a combination thereof, wherein the guide polynucleotide does not solely comprise ribonucleic acids.

[0322] 2. The guide polynucleotide of embodiment 1 wherein the first nucleotide sequence domain and the second nucleotide sequence domain are located on a single molecule.

[0323] 3. The guide polynucleotide of embodiment 1 wherein the second nucleotide sequence domain comprises two separate molecules that are capable of hybridizing along a region of complementarity.

[0324] 4. The guide polynucleotide of any one of embodiments 1-3, wherein the first nucleotide sequence domain is a DNA sequence and the second nucleotide sequence domain is selected from the group consisting of a DNA sequence, a RNA sequence, and a combination thereof.

[0325] 5. The guide polynucleotide of embodiment 1 wherein the first nucleotide sequence domain and the second nucleotide sequence domain are DNA sequences.

[0326] 6. The guide polynucleotide of embodiment 1, wherein the first nucleotide sequence domain and/or the second nucleotide sequence domain comprises at least one modification, wherein said at least one modification is selected from 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.

[0327] 7. The guide polynucleotide of embodiment 1, wherein the first nucleotide sequence domain and/or the second nucleotide sequence domain comprises at least one modification that provides for an additional beneficial feature, wherein said at least one modification is selected from 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.

[0328] 8. The guide polynucleotide of embodiment 7, 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.

[0329] 9. A plant or seed comprising the guide polynucleotide of any one of embodiments 1-8

[0330] 10. A guide polynucleotide/Cas endonuclease complex wherein the guide polynucleotide comprises (i) a first nucleotide sequence domain that is complementary to a nucleotide sequence in a target DNA; and (ii) a second nucleotide sequence domain that interacts with a Cas endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site.

[0331] 11. The guide polynucleotide/Cas endonuclease complex of embodiment 10, wherein the first nucleotide sequence domain and the second nucleotide sequence domain of the guide polynucleotide are composed of deoxyribonucleic acids (DNA), ribonucleic acids (RNA), or a combination thereof, wherein the guide polynucleotide does not solely comprise ribonucleic acids.

[0332] 12. The guide polynucleotide/Cas endonuclease complex of embodiment 10, wherein the first nucleotide sequence domain and/or the second nucleotide sequence domain of said guide polynucleotide comprises at least one modification that provides for an additional beneficial feature, wherein said at least one modification is selected from 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.

[0333] 13. The guide polynucleotide/Cas endonuclease complex of any one of embodiments 10-12, wherein the Cas endonuclease is a Cas9 endonuclease.

[0334] 14. A plant or seed comprising the guide polynucleotide/Cas endonuclease complex of any of one of embodiments 10-13.

[0335] 15. A method for modifying a target site in the genome of a cell, the method comprising introducing a guide polynucleotide into a cell having a Cas endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site.

[0336] 16. A method for modifying a target site in the genome of a cell, the method comprising introducing a guide polynucleotide and a Cas endonuclease into a cell, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site.

[0337] 17. The method of any one of embodiments 15-16, further comprising introducing a donor DNA to said cell, wherein said donor DNA comprises a polynucleotide of interest.

[0338] 18. The method of any one of embodiments 15-17, further comprising identifying at least one cell that has a modification at said target, wherein the modification at said target site is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

[0339] 19. A method for introducing a polynucleotide of interest into a target site in the genome of a cell, the method comprising:

[0340] a) introducing a guide polynucleotide, a donor DNA and a Cas endonuclease into a cell, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site;

[0341] b) contacting the cell of (a) with a donor DNA comprising a polynucleotide of interest; and,

[0342] c) identifying at least one cell from (b) comprising in its genome the polynucleotide of interest integrated at said target site.

[0343] 20. The method of embodiment 19, wherein the donor DNA and Cas endonuclease are introduced into said cell using at least one recombinant DNA construct capable of expressing the donor DNA and/or the Cas endonuclease.

[0344] 21. The method of any one of embodiments 15-20, wherein the guide polynucleotide is introduced directly by particle bombardment.

[0345] 22. The method of any one of embodiments 15-20, wherein the guide polynucleotide is introduced via particle bombardment or *Agrobacterium* transformation of a recombinant DNA construct comprising a U6 polymerase III

[0346] 23. The method of any one of embodiments 15-20, wherein the guide polynucleotide is a single guide polynucleotide comprising a variable targeting domain and a cas endonuclease recognition domain.

[0347] 24. The method of any one of embodiments 15-20, wherein the guide polynucleotide is a duplex guide polynucleotide comprising a crNucleotide molecule and a tracrNucleotide molecule.

[0348] 25. A method for modifying a target site in the genome of a cell, the method comprising:

[0349] a) introducing into a cell a crNucleotide, a first recombinant DNA construct capable of expressing a tracrRNA, and a second recombinant DNA capable of expressing a Cas endonuclease, wherein said crNucleotide is a deoxyribonucleotide sequence or a combination of a deoxyribonucleotide and ribonucleotide sequence, wherein said crNucleotide, said tracrRNA and said Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site; and,

[0350] b) identifying at least one cell that has a modification at said target site, wherein the modification is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

[0351] 26. A method for modifying a target site in the genome of a cell, the method comprising:

[0352] a) introducing into a cell a tracrNucleotide, a first recombinant DNA construct capable of expressing a crRNA and a second recombinant DNA capable of expressing a Cas endonuclease, wherein said tracrNucleotide is selected a deoxyribonucleotide sequence or a combination of a deoxyribonucleotide and ribonucleotide sequence, wherein said tracrNucleotide, said crRNA and said Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site; and,

[0353] b) identifying at least one cell that has a modification at said target site, wherein the modification is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

[0354] 27. A method for introducing a polynucleotide of interest into a target site in the genome of a cell, the method comprising:

[0355] a) introducing into a cell a first recombinant DNA construct capable of expressing a guide polynucleotide, and a second recombinant DNA construct capable of expressing a Cas endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site;

[0356] b) contacting the cell of (a) with a donor DNA comprising a polynucleotide of interest; and,

[0357] c) identifying at least one cell from (b) comprising in its genome the polynucleotide of interest integrated at said target site.

[0358] 28. A method for editing a nucleotide sequence in the genome of a cell, the method comprising introducing a guide polynucleotide, a polynucleotide modification template and at least one Cas endonuclease into a cell, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein the Cas endonuclease introduces a double-strand break at a target site in the genome of said cell, wherein said polynucleotide modification template comprises at least one nucleotide modification of said nucleotide sequence.

[0359] 29. The method of any one of embodiments 15-28, wherein the cell is selected from the group consisting of a non-human animal, bacterial, fungal, insect, yeast, and a plant cell.

[0360] 30. The method of embodiment 29, wherein the plant cell is selected from the group consisting of a monocot and dicot cell.

[0361] 31. The method of embodiment 29, wherein the plant cell is selected from the group consisting of maize, rice, sorghum, rye, barley, wheat, millet, oats, sugarcane, turfgrass, or switchgrass, soybean, canola, alfalfa, sunflower, cotton, tobacco, peanut, potato, tobacco, *Arabidopsis*, and safflower cell.

[0362] 32. A plant or seed comprising a guide polynucleotide and a Cas9 endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said Cas9 endonuclease and guide polynucleotide are capable of forming a complex and creating a double strand break in a genomic target site of said plant.

[0363] 33. A plant or seed comprising a recombinant DNA construct and a guide polynucleotide, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said recombinant DNA construct comprises a promoter operably linked to a nucleotide sequence encoding a plant optimized Cas endonuclease, wherein said plant optimized Cas endonuclease and guide polynucleotide are capable of forming a complex and creating a double strand break in a genomic target site of said plant.

[0364] 34. The plant of any one of embodiments 32-33, further comprising a polynucleotide of interest integrated into said genomic target site of said plant.

[0365] 35. The plant or seed of any one of embodiments 32-33 further comprising a modification at said genomic target site, wherein the modification is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

[0366] 36. A plant or seed comprising at least one altered target sequence, wherein the at least one altered target sequence originated from a corresponding target sequence that was recognized and cleaved by a guide polynucleotide/Cas endonuclease complex, wherein the Cas endonuclease is capable of introducing a double-strand break at said target site in the plant genome, wherein said guide polynucleotide does not solely comprise ribonucleic acids.

[0367] 37. A plant or seed comprising a modified nucleotide sequence, wherein the modified nucleotide sequence was produced by providing a guide polynucleotide, a polynucleotide modification template and at least one Cas endonuclease to a cell, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein the Cas endonuclease is capable of introducing a double-strand break at a target site in the plant genome, wherein said polynucleotide modification template comprises at least one nucleotide modification of said nucleotide sequence.

[0368] 38. The plant or plant cell of embodiment 29 wherein the at least one nucleotide modification is not a modification at said target site.

[0369] 39. The plant of any one of embodiments 32-38, wherein the plant is a monocot or a dicot.

[0370] 40. The plant of embodiment 39, wherein the monocot is selected from the group consisting of maize, rice, sorghum, rye, barley, wheat, millet, oats, sugarcane, turfgrass, or switchgrass.

[0371] 41. The plant of embodiment 39, wherein the dicot is selected from the group consisting of soybean, canola, alfalfa, sunflower, cotton, tobacco, peanut, potato, tobacco, *Arabidopsis*, or safflower.

[0372] 42. A method for selecting a plant comprising an altered target site in its plant genome, the method comprising: a) obtaining a first plant comprising at least one Cas endonuclease capable of introducing a double strand break at a target site in the plant genome; b) obtaining a second plant comprising a guide polynucleotide that is capable of forming a complex with the Cas endonuclease of (a), wherein the guide polynucleotide does not solely comprise ribonucleic acids, c) crossing the first plant of (a) with the second plant of (b); d) evaluating the progeny of (c) for an alteration in the target site and e) selecting a progeny plant that possesses the desired alteration of said target site.

[0373] A method for selecting a plant comprising an altered target site in its plant genome, the method comprising: a) obtaining a first plant comprising at least one Cas endonuclease capable of introducing a double strand break at a target site in the plant genome; b) obtaining a second plant comprising a guide polynucleotide and a donor DNA, wherein the guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide is capable of forming a complex with the Cas endonuclease of (a), wherein said donor DNA comprises a polynucleotide of interest; c) crossing the first plant of (a) with the second plant of (b); d) evaluating the progeny of (c) for an alteration in the target site and e) selecting a progeny plant that comprises the polynucleotide of interest inserted at said target site.

EXAMPLES

[0374] The present disclosure is further defined in the following Examples, in which parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating embodiments of the disclosure, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this disclosure, and without departing from the spirit and scope thereof, can make various changes and modifications of the disclosure to adapt it to various usages and conditions. Such modifications are also intended to fall within the scope of the appended embodiments.

Example 1

Maize Optimized Expression Cassettes for a Duplex Guide Polynucleotide/Cas Endonuclease System for Genome Modification in Maize Plants

[0375] In this example, expression cassettes for the Cas9 endonuclease and the mature fully processed naturally occurring CRISPR—RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA) belonging to the type II adaptive viral immune system from *S. pyogenes* as described in Deltcheva et al. (2011) *Nature* 471:602-7 and Jinek et al. (2012) *Science* 337:816-21 are maize optimized to examine their use as genome engineering tools in maize.

[0376] As shown in FIG. **1**A, the crNucleotide and tracrNucleotide molecules composed entirely of RNA nucleotides in this example, form a duplex comprised of a first nucleotide sequence domain referred to as the "variable targeting" (VT) domain, and a second nucleotide sequence domain, referred to as the "Cas endonuclease recognition" (CER) domain. The CER domain of a crRNA and tracrRNA polynucleotide duplex comprises two separate molecules (a nucleotide sequence 3' of the VT domain located on the crNucleotide and a tracrNucleotide) that are hybridized along a region of complementarity (FIG. **1**A). The VT domain helps facilitate DNA target site recognition while the CER domain promotes recognition by the Cas9 protein. Along with the required protospacer adjacent motif (PAM) sequence, both domains of the crRNA/tracrRNA polynucleotide duplex function to guide Cas endonuclease DNA target site cleavage and will herein be referred to as a "duplex guide polynucleotide", "duplex guide RNA" or "crRNA/tracrRNA duplex" as described in Example 1 of U.S. provisional application 61/868,706, filed Aug. 22, 2013.

[0377] To test a duplex guide polynucleotide/Cas endonuclease system in maize, the Cas9 gene from *Streptococcus pyogenes* M1 GAS (SF370) (SEQ ID NO: 1) was maize codon optimized per standard techniques known in the art and the potato ST-LS1 intron (SEQ ID NO: 2) was introduced in order to eliminate its expression in *E. coli* and *Agrobacterium* (FIG. **2**A). To facilitate nuclear localization of the Cas9 protein in maize cells, *Simian virus* 40 (SV40) monopartite (MAPKKKRKV, SEQ ID NO: 3) and *Agrobacterium tumefaciens* bipartite VirD2 T-DNA border endonuclease (KRPRDRHDGELGGRKRAR, SEQ ID NO: 4) nuclear localization signals were incorporated at the amino and carboxyl-termini of the Cas9 open reading frame, respectively (FIG. **2**A). The maize optimized Cas9 gene was operably linked to a maize constitutive or regulated promoter by standard molecular biological techniques. An example of the maize optimized Cas9 expression cassette (SEQ ID NO: 5) is illustrated in FIG. **2**A containing a maize optimized Cas9 gene with a ST-LS1 intron, SV40 amino terminal nuclear localization signal (NLS) and VirD2 carboxyl terminal NLS driven by a plant Ubiquitin promoter.

[0378] To confer efficient crRNA and tracrRNA expression in maize cells so that crRNA/tracrRNA polynucleotide duplexes may guide the Cas9 protein to cleave DNA target sites in vivo, the maize U6 polymerase III promoter (SEQ ID NO: 6) and maize U6 polymerase III terminator (TTTTTTTT) residing on chromosome 8 were isolated and operably fused as 5' and 3' terminal fusions, respectively, to both the crRNA and tracrRNA DNA coding sequences using standard molecular biology techniques generating expression cassettes as illustrated in FIG. **2**B and FIG. **2**C. Sequences of the resulting maize optimized crRNA and tracrRNA expression cassettes may be found in SEQ ID NO: 8 (crRNA expression cassette with a VT domain targeting the LIGCas-3 target site (Table 1) and SEQ ID NO: 9 (tracrRNA expression cassette).

[0379] As shown in FIG. **3**A, the crRNA molecule requires a region of complementarity to the DNA target (VT domain) that is approximately 12-30 nucleotides in length and upstream of a PAM sequence for target site recognition and cleavage (Gasiunas et al. (2012) *Proc. Natl. Acad. Sci. USA* 109:E2579-86, Jinek et al. (2012) *Science* 337:816-21, Mali et al. (2013) *Science* 339:823-26, and Cong et al. (2013) *Science* 339:819-23). To facilitate the rapid introduction of maize genomic DNA target sequences into the crRNA expression construct, two Type IIS BbsI restriction endonuclease target sites were introduced in an inverted tandem orientation with cleavage orientated in an outward direction as described in Cong et al. (2013) *Science* 339:819-23. Upon cleavage, the Type IIS restriction endonuclease excises its target sites from the crRNA expression plasmid, generating overhangs allowing for the in-frame directional cloning of duplexed oligos containing the desired maize genomic DNA target site into the VT domain. In the example shown, only target sequences starting with a G nucleotide were used to promote favorable polymerase III expression of the crRNA.

[0380] Expression of both the Cas endonuclease gene and the crRNA and tracrRNA molecules then allows for the formation of the duplex guide RNA/Cas endonuclease system (also referred to as crRNA/tracrRNA/Cas endonuclease complex) depicted in FIG. **3**A (SEQ ID NOs: 10-11).

Example 2

The Duplex Guide RNA/Cas Endonuclease System Cleaves Chromosomal DNA in Maize and Introduces Mutations by Imperfect Non-Homologous End-Joining

[0381] To test whether the maize optimized duplex guide RNA/Cas endonuclease system described in Example 1 could recognize, cleave, and mutate maize chromosomal DNA through imprecise non-homologous end-joining (NHEJ) repair pathways, three different genomic target sequences were targeted for cleavage (see Table 1) and examined by deep sequencing for the presence of NHEJ mutations.

TABLE 1

Maize genomic target sequences introduced into the crRNA expression cassette.

Locus	Location	Cas RNA System Used	Target Site Designation	Maize Genomic Target Site Sequence	PAM Sequence	SEQ ID NO:
LIG	Chr. 2: 28.45cM	crRNA/ tracrRNA	LIGCas-1	GTACCGTACGTGCCCCGGCGG	AGG	12
		crRNA/ tracrRNA	LIGCas-2	GGAATTGTACCGTACGTGCCC	CGG	13
		crRNA/ tracrRNA	LIGCas-3	GCGTACGCGTACGTGTG	AGG	14

LIG = approximately 600 bp upstream of the Liguleless 1 gene start codon

[0382] The maize optimized Cas9 endonuclease expression cassette, crRNA expression cassettes containing the specific maize VT domains complementary to the antisense strand of the maize genomic target sequences listed in Table 1 and tracrRNA expression cassette were co-delivered to 60-90 Hi-II immature maize embryos by particle-mediated delivery (see Example 7) in the presence of BBM and WUS2 genes (see Example 8). Hi-II maize embryos transformed with the Cas9 and long guide RNA expression cassettes (as described in U.S. provisional patent application 61/868,706, filed on Aug. 22, 2013) targeting the LIGCas-3 genomic target site for cleavage served as a positive control and embryos transformed with only the Cas9 expression cassette served as a negative control. After 7 days, the 20-30 most uniformly transformed embryos from each treatment were pooled and total genomic DNA was extracted. The region surrounding the intended target site was PCR amplified with Phusion® HighFidelity PCR Master Mix (New England Biolabs, M0531 L) adding on the sequences necessary for amplicon-specific barcodes and Illumnia sequencing using "tailed" primers through two rounds of PCR. The primers used in the primary PCR reaction are shown in Table 2 and the primers used in the secondary PCR reaction were AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACG (forward, SEQ ID NO: 21) and CAAGCAGAAGACGGCATA (reverse, SEQ ID NO: 22).

TABLE 2

PCR primer sequences

Target Site	Cas RNA System Used	Primer Orientation	Primary PCR Primer Sequence	SEQ ID NO:
LIGCas-1	crRNA/ tracrRNA	Forward	CTACACTCTTTCCCTACACGACGCTCTTCCGATCTTCCTCTGTAACGATTTACGCACCTGCTG	15
LIGCas-1	crRNA/ tracrRNA	Reverse	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGCAAATGAGTAGCAGCGCACGTAT	16
LIGCas-2	crRNA/ tracrRNA	Forward	CTACACTCTTTCCCTACACGACGCTCTTCCGATCTGAAGCTGTAACGATTTACGCACCTGCTG	17
LIGCas-2	crRNA/ tracrRNA	Reverse	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGCAAATGAGTAGCAGCGCACGTAT	16
LIGCas-3	crRNA/ tracrRNA	Forward	CTACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGGCGCAAATGAGTAGCAGCGCAC	18
LIGCas-3	crRNA/ tracrRNA	Reverse	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCACCTGCTGGGAATTGTACCGTA	19
LIGCas-3	Long guide RNA	Forward	CTACACTCTTTCCCTACACGACGCTCTTCCGATCTTTCCCGCAAATGAGTAGCAGCGCAC	20
LIGCas-3	Long guide RNA	Reverse	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCACCTGCTGGGAATTGTACCGTA	19

[0383] The resulting PCR amplifications were purified with a Qiagen PCR purification spin column, concentration measured with a Hoechst dye-based fluorometric assay, combined in an equimolar ratio, and single read 100 nucleotide-length deep sequencing was performed on Illumina's MiSeq Personal Sequencer with a 30-40% (v/v) spike of PhiX control v3 (Illumina, FC-110-3001) to off-set sequence bias. Only those reads with a ≥1 nucleotide indel arising within the 10 nucleotide window centered over the expected site of cleavage and not found in a similar level in the negative control were classified as NHEJ mutations. NHEJ mutant reads with the same mutation were counted and collapsed into a single read and the top 10 most prevalent mutations were visually confirmed as arising within the expected site of cleavage. The total numbers of visually confirmed NHEJ mutations were then used to calculate the % mutant reads based on the total number of reads of an appropriate length containing a perfect match to the barcode and forward primer.

[0384] The frequency of NHEJ mutations recovered by deep sequencing for the duplex guide RNA/Cas endonuclease system targeting the three LIGCas targets (SEQ ID NOs: 12-14) compared to the single long guide RNA/Cas endonuclease system targeting the same locus is shown in Table 3. The ten most prevalent types of NHEJ mutations recovered based on the duplex guide RNA/Cas endonuclease system are shown in FIG. **4**A (corresponding to SEQ ID NOs: 24-33 wherein SEQ ID NO: 23 is the reference maize sequence comprising the LIGCas-1 target site), FIG. **4**B (corresponding to SEQ ID NOs: 34-43 wherein SEQ ID NO: 23 is the reference maize sequence comprising the LIGCas-2 target site) and FIG. **4**C (corresponding to SEQ ID NOs: 45-54, wherein SEQ ID NO: 44 is the reference maize sequence comprising the LIGCas-3 target site).

[0385] Taken together, this data indicates that the maize optimized duplex guide RNA/Cas endonuclease system described herein cleaves maize chromosomal DNA and generates imperfect NHEJ mutations.

TABLE 3

Percent (%) mutant reads at maize Liguleless 1 target locus produced by duplex guide RNA/Cas endonuclease system compared to the long guide RNA/Cas endonuclease system

System	Total Number of Reads	Number of Mutant Reads	% Mutant Reads
Cas9 Only Control	1,744,427	0	0.00%
LIGCas-3 long guide RNA	1,596,955	35,300	2.21%
LIGCas-1 crRNA/tracrRNA	1,803,163	4,331	0.24%
LIGCas-2 crRNA/tracrRNA	1,648,743	3,290	0.20%
LIGCas-3 crRNA/tracrRNA	1,681,130	2,409	0.14%

Example 3

Deoxyribonucleic Acid (DNA) can be Used to Guide the Cas9 Protein to Cleave Maize Chromosomal DNA and Introduce Mutations by Imperfect Non-Homologous End-Joining

[0386] As previously described in Gasiunas et al. (2012) *Proc. Natl. Acad. Sci. USA* 109:E2579-86, Jinek et al. (2012) *Science* 337:816-21, Mali et al. (2013) *Science* 339:823-26 and Gong et al. (2013) *Science* 339:819-23, ribonucleic acids or RNA have been the only molecules described to guide a Cas9 endonuclease to recognize and cleave a specific DNA target site. In this example, we provide evidence that a new class of molecules, deoxyribonucleic acids (DNA), can also be used to guide a Cas endonuclease to recognize and cleave chromosomal DNA target sites resulting in the recovery of imperfect NHEJ mutations.

[0387] In this example, we used a duplex guide polynucleotide comprising of a first nucleotide sequence domain, referred to as the "variable targeting" (VT) domain, and a second nucleotide sequence domain, referred to as the "Cas endonuclease recognition" (CER) domain, wherein the variable targeting domain is a contiguous stretch of deoxyribonucleic acids (DNA). The CER domain of the duplex guide polynucleotide comprised two separate molecules, one DNA molecule that was linked to the VT domain of the crNucleotide molecule (FIG. **1**B) and hybridized along a region of complementarity to a second molecule (the tracrNucleotide, FIG. **1**A) consisting of a contiguous stretch of ribonucleic acids (RNA) nucleotides (referred to as tracrRNA). In this example the crNucleotide of the duplex guide polynucleotide (FIG. **1**A) consisted solely of DNA nucleotides and is herein referred to as crDNA.

[0388] The crDNA sequence containing VT domain targeting the LIGCas-3 target site (Table 1) (SEQ ID No: 55) was synthesized at Integrated DNA Technologies, Inc. with a 5' phosphate group and purified by PAGE and then used to test if a duplex guide crDNA-tracrRNA polynucleotide/Cas endonuclease complex as illustrated in FIG. **5** may recognize and cleave maize chromosomal DNA target sites resulting in the recovery of NHEJ mutations.

[0389] To determine the optimal delivery concentration for the synthetic crDNA molecules, different concentrations of crDNA (20 ng, 50 ng, 100 ng, 1 μg and 5 μg) were co-delivered along with a maize optimized tracrRNA and Cas9 expression cassettes to 60-90 Hi-II immature maize embryos and assayed for the presence of NHEJ mutations as described in Example 2. Embryos transformed with only the Cas9 and tracrRNA expression cassettes served as a negative control. As shown in Table 4, NHEJ mutations were detected with an optimal crDNA delivery concentration near 50 ng.

[0390] To compare the NHEJ mutational activity of the duplex guide crDNA-tracrRNA polynucleotide/Cas endonuclease complex (FIG. **5**) with a duplex guide RNA (crRNA-tracrRNA)/Cas endonuclease complex (FIG. **3**A), a crRNA comprising a VT domain targeting the LIGCas-3 target site (Table 1) was synthesized at Bio-Synthesis, Inc. with a 5' phosphate group and purified by PAGE (SEQ ID NO: 10). 50 ng of both the synthetic crDNA and crRNA were then independently co-delivered along with the maize optimized tracrRNA and Cas9 DNA expression cassettes and assayed for NHEJ mutations as described previously. The transformation experiment was performed twice to demonstrate reproducibility. Negative controls consisted of Hi-II maize embryos transformed with 50 ng of crDNA, the Cas9 expression cassette or 50 ng of crDNA plus the tracrRNA expression cassette.

[0391] As shown in Tables 4 and 5, NHEJ mutations resulting from the duplex guide crDNA-tracRNA polynucleotide/Cas endonuclease system were identified when the crDNA was delivered in combination with a tracrRNA and Cas9 DNA expression cassettes compared to the absence of NHEJ mutations in the Cas9 only, crDNA only, crDNA plus tracrRNA only and tracrRNA plus Cas9 only controls. The top 3 most abundant crRNA NHEJ mutations are shown in FIG. **6** A (SEQ ID NO:56, 57, 58, wherein SEQ ID NO:44 is the unmodified reference sequence for LIGCas-3 locus) and the top 3 most abundant crDNA NHEJ mutations are shown in FIG. **6** B (SEQ ID NO: 59, 60, 61, wherein SEQ ID NO:44 is the unmodified reference sequence for LIGCas-3 locus) identified are compared and shown in FIG. **6** (SEQ ID NOs: 44, 56-61).

[0392] Taken together, this data indicates that deoxyribonucleic acids (DNA) may also be used to guide Cas endonucleases in a duplex guide crDNA-tracRNA polynucleotide/

Cas endonuclease complex (FIG. **4**) to cleave maize chromosomal DNA resulting in imprecise NHEJ mutations.

TABLE 4

Percent (%) mutant reads at maize Liguleless 1 target locus produced by different concentrations of transiently delivered crDNA molecules co-delivered with tracrRNA and Cas9 DNA expression cassettes

Amount of crDNA Delivered	DNA Expression Cassettes Delivered	Total Number of Reads	Number of Mutant Reads
0	Cas9, tracrRNA	1,046,553	0
20 ng	Cas9, tracrRNA	926,915	0
50 ng	Cas9, tracrRNA	1,032,080	18
100 ng	Cas9, tracrRNA	860,565	8
1 μg	Cas9, tracrRNA	398,996	0
5 μg	Cas9, tracrRNA	394,959	0

TABLE 5

Comparison of percent (%) mutant reads at maize Liguleless 1 target locus produced by transiently delivered crDNA or crRNA molecules co-delivered with tracrRNA and Cas9 DNA expression cassettes

Synthetic crDNA or crRNA Delivered	Trans-formation Replicate	DNA Expression Cassettes Delivered	Total Number of Reads	Number of Mutant Reads
—	1	Cas9	1,151,532	0
crDNA	1	—	1,234,489	0
crDNA	1	tracrRNA	666,151	0
crD NA	1	tracrRNA, Cas9	1,046,189	40
crDNA	2	tracrRNA, Cas9	913,430	39
crRNA	1	tracrRNA, Cas9	1,217,740	136
crRNA	2	tracrRNA, Cas9	1,028,995	281

Example 4

Modifying Nucleic Acid Component(s) of the Guide Polynucleotide/Cas Endonuclease System to Increase Cleavage Activity and Specificity

[0393] In this example, modifying the nucleotide base, phosphodiester bond linkage or molecular topography of the guiding nucleic acid component(s) of the guide polynucleotide/Cas endonuclease system is described for increasing cleavage activity and specificity.

[0394] As shown in FIG. **1**A and FIG. **1**B, the nucleic acid component(s) of the guide polynucleotide include a variable targeting (VT) domain and a Cas endonuclease recognition (CER) domain. The VT domain is responsible for interacting with the DNA target site through direct nucleotide-nucleotide base pairings while the CER domain is required for proper Cas endonuclease recognition (FIG. **3**A and FIG. **3**B). Along with the required PAM sequence, both domains of the nucleic acid component(s) of the guide polynucleotide/Cas endonuclease system function to link DNA target site recognition with Cas endonuclease target site cleavage (FIG. **3**A and FIG. **3**B).

[0395] Given the direct interaction of the VT domain with the DNA target site, nucleotide base modifications within the VT domain can be utilized to alter the nucleotide-nucleotide base pairing relationships facilitating Cas endonuclease target site recognition. Such modifications can be used to strengthen the binding affinity to the complementary DNA target sequence enhancing guide polynucleotide/Cas endonuclease target site recognition and/or specificity. Non-limiting examples of nucleotide base modifications that can enhance target site binding affinity and/or specificity when introduced into the VT domain of a guide polynucleotide are listed in Table 6. These modifications can be used individually or in combination within the VT domain.

TABLE 6

Nucleotide base modifications to enhance nucleotide base pairing with complementary DNA target sequence

Modification	Effect
Deoxyribonucleic Acid	More specific hybridization to complementary DNA target sequence[1]
Locked Nucleic Acid	Increased binding affinity and more specific hybridization to complementary DNA target sequence[2]
5-methyl dC	Increase binding affinity to complementary DNA target sequence[3]
2,6-Diaminopurine	Increase binding affinity to complementary DNA target sequence[4]
2'-Fluoro A or U	Increase binding affinity to complementary DNA target sequence[5]

[0396] Nucleotide modifications similar to those shown in Table 6 can also be made in the CER domain of the single or duplex guide polynucleotide (FIG. **1**A and FIG. **1**B). These modifications may act to strengthen or stabilize inter-molecular interactions in the CER domain of a duplexed crNucleotide molecule (for example, but not limiting to crRNA, crDNA or a combination thereof) and tracrNucleotide molecule (for example tracrRNA, tracrDNA, or a combination thereof)(FIG. **3**A). These modifications can also help recapitulate crNucleotide and tracrNucleotide (such as for example but not limiting to crRNA/tracrRNA; crDNA/tracrNA; crRNA/tracrDNA, crDNA/tracrDNA) structures required for proper Cas endonuclease recognition in the secondary structure of guide polynucleotides being comprised of a single molecule (FIG. **3**A and FIG. **3**B).

[0397] Nucleic acids expressed or delivered transiently to cells are subject to turnover or degradation. To increase the effective lifespan or stability of the nucleic acid component(s) of the guide polynucleotide/Cas endonuclease system in vivo, nucleotide and/or phosphodiester bond modifications may be introduced to reduce unwanted degradation. Examples of nuclease resistant nucleotide and phosphodiester bond modifications are shown in Table 7 and may be introduced in any one of the VT and/or CER domains of the guide polynucleotide. Modifications may be introduced at the 5' and 3' ends of any one of the nucleic acid residues comprising the VT or CER domains to inhibit exonuclease cleavage activity, can be introduced in the middle of the nucleic acid sequence comprising the VT or CER domains to slow endonuclease cleavage activity or can be introduced throughout the nucleic acid sequences comprising the VT or CER domains to provide protection from both exo- and endo-nucleases.

TABLE 7

Nucleotide base and phosphodiester bond modifications to decrease unwanted nuclease degradation.

Modification	Effect
Deoxyribonucleic Acid	Less susceptible to nuclease degradation than RNA[1]

TABLE 7-continued

Nucleotide base and phosphodiester bond modifications to decrease unwanted nuclease degradation.

Modification	Effect
Locked Nucleic Acid	Very resistant to nuclease cleavage[2]
2'-Fluoro A or U	Increased resistance to nuclease cleavage[3]
2'-O—Methyl RNA Bases	Resistant to ribonucleases and 5-10 fold more resistant to DNases than DNA[4]
Phosphorothioate bond	Very resistant to nuclease cleavage[5]

[0398] To provide resistance against turnover or degradation in cells, the nucleic acid component(s) of the guide polynucleotide may also be circularized where the 5' and 3' ends are covalently joined together. Circular RNA can be more resistant to nuclease degradation than linear RNA and can persist in cells long after corresponding linear transcripts (Jeck et al. (2013) *RNA* 19:141-157).

[0399] Modifications to any one of the guide polynucleotide nucleic acid components may also be introduced to increase their permeability or delivery into cells. Such modifications would include, but not be limited to, linkage to cholesterol, polyethylene glycol and spacer 18 (hexaethylene glycol chain).

[0400] Many of the above mentioned modified guide polynucleotides can be synthesized and delivered transiently by biolistic particle-mediated transformation, transfection or electroporation. The remaining components of the guide polynucleotide/Cas endonuclease system needed to form a functional complex capable of binding and/or cleaving a chromosomal DNA target site may be co-delivered as any combination of DNA expression cassettes, RNA, mRNA (5'-capped and polyadenylated) or protein. Cell lines or transformants may also be established stably expressing all but one or two of the components needed to form a functional guide polynucleotide/Cas endonuclease complex so that upon transient delivery of the above mentioned modified nucleic acid guide(s) a functional guide polynucleotide/Cas endonuclease complex may form. Modified guide polynucleotides described above may also be delivered simultaneously in multiplex to target multiple chromosomal DNA sequences for cleavage or nicking.

[0401] The above mentioned modified guide polynucleotides may be used in plants, animals, yeast and bacteria or in any organism subject to genome modification with the guide polynucleotide/Cas endonuclease system and be used to introduce imprecise NHEJ mutations into chromosomal DNA, excise chromosomal DNA fragments comprised of either transgenic or endogenous DNA, edit codon composition of native or transgenic genes by homologous recombination repair with a donor DNA repair template(s) and site-specifically insert transgenic or endogenous DNA sequences by homologous recombination repair with a donor DNA repair template(s).

Example 5

Examining the Effect of Nucleotide Base and Phosphodiester Bond Modifications to the Guide Polynucleotide Component of the Guide Polynucleotide/Cas Endonuclease System in Maize

[0402] In this example, some of the nucleotide base and phosphodiester bond modifications described in Example 4 are introduced into the VT domain and/or CER domain of a crNucleotide and methods for evaluating the impact of these modifications on the ability of a duplexed guide polynucleotide/Cas endonuclease system to cleave maize chromosomal DNA will be discussed.

[0403] As illustrated in Table 8, nucleotide base and phosphodiester linkage modifications were introduced individually or in combination into the VT domain and the CER domain of the crNucleotide (crRNA or crDNA) component of the duplexed guide polynucleotide/Cas endonuclease system targeting the LIGCas-3 site (see Table 1) for cleavage. Although a number of different nucleotide base and phosphodiester linkage modifications are examined in combination here, other possible combinations may be envisioned.

[0404] Locked Nucleic Acid (+), 5-Methyl dC (iMe-dC) and 2,6-Diaminopurine (i6diPr) nucleotide base modifications made in the VT domain are introduced to increase the binding affinity to the complementary DNA target sequence and in the case of the Locked Nucleic Acid modifications to also increase resistance to in vivo nucleases. All other modifications designed in both the VT and CER domains at the 5' and 3' ends or throughout the crRNA or crDNA sequence are introduced to decrease the effect of in vivo nucleases and increase the effective lifespan of the crRNA or crDNA component.

[0405] To examine the effect that the modified crRNA or crDNA components described in Table 8 have on the ability of their associated modified guide polynucleotide/Cas endonuclease complex to recognize and cleave the LIGCas-3 site (see Table 1), the modified crRNA and crDNA molecules are co-delivered to Hill immature maize embryos with tracrRNA and Cas9 expression cassettes as described in Example 3. Unmodified crRNA or crDNA molecules co-delivered with tracrRNA and Cas9 expression cassettes serve as comparators. Negative controls consist of immature maize embryos transformed with only the corresponding modified crRNA or crDNA or Cas9 expression cassette. Frequencies of imperfect NHEJ mutations, assayed as described in Example 2, are used to evaluate the effect of each crRNA or crDNA modification on Cas endonuclease cleavage activity relative to the comparable unmodified crRNA or crDNA experiments.

TABLE 8

crRNA and crDNA nucleotide base and phosphodiester linkage modifications.

Nucleic Acid		crRNA or crDNA Sequence and Corresponding Modification[1]	
Type	Modification	VT Domain	CER Domain
crRNA	None	GCGUACGCGUACGUGUG (SEQ ID NO: 62)	GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO: 63)

TABLE 8 -continued

crRNA and crDNA nucleotide base and phosphodiester linkage modifications.

Nucleic Acid		crRNA or crDNA Sequence and Corresponding Modification[1]	
Type	Modification	VT Domain	CER Domain
crRNA	Phosphorothioate bonds near ends	G*C*G*UACGCGUACGUGUG (SEQ ID NO: 64)	GUUUUAGAGCUAUGCUGUU*U*U*G (SEQ ID NO: 65)
crRNA	2'-O-Methyl RNA nucleotides at ends	mGmCmGUACGCGUACGUGUG (SEQ ID NO: 66)	GUUUUAGAGCUAUGCUGUUmUmUmG (SEQ ID NO: 67)
crRNA	2'-O-Methyl RNA nucleotides for each nucleotide	mGmCmGmUmAmCmGmCmG mUmAmCmGmUmGmUmG (SEQ ID NO: 68)	mGmUmUmUmUmAmGmAmGmC mUmAmUmGmCmUmGmUmUmUmG (SEQ ID NO: 69)
crDNA	None	GCGTACGCGTACGTGTG (SEQ ID NO: 70)	GTTTTAGAGCTATGCTGTTTTG (SEQ ID NO: 71)
crDNA	1 Locked Nucleic Acid nucleotide in the variable targeting domain	GCGTACGCGTACGTG+TG (SEQ ID NO: 72)	GTTTTAGAGCTATGCTGTTTTG (SEQ ID NO: 71)
crDNA	3 Locked Nucleic Acid nucleotides in the variable targeting domain	GCGTACGCGTA+CG+TG+TG (SEQ ID NO: 73)	GTTTTAGAGCTATGCTGTTTTG (SEQ ID NO: 71)
crDNA	6 Locked Nucleic Acid nucleotides in the variable targeting domain	GCGTA+CG+CG+TA+CG+TG+TG (SEQ ID NO: 74)	GTTTTAGAGCTATGCTGTTTTG (SEQ ID NO: 71)
crDNA	3 Locked Nucleic Acid nucleotides in the variable targeting domain plus Phosphorothioate bonds near ends	G*C*G*TACGCGTA+CG+TG+TG (SEQ ID NO: 75)	GTTTTAGAGCTATGCTGTT*T*T*G (SEQ ID NO: 76)
crDNA	One 5-Methyl dC nucleotide in variable targeting domain	GCGTACGCGTA/iMe-dC/GTGTG (SEQ ID NO: 77)	GTTTTAGAGCTATGCTGTTTTG (SEQ ID NO: 71)
crDNA	Three 5-Methyl dC nucleotides in the variable targeting domain	GCGTA/iMe-dC/G/iMe-dC/ GTA/iMe-dC/GTGTG (SEQ ID NO: 78)	GTTTTAGAGCTATGCTGTTTTG (SEQ ID NO: 71)
crDNA	One 2,6-Diaminopurine nucleotide in the variable targeting domain	GCGTACGCGT/i6diPr/CGTGTG (SEQ ID NO: 79)	GTTTTAGAGCTATGCTGTTTTG (SEQ ID NO: 71)
crDNA	Two 2,6-Diaminopurine nucleotides in the variable targeting domain	GCGT/i6diPr/CGCGT/i6diPr/ CGTGTG (SEQ ID NO: 80)	GTTTTAGAGCTATGCTGTTTTG (SEQ ID NO: 71)
crDNA	Locked Nucleic Acid nucleotides at ends	+G+C+GTACGCGTACGTGTG (SEQ ID NO: 81)	GTTTTAGAGCTATGCTGTT+T+T+G (SEQ ID NO: 82)
crDNA	Phosphorothioate bonds near ends	G*C*G*TACGCGTACGTGTG (SEQ ID NO: 83)	GTTTTAGAGCTATGCTGTT*T*T*G (SEQ ID NO: 76)
crDNA	2'-O-Methyl RNA nucleotides at ends	mGmCmGTACGCGTACGTGTG (SEQ ID NO: 84)	GTTTTAGAGCTATGCTGTTmUmUmG (SEQ ID NO: 85)

TABLE 8 -continued

crRNA and crDNA nucleotide base and phosphodiester linkage modifications.

Nucleic Acid Type	Modification	crRNA or crDNA Sequence and Corresponding Modification[1] VT Domain	CER Domain
crDNA	2'-O-Methyl RNA nucleotides at each nucleotide except T	mGmCmGTmAmCmGmCmGTm AmCmGTmGTmG (SEQ ID NO: 86)	mGTTTTmAmGmAmGmCTmATm GmCTmGTTTTmG (SEQ ID NO: 87)

[1] "+" before nucleotide denotes lock nucleic acid base modification,
"*" after nucleotide denotes Phosphorothioate bond backbone modification,
"m" before nucleotide denotes 2'-O-Methyl RNA base modification,
"iMe-dC" denotes 5-Methyl dC base modification and
"i6diPr" denotes 2,6-Diaminopurine base modification

Example 6

Methods to Examine the Effect of Modifications to the Nucleic Acid Component(s) of the Guide Polynucleotide/Cas Endonuclease System in Yeast

[0406] In this example, yeast screening methods are devised to identify optimal modifications to the nucleic acid component(s) of the guide polynucleotide/Cas endonuclease system that result in enhanced cleavage activity.

a. ADE:URA3:DE2 Yeast Screening Strain

[0407] To identify optimal modification(s) or combinations thereof to the nucleic acid component(s) of the guide polynucleotide/Cas endonuclease system outlined in Example 4, a *Saccharomyces cerevisiae* strain is developed to carefully monitor the cleavage activity of the guide polynucleotide/Cas endonuclease system. This is accomplished by replacing the native ADE2 gene on chromosome 15 of yeast strain BY4247 with a non-functional partially duplicated ADE2 gene disrupted by the yeast URA3 gene (ADE:URA3:DE2) as shown in FIG. **7**. A guide polynucleotide/Cas endonuclease target site adjacent to the appropriate PAM sequence is then designed against the implanted URA3 gene so that upon cleavage the disrupted ADE2 gene containing 305 bp of duplicated overlapping sequence can be repaired by intramolecular homologous recombination pathways resulting in the loss of the URA3 gene and the gain of a functional ADE2 gene as shown in FIG. **8**. Media containing 5-Fluoroorotic Acid (5-FOA) or media deficient in adenine can then be used to select for cells where cleavage has occurred. The frequency of yeast cells recovered after selection can then be used to quantify the cleavage efficiency of the guide polynucleotide/Cas endonuclease system when examining different modifications to the nucleic acid component(s) of the guide polynucleotide/Cas endonuclease system.

[0408] Yeast cells containing a functional ADE2 gene as a result of cleavage and repair of the ADE:URA3:DE2 locus can also be subject to a visual phenotypic screen for cleavage activity. In the absence of 5-FOA or adenine minus selection, functional ADE2 gene products result in a white phenotype while non-functional products result in a red phenotype (Ugolini et al. (1996) *Curr. Genet.* 30:485-492). To visualize the white or red phenotype, individual yeast cell transformants can be plated on solid media and allowed to grow into a colony large enough to inspect visually. The amount of white to red sectoring provides an indication as to the amount of cleavage activity. Since the sectoring phenotype is a qualitative measure, a 0-4 numerical scoring system can be implemented. As shown in FIG. **9**, a score of 0 indicates that no white sectors (no target site cleavage) were observed; a score of 4 indicates completely white colonies (complete cutting of the recognition site); scores of 1-3 indicate intermediate white sectoring phenotypes (and intermediate degrees of target site cleavage).

B. Cas9 Component of the Guide Polynucleotide/Cas Endonuclease System

[0409] To stably express the Cas endonuclease for pairing with the transiently delivered modified nucleic acid component(s) described in Example 4, the Cas9 gene from *Streptococcus pyogenes* M1 GAS (SF370) can be *S. cerevisiae* codon optimized per standard techniques known in the art (SEQ ID NO: 88) and a SV40 (Simian virus 40) nuclear localization signal (SRADPKKKRKV, SEQ ID NO: 7) can be incorporated at the carboxyl terminal to facilitate nuclear localization. The resulting Cas9 open reading frame will then be operably fused to the yeast inducible GAL1 promoter and CYC1 terminator. The resulting Cas9 expression cassette will then be placed into a CEN6 autonomously replicating yeast vector containing a LEU2 selectable marker.

[0410] To be able to test transient delivery of both the Cas9 component and the modified nucleic acid component(s) described in Example 4, the Cas9 gene can also be delivered as mRNA. To generate *S. cerevisiae* optimized Cas9 mRNA, PCR can be used to amplify the *S. cerevisiae* optimized Cas9 open reading frame and associated nuclear localization signal tailing on the required T7 promoter sequence (TAATACGACTCACTATAGGG, SEQ ID NO: 89) just 5' of the translation ATG start site. The resulting linear template containing the T7 promoter can then be used to transcribe uncapped or capped Cas9 mRNA with or without polyadenylation in vitro.

[0411] The Cas9 component can also be delivered transiently as protein and paired with the modified nucleic acid component(s). Cas9 protein with associated carboxyl-terminal nuclear localization signal can be expressed and purified per standard techniques similar to that described by Fonfara et al. (2013) *Nucl. Acids Res.* doi:10.1093/nar/gkt1074 or by other methods.

C. Nucleic Acid Component(s) of the Guide Polynucleotide/Cas Endonuclease System

[0412] It can be advantageous to pair the transient delivery of modified crRNA or tracrRNA components with the corresponding stably expressed unmodified crRNA or tracrRNA.

To facilitate stable expression of crRNA and tracrRNA in yeast, *S. cerevisiae* optimized crRNA and tracrRNA expression cassettes can be generated. The yeast RNA polymerase III SNR52 promoter and SUP4 terminator can be operably fused to the ends of DNA fragments encoding the appropriate crRNA and tracrRNA sequences required for recognition by the *S. pyogenes* Cas9 protein. All crRNA expression cassettes will contain the ADE:URA3:DE2 target sequence (GCAGACATTACGAATGCACA, SEQ ID NO: 90) in the VT domain and target the ADE:URA3:DE2 locus for cleavage. The resulting expression cassettes will then be placed into a CEN6 autonomously replicating yeast vector containing a HIS3 selectable marker.

[0413] To deliver unmodified crRNA, tracrRNA or guide RNA transiently, PCR can be used to amplify the corresponding crRNA, tracrRNA or guide RNA sequence tailing on the required T7 promoter sequence (TAATACGACTCACTATAGGG, SEQ ID NO: 89) just 5' of the transcriptional start site. The resulting linear template containing the T7 promoter can then be used to transcribe the corresponding crRNA, tracrRNA or long guide RNA.

[0414] Modified nucleic acid components(s) of the guide polynucleotide/Cas endonuclease system as outlined in Example 4 will also be transiently delivered. Nucleotide base and/or phosphodiester bond modifications similar to those illustrated in Example 5 Table 8 can be introduced individually or in combination into the crRNA, crDNA, tracrRNA, tracrDNA, long guide RNA or long guide DNA nucleic acid components of the guide polynucleotide/Cas endonuclease system and synthesized per standard techniques.

[0415] Circular RNAs, also discussed in Example 4, containing the necessary VT and CER domains capable of forming a functional complex with the Cas endonuclease can be generated in vitro as described by Diegleman et al. (1998) Nucl. Acids Res. 26:3235-3241 and delivered transiently to yeast cells.

D. Transformation of Guide Polynucleotide/Cas Endonuclease Components into the ADE:URA3:DE2 Yeast Strain

[0416] Components of the guide polynucleotide/Cas endonuclease system can be delivered to ADE:URA3:DE2 yeast cells using standard lithium acetate, polyethylene glycol (PEG), electroporation or biolistic transformation methods and monitored for their ability to cleave the ADE:URA3:DE2 target. The yeast optimized guide polynucleotide/Cas endonuclease components discussed in Example 6 sections B and C can be delivered as expression cassettes on low copy autonomously replicating plasmid DNA vectors, as non-replicating transient molecules (such as mRNA, protein, RNA or modified guide nucleic acids) or in any combination of plasmid DNA vector expression cassette(s) and transient molecule(s).

Example 7

Transformation of Maize Immature Embryos

[0417] Transformation can be accomplished by various methods known to be effective in plants, including particle-mediated delivery, *Agrobacterium*-mediated transformation, PEG-mediated delivery, and electroporation.

[0418] a. Particle-Mediated Delivery

[0419] Transformation of maize immature embryos using particle delivery is performed as follows. Media recipes follow below.

[0420] The ears are husked and surface sterilized in 30% Clorox bleach plus 0.5% Micro detergent for 20 minutes, and rinsed two times with sterile water. The immature embryos are isolated and placed embryo axis side down (scutellum side up), 25 embryos per plate, on 560Y medium for 4 hours and then aligned within the 2.5-cm target zone in preparation for bombardment. Alternatively, isolated embryos are placed on 560 L (Initiation medium) and placed in the dark at temperatures ranging from 26° C. to 37° C. for 8 to 24 hours prior to placing on 560Y for 4 hours at 26° C. prior to bombardment as described above.

[0421] Plasmids containing the double strand brake inducing agent and donor DNA are constructed using standard molecular biology techniques and co-bombarded with plasmids containing the developmental genes ODP2 (AP2 domain transcription factor ODP2 (Ovule development protein 2); US20090328252 A1) and Wushel (US2011/0167516).

[0422] The plasmids and DNA of interest are precipitated onto 0.6 μm (average diameter) gold pellets using a water-soluble cationic lipid transfection reagent as follows. DNA solution is prepared on ice using 1 μg of plasmid DNA and optionally other constructs for co-bombardment such as 50 ng (0.5 μl) of each plasmid containing the developmental genes ODP2 (AP2 domain transcription factor ODP2 (Ovule development protein 2); US20090328252 A1) and Wushel. To the pre-mixed DNA, 20 μl of prepared gold particles (15 mg/ml) and 5 μl of the a water-soluble cationic lipid transfection reagent is added in water and mixed carefully. Gold particles are pelleted in a microfuge at 10,000 rpm for 1 min and supernatant is removed. The resulting pellet is carefully rinsed with 100 ml of 100% EtOH without resuspending the pellet and the EtOH rinse is carefully removed. 105 μl of 100% EtOH is added and the particles are resuspended by brief sonication. Then, 10 μl is spotted onto the center of each macrocarrier and allowed to dry about 2 minutes before bombardment.

[0423] Alternatively, the plasmids and DNA of interest are precipitated onto 1.1 μm (average diameter) tungsten pellets using a calcium chloride ($CaCl_2$) precipitation procedure by mixing 100 μl prepared tungsten particles in water, 10 μl (1 μg) DNA in Tris EDTA buffer (1 μg total DNA), 100 μl 2.5 M CaC12, and 10 μl 0.1 M spermidine. Each reagent is added sequentially to the tungsten particle suspension, with mixing. The final mixture is sonicated briefly and allowed to incubate under constant vortexing for 10 minutes. After the precipitation period, the tubes are centrifuged briefly, liquid is removed, and the particles are washed with 500 ml 100% ethanol, followed by a 30 second centrifugation. Again, the liquid is removed, and 105 μl 100% ethanol is added to the final tungsten particle pellet. For particle gun bombardment, the tungsten/DNA particles are briefly sonicated. 10 μl of the tungsten/DNA particles is spotted onto the center of each macrocarrier, after which the spotted particles are allowed to dry about 2 minutes before bombardment.

[0424] The sample plates are bombarded at level #4 with a Biorad Helium Gun. All samples receive a single shot at 450 PSI, with a total of ten aliquots taken from each tube of prepared particles/DNA.

[0425] Following bombardment, the embryos are incubated on 560P (maintenance medium) for 12 to 48 hours at temperatures ranging from 26C to 37C, and then placed at 26C. After 5 to 7 days the embryos are transferred to 560R selection medium containing 3 mg/liter Bialaphos, and subcultured every 2 weeks at 26C. After approximately 10 weeks of selection, selection-resistant callus clones are transferred to 288J medium to initiate plant regeneration. Following somatic embryo maturation (2-4 weeks), well-developed somatic embryos are transferred to medium for germination and transferred to a lighted culture room. Approximately 7-10 days later, developing plantlets are transferred to 272V hormone-free medium in tubes for 7-10 days until plantlets are well established. Plants are then transferred to inserts in flats (equivalent to a 2.5" pot) containing potting soil and grown for 1 week in a growth chamber, subsequently grown an additional 1-2 weeks in the greenhouse, then transferred to Classic 600 pots (1.6 gallon) and grown to maturity. Plants are monitored and scored for transformation efficiency, and/or modification of regenerative capabilities.

[0426] Initiation medium (560 L) comprises 4.0 g/l N6 basal salts (SIGMA C-1416), 1.0 ml/l Eriksson's Vitamin Mix (1000×SIGMA-1511), 0.5 mg/l thiamine HCl, 20.0 g/l sucrose, 1.0 mg/l 2,4-D, and 2.88 g/l L-proline (brought to volume with D-1H2O following adjustment to pH 5.8 with KOH); 2.0 g/l Gelrite (added after bringing to volume with D-1H2O); and 8.5 mg/l silver nitrate (added after sterilizing the medium and cooling to room temperature).

[0427] Maintenance medium (560P) comprises 4.0 g/l N6 basal salts (SIGMA C-1416), 1.0 ml/l Eriksson's Vitamin Mix (1000×SIGMA-1511), 0.5 mg/l thiamine HCl, 30.0 g/l sucrose, 2.0 mg/l 2,4-D, and 0.69 g/l L-proline (brought to volume with D-I H2O following adjustment to pH 5.8 with KOH); 3.0 g/l Gelrite (added after bringing to volume with D-1H2O); and 0.85 mg/l silver nitrate (added after sterilizing the medium and cooling to room temperature).

[0428] Bombardment medium (560Y) comprises 4.0 g/l N6 basal salts (SIGMA C-1416), 1.0 ml/l Eriksson's Vitamin Mix (1000×SIGMA-1511), 0.5 mg/l thiamine HCl, 120.0 g/l sucrose, 1.0 mg/l 2,4-D, and 2.88 g/l L-proline (brought to volume with D-I H2O following adjustment to pH 5.8 with KOH); 2.0 g/l Gelrite (added after bringing to volume with D-1H2O); and 8.5 mg/l silver nitrate (added after sterilizing the medium and cooling to room temperature).

[0429] Selection medium (560R) comprises 4.0 g/l N6 basal salts (SIGMA C-1416), 1.0 ml/l Eriksson's Vitamin Mix (1000×SIGMA-1511), 0.5 mg/l thiamine HCl, 30.0 g/l sucrose, and 2.0 mg/l 2,4-D (brought to volume with D-1H2O following adjustment to pH 5.8 with KOH); 3.0 g/l Gelrite (added after bringing to volume with D-1H2O); and 0.85 mg/l silver nitrate and 3.0 mg/l bialaphos (both added after sterilizing the medium and cooling to room temperature).

[0430] Plant regeneration medium (288J) comprises 4.3 g/l MS salts (GIBCO 11117-074), 5.0 ml/l MS vitamins stock solution (0.100 g nicotinic acid, 0.02 g/l thiamine HCL, 0.10 g/l pyridoxine HCL, and 0.40 g/l glycine brought to volume with polished D-1H2O) (Murashige and Skoog (1962) Physiol. Plant. 15:473), 100 mg/l myo-inositol, 0.5 mg/l zeatin, 60 g/l sucrose, and 1.0 ml/l of 0.1 mM abscisic acid (brought to volume with polished D-1H2O after adjusting to pH 5.6); 3.0 g/l Gelrite (added after bringing to volume with D-1H2O); and 1.0 mg/l indoleacetic acid and 3.0 mg/l bialaphos (added after sterilizing the medium and cooling to 60° C.). Hormone-free medium (272V) comprises 4.3 g/l MS salts (GIBCO 11117-074), 5.0 ml/l MS vitamins stock solution (0.100 g/l nicotinic acid, 0.02 g/l thiamine HCL, 0.10 g/l pyridoxine HCL, and 0.40 g/l glycine brought to volume with polished D-1H2O), 0.1 g/l myo-inositol, and 40.0 g/l sucrose (brought to volume with polished D-1H2O after adjusting pH to 5.6); and 6 g/l bacto-agar (added after bringing to volume with polished D-1H2O), sterilized and cooled to 60° C.

[0431] b. *Agrobacterium*-Mediated Transformation

[0432] *Agrobacterium*-mediated transformation was performed essentially as described in Djukanovic et al. (2006) *Plant Biotech J*4:345-57. Briefly, 10-12 day old immature embryos (0.8-2.5 mm in size) were dissected from sterilized kernels and placed into liquid medium (4.0 g/L N6 Basal Salts (Sigma C-1416), 1.0 ml/L Eriksson's Vitamin Mix (Sigma E-1511), 1.0 mg/L thiamine HCl, 1.5 mg/L 2,4-D, 0.690 g/L L-proline, 68.5 g/L sucrose, 36.0 g/L glucose, pH 5.2). After embryo collection, the medium was replaced with 1 ml *Agrobacterium* at a concentration of 0.35-0.45 OD550. Maize embryos were incubated with *Agrobacterium* for 5 min at room temperature, then the mixture was poured onto a media plate containing 4.0 g/L N6 Basal Salts (Sigma C-1416), 1.0 ml/L Eriksson's Vitamin Mix (Sigma E-1511), 1.0 mg/L thiamine HCl, 1.5 mg/L 2,4-D, 0.690 g/L L-proline, 30.0 g/L sucrose, 0.85 mg/L silver nitrate, 0.1 nM acetosyringone, and 3.0 g/L Gelrite, pH 5.8. Embryos were incubated axis down, in the dark for 3 days at 20° C., then incubated 4 days in the dark at 28° C., then transferred onto new media plates containing 4.0 g/L N6 Basal Salts (Sigma C-1416), 1.0 ml/L Eriksson's Vitamin Mix (Sigma E-1511), 1.0 mg/L thiamine HCl, 1.5 mg/L 2,4-D, 0.69 g/L L-proline, 30.0 g/L sucrose, 0.5 g/L MES buffer, 0.85 mg/L silver nitrate, 3.0 mg/L Bialaphos, 100 mg/L carbenicillin, and 6.0 g/L agar, pH 5.8. Embryos were subcultured every three weeks until transgenic events were identified. Somatic embryogenesis was induced by transferring a small amount of tissue onto regeneration medium (4.3 g/L MS salts (Gibco 11117), 5.0 ml/L MS Vitamins Stock Solution, 100 mg/L myo-inositol, 0.1 μM ABA, 1 mg/L IAA, 0.5 mg/L zeatin, 60.0 g/L sucrose, 1.5 mg/L Bialaphos, 100 mg/L carbenicillin, 3.0 g/L Gelrite, pH 5.6) and incubation in the dark for two weeks at 28° C. All material with visible shoots and roots were transferred onto media containing 4.3 g/L MS salts (Gibco 11117), 5.0 ml/L MS Vitamins Stock Solution, 100 mg/L myo-inositol, 40.0 g/L sucrose, 1.5 g/L Gelrite, pH 5.6, and incubated under artificial light at 28° C. One week later, plantlets were moved into glass tubes containing the same medium and grown until they were sampled and/or transplanted into soil.

Example 8

Transient Expression of BBM Enhances Transformation

[0433] Parameters of the transformation protocol can be modified to ensure that the BBM activity is transient. One such method involves precipitating the BBM-containing plasmid in a manner that allows for transcription and expression, but precludes subsequent release of the DNA, for example, by using the chemical PEI. In one example, the BBM plasmid is precipitated onto gold particles with PEI, while the transgenic expression cassette (UBI::moPAT~GFPm::PinII; moPAT is the maize optimized PAT gene) to be integrated is precipitated onto gold particles using the standard calcium chloride method.

[0434] Briefly, gold particles were coated with PEI as follows. First, the gold particles were washed. Thirty-five mg of gold particles, 1.0 in average diameter (A.S.I. #162-0010), were weighed out in a microcentrifuge tube, and 1.2 ml absolute EtOH was added and vortexed for one minute. The tube was incubated for 15 minutes at room temperature and then centrifuged at high speed using a microfuge for 15 minutes at 4° C. The supernatant was discarded and a fresh 1.2 ml aliquot of ethanol (EtOH) was added, vortexed for one minute, centrifuged for one minute, and the supernatant again discarded (this is repeated twice). A fresh 1.2 ml aliquot of EtOH was added, and this suspension (gold particles in EtOH) was stored at −20° C. for weeks. To coat particles with polyethylimine (PEI; Sigma #P3143), 250 μl of the washed gold particle/EtOH mix was centrifuged and the EtOH discarded. The particles were washed once in 100 μl ddH2O to remove residual ethanol, 250 μl of 0.25 mM PEI was added, followed by a pulse-sonication to suspend the particles and then the tube was plunged into a dry ice/EtOH bath to flash-freeze the suspension, which was then lyophilized overnight. At this point, dry, coated particles could be stored at −80° C. for at least 3 weeks. Before use, the particles were rinsed 3 times with 250 μl aliquots of 2.5 mM HEPES buffer, pH 7.1, with 1× pulse-sonication, and then a quick vortex before each centrifugation. The particles were then suspended in a final volume of 250 μl HEPES buffer. A 25 μl aliquot of the particles was added to fresh tubes before attaching DNA. To attach uncoated DNA, the particles were pulse-sonicated, then 1 μg of DNA (in 5 μl water) was added, followed by mixing by pipetting up and down a few times with a Pipetteman and incubated for 10 minutes. The particles were spun briefly (i.e. 10 seconds), the supernatant removed, and 60 μl EtOH added. The particles with PEI-precipitated DNA-1 were washed twice in 60 μl of EtOH. The particles were centrifuged, the supernatant discarded, and the particles were resuspended in 45 μl water. To attach the second DNA (DNA-2), precipitation using a water-soluble cationic lipid transfection reagent was used. The 45 μl of particles/DNA-1 suspension was briefly sonicated, and then 5 μl of 100 ng/μl of DNA-2 and 2.5 μl of the water-soluble cationic lipid transfection reagent were added. The solution was placed on a rotary shaker for 10 minutes, centrifuged at 10,000 g for 1 minute. The supernatant was removed, and the particles resuspended in 60 μl of EtOH. The solution was spotted onto macrocarriers and the gold particles onto which DNA-1 and DNA-2 had been sequentially attached were delivered into scutellar cells of 10 DAP Hi-II immature embryos using a standard protocol for the PDS-1000. For this experiment, the DNA-1 plasmid contained a UBI::RFP::pinII expression cassette, and DNA-2 contained a UBI::CFP::pinII expression cassette. Two days after bombardment, transient expression of both the CFP and RFP fluorescent markers was observed as numerous red & blue cells on the surface of the immature embryo. The embryos were then placed on non-selective culture medium and allowed to grow for 3 weeks before scoring for stable colonies. After this 3-week period, 10 multicellular, stably-expressing blue colonies were observed, in comparison to only one red colony. This demonstrated that PEI-precipitation could be used to effectively introduce DNA for transient expression while dramatically reducing integration of the PEI-introduced DNA and thus reducing the recovery of RFP-expressing transgenic events. In this manner, PEI-precipitation can be used to deliver transient expression of BBM and/or WUS2.

[0435] For example, the particles are first coated with UBI::BBM::pinII using PEI, then coated with UBI::moPAT~YFP using a water-soluble cationic lipid transfection reagent, and then bombarded into scutellar cells on the surface of immature embryos. PEI-mediated precipitation results in a high frequency of transiently expressing cells on the surface of the immature embryo and extremely low frequencies of recovery of stable transformants Thus, it is expected that the PEI-precipitated BBM cassette expresses transiently and stimulates a burst of embryogenic growth on the bombarded surface of the tissue (i.e. the scutellar surface), but this plasmid will not integrate. The PAT~GFP plasmid released from the Ca++/gold particles is expected to integrate and express the selectable marker at a frequency that results in substantially improved recovery of transgenic events. As a control treatment, PEI-precipitated particles containing a UBI::GUS::pinII (instead of BBM) are mixed with the PAT~GFP/Ca++ particles. Immature embryos from both treatments are moved onto culture medium containing 3 mg/l bialaphos. After 6-8 weeks, it is expected that GFP+, bialaphos-resistant calli will be observed in the PEI/BBM treatment at a much higher frequency relative to the control treatment (PEI/GUS).

[0436] As an alternative method, the BBM plasmid is precipitated onto gold particles with PEI, and then introduced into scutellar cells on the surface of immature embryos, and subsequent transient expression of the BBM gene elicits a rapid proliferation of embryogenic growth. During this period of induced growth, the explants are treated with *Agrobacterium* using standard methods for maize (see Example 1), with T-DNA delivery into the cell introducing a transgenic expression cassette such as UBI::moPAT~GFPm::pinII. After co-cultivation, explants are allowed to recover on normal culture medium, and then are moved onto culture medium containing 3 mg/l bialaphos. After 6-8 weeks, it is expected that GFP+, bialaphos-resistant calli will be observed in the PEI/BBM treatment at a much higher frequency relative to the control treatment (PEI/GUS).

[0437] It may be desirable to “kick start” callus growth by transiently expressing the BBM and/or WUS2 polynucleotide products. This can be done by delivering BBM and WUS2 5'-capped polyadenylated RNA, expression cassettes containing BBM and WUS2 DNA, or BBM and/or WUS2 proteins. All of these molecules can be delivered using a biolistics particle gun. For example 5'-capped polyadenylated BBM and/or WUS2 RNA can easily be made in vitro using Ambion’s mMessage mMachine kit. RNA is co-delivered along with DNA containing a polynucleotide of interest and a marker used for selection/screening such as Ubi::moPAT~GFPm::PinII. It is expected that the cells receiving the RNA will immediately begin dividing more rapidly and a large portion of these will have integrated the agronomic gene. These events can further be validated as being transgenic clonal colonies because they will also express the PAT~GFP fusion protein (and thus will display green fluorescence under appropriate illumination). Plants regenerated from these embryos can then be screened for the presence of the polynucleotide of interest.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 96

<210> SEQ ID NO 1
<211> LENGTH: 4107
<212> TYPE: DNA
<213> ORGANISM: Streptococcus pyogenes M1 GAS (SF370)

<400> SEQUENCE: 1

atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg 60

atcactgatg aatataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120

cacagtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcggaa 180

gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240

tatctacagg agatttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga 300

cttgaagagt ctttttggt ggaagaagac aagaagcatg aacgtcatcc tattttgga 360

aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420

aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcat 480

atgattaagt ttcgtggtca tttttgatt gagggagatt taaatcctga taatagtgat 540

gtggacaaac tatttatcca gttggtacaa acctacaatc aattatttga agaaaaccct 600

attaacgcaa gtggagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660

cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttgggaat 720

ctcattgctt tgtcattggg tttgacccct aattttaaat caaattttga tttggcagaa 780

gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840

caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt 900

ttactttcag atatcctaag agtaaatact gaaataacta aggctcccct atcagcttca 960

atgattaaac gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020

caacaacttc cagaaaagta taaagaaatc tttttgatc aatcaaaaaa cggatatgca 1080

ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140

gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200

aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat 1260

gctattttga gaagacaaga agacttttat ccattttaa aagacaatcg tgagaagatt 1320

gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380

cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa 1440

gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500

aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt 1560

tataacgaat tgacaaaggt caaatatgtt actgaaggaa tgcgaaaacc agcatttctt 1620

tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680

gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt 1740

tcaggagttg aagatagatt taatgcttca ttaggtacct accatgattt gctaaaaatt 1800

attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt 1860

ttaacattga ccttatttga agatagggag atgattgagg aaagacttaa aacatatgct 1920

cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980

-continued

```
cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta   2040

gatttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat   2100

agtttgacat ttaaagaaga cattcaaaaa gcacaagtgt ctggacaagg cgatagttta   2160

catgaacata ttgcaaattt agctggtagc cctgctatta aaaaaggtat tttacagact   2220

gtaaaagttg ttgatgaatt ggtcaaagta atggggcggc ataagccaga aaatatcgtt   2280

attgaaatgg cacgtgaaaa tcagacaact caaaagggcc agaaaaattc gcgagagcgt   2340

atgaaacgaa tcgaagaagg tatcaaagaa ttaggaagtc agattcttaa agagcatcct   2400

gttgaaaata ctcaattgca aaatgaaaag ctctatctct attatctcca aaatggaaga   2460

gacatgtatg tggaccaaga attagatatt aatcgtttaa gtgattatga tgtcgatcac   2520

attgttccac aaagtttcct taaagacgat tcaatagaca ataaggtctt aacgcgttct   2580

gataaaaatc gtggtaaatc ggataacgtt ccaagtgaag aagtagtcaa aaagatgaaa   2640

aactattgga gacaacttct aaacgccaag ttaatcactc aacgtaagtt tgataattta   2700

acgaaagctg aacgtggagg tttgagtgaa cttgataaag ctggttttat caaacgccaa   2760

ttggttgaaa ctcgccaaat cactaagcat gtggcacaaa ttttggatag tcgcatgaat   2820

actaaatacg atgaaaatga taaacttatt cgagaggtta aagtgattac cttaaaatct   2880

aaattagttt ctgacttccg aaaagatttc caattctata aagtacgtga gattaacaat   2940

taccatcatg cccatgatgc gtatctaaat gccgtcgttg gaactgcttt gattaagaaa   3000

tatccaaaac ttgaatcgga gtttgtctat ggtgattata aagtttatga tgttcgtaaa   3060

atgattgcta agtctgagca agaaataggc aaagcaaccg caaaatattt cttttactct   3120

aatatcatga acttcttcaa aacagaaatt acacttgcaa atggagagat tcgcaaacgc   3180

cctctaatcg aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagatttt   3240

gccacagtgc gcaaagtatt gtccatgccc caagtcaata ttgtcaagaa aacagaagta   3300

cagacaggcg gattctccaa ggagtcaatt ttaccaaaaa gaaattcgga caagcttatt   3360

gctcgtaaaa aagactggga tccaaaaaaa tatggtggtt ttgatagtcc aacggtagct   3420

tattcagtcc tagtggttgc taaggtggaa aaagggaaat cgaagaagtt aaaatccgtt   3480

aaagagttac tagggatcac aattatggaa agaagttcct ttgaaaaaaa tccgattgac   3540

tttttagaag ctaaaggata taaggaagtt aaaaaagact taatcattaa actacctaaa   3600

tatagtcttt ttgagttaga aaacggtcgt aaacggatgc tggctagtgc cggagaatta   3660

caaaaaggaa atgagctggc tctgccaagc aaatatgtga atttttttata tttagctagt   3720

cattatgaaa agttgaaggg tagtccagaa gataacgaac aaaaacaatt gtttgtggag   3780

cagcataagc attatttaga tgagattatt gagcaaatca gtgaattttc taagcgtgtt   3840

attttagcag atgccaattt agataaagtt cttagtgcat ataacaaaca tagagacaaa   3900

ccaatacgtg aacaagcaga aaatattatt catttattta cgttgacgaa tcttggagct   3960

cccgctgctt ttaaatattt tgatacaaca attgatcgta aacgatatac gtctacaaaa   4020

gaagttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt   4080

gatttgagtc agctaggagg tgactga                                       4107
```

<210> SEQ ID NO 2
<211> LENGTH: 189
<212> TYPE: DNA

-continued

<213> ORGANISM: Solanum tuberosum

<400> SEQUENCE: 2

gtaagtttct gcttctacct ttgatatata tataataatt atcattaatt agtagtaata 60

taatatttca aatatttttt tcaaaataaa agaatgtagt atatagcaat tgcttttctg 120

tagtttataa gtgtgtatat tttaatttat aacttttcta atatatgacc aaaacatggt 180

gatgtgcag 189

<210> SEQ ID NO 3
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Simian virus 40

<400> SEQUENCE: 3

Met Ala Pro Lys Lys Lys Arg Lys Val
1 5

<210> SEQ ID NO 4
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Agrobacterium tumefaciens

<400> SEQUENCE: 4

Lys Arg Pro Arg Asp Arg His Asp Gly Glu Leu Gly Gly Arg Lys Arg
1 5 10 15

Ala Arg

<210> SEQ ID NO 5
<211> LENGTH: 6717
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Maize optimized Cas9 expression cassette

<400> SEQUENCE: 5

gtgcagcgtg acccggtcgt gcccctctct agagataatg agcattgcat gtctaagtta 60

taaaaaatta ccacatattt tttttgtcac acttgtttga agtgcagttt atctatcttt 120

atacatatat ttaaacttta ctctacgaat aatataatct atagtactac aataatatca 180

gtgttttaga gaatcatata aatgaacagt tagacatggt ctaaaggaca attgagtatt 240

ttgacaacag gactctacag ttttatcttt ttagtgtgca tgtgttctcc tttttttttg 300

caaatagctt cacctatata atacttcatc cattttatta gtacatccat ttagggttta 360

gggttaatgg tttttataga ctaatttttt tagtacatct attttattct attttagcct 420

ctaaattaag aaaactaaaa ctctatttta gtttttttat ttaataattt agatataaaa 480

tagaataaaa taaagtgact aaaaattaaa caaatacct ttaagaaatt aaaaaaacta 540

aggaaacatt tttcttgttt cgagtagata atgccagcct gttaaacgcc gtcgacgagt 600

ctaacggaca ccaaccagcg aaccagcagc gtcgcgtcgg gccaagcgaa gcagacggca 660

cggcatctct gtcgctgcct ctggacccct ctcgagagtt ccgctccacc gttggacttg 720

ctccgctgtc ggcatccaga aattgcgtgg cggagcggca gacgtgagcc ggcacggcag 780

gcggcctcct cctcctctca cggcaccggc agctacgggg gattcctttc ccaccgctcc 840

ttcgctttcc cttcctcgcc cgccgtaata aatagacacc ccctccacac cctctttccc 900

-continued

caacctcgtg ttgttcggag cgcacacaca cacaaccaga tctcccccaa atccacccgt 960

cggcacctcc gcttcaaggt acgccgctcg tcctcccccc cccccctctc taccttctct 1020

agatcggcgt tccggtccat gcatggttag ggcccggtag ttctacttct gttcatgttt 1080

gtgttagatc cgtgtttgtg ttagatccgt gctgctagcg ttcgtacacg gatgcgacct 1140

gtacgtcaga cacgttctga ttgctaactt gccagtgttt ctctttgggg aatcctggga 1200

tggctctagc cgttccgcag acgggatcga tttcatgatt ttttttgttt cgttgcatag 1260

ggtttggttt gcccttttcc tttatttcaa tatatgccgt gcacttgttt gtcgggtcat 1320

cttttcatgc ttttttttgt cttggttgtg atgatgtggt ctggttgggc ggtcgttcta 1380

gatcggagta gaattctgtt tcaaactacc tggtggattt attaattttg gatctgtatg 1440

tgtgtgccat acatattcat agttacgaat tgaagatgat ggatggaaat atcgatctag 1500

gataggtata catgttgatg cgggttttac tgatgcatat acagagatgc tttttgttcg 1560

cttggttgtg atgatgtggt gtggttgggc ggtcgttcat tcgttctaga tcggagtaga 1620

atactgtttc aaactacctg gtgtatttat taattttgga actgtatgtg tgtgtcatac 1680

atcttcatag ttacgagttt aagatggatg gaaatatcga tctaggatag gtatacatgt 1740

tgatgtgggt tttactgatg catatacatg atggcatatg cagcatctat tcatatgctc 1800

taaccttgag tacctatcta ttataataaa caagtatgtt ttataattat tttgatcttg 1860

atatacttgg atgatggcat atgcagcagc tatatgtgga tttttttagc cctgccttca 1920

tacgctattt atttgcttgg tactgtttct tttgtcgatg ctcaccctgt tgtttggtgt 1980

tacttctgca ggtcgactct agaggatcca tggcaccgaa gaagaagcgc aaggtgatgg 2040

acaagaagta cagcatcggc ctcgacatcg gcaccaactc ggtgggctgg gccgtcatca 2100

cggacgaata taaggtcccg tcgaagaagt tcaaggtcct cggcaataca gaccgccaca 2160

gcatcaagaa aaacttgatc ggcgccctcc tgttcgatag cggcgagacc gcggaggcga 2220

ccaggctcaa gaggaccgcc aggagacggt acactaggcg caagaacagg atctgctacc 2280

tgcaggagat cttcagcaac gagatggcga aggtggacga ctccttcttc caccgcctgg 2340

aggaatcatt cctggtggag gaggacaaga agcatgagcg gcacccaatc ttcggcaaca 2400

tcgtcgacga ggtaagtttc tgcttctacc tttgatatat atataataat tatcattaat 2460

tagtagtaat ataatatttc aaatattttt ttcaaaataa aagaatgtag tatatagcaa 2520

ttgcttttct gtagtttata agtgtgtata ttttaattta taacttttct aatatatgac 2580

caaaacatgg tgatgtgcag gtggcctacc acgagaagta cccgacaatc taccacctcc 2640

ggaagaaact ggtggacagc acagacaagg cggacctccg gctcatctac cttgccctcg 2700

cgcatatgat caagttccgc ggccacttcc tcatcgaggg cgacctgaac ccggacaact 2760

ccgacgtgga caagctgttc atccagctcg tgcagacgta caatcaactg ttcgaggaga 2820

accccataaa cgctagcggc gtggacgcca aggccatcct ctcggccagg ctctcgaaat 2880

caagaaggct ggagaacctt atcgcgcagt tgccaggcga aaagaagaac ggcctcttcg 2940

gcaaccttat tgcgctcagc ctcggcctga cgccgaactt caaatcaaac ttcgacctcg 3000

cggaggacgc caagctccag ctctcaaagg acacctacga cgacgacctc gacaacctcc 3060

tggcccagat aggagaccag tacgcggacc tcttcctcgc cgccaagaac ctctccgacg 3120

ctatcctgct cagcgacatc cttcgggtca acaccgaaat taccaaggca ccgctgtccg 3180

-continued

ccagcatgat taaacgctac gacgagcacc atcaggacct cacgctgctc aaggcactcg 3240

tccgccagca gctccccgag aagtacaagg agatcttctt cgaccaatca aaaaacggct 3300

acgcgggata tatcgacggc ggtgccagcc aggaagagtt ctacaagttc atcaaaccaa 3360

tcctggagaa gatggacggc accgaggagt tgctggtcaa gctcaacagg gaggacctcc 3420

tcaggaagca gaggaccttc gacaacggct ccatcccgca tcagatccac ctgggcgaac 3480

tgcatgccat cctgcggcgc caggaggact tctacccgtt cctgaaggat aaccgggaga 3540

agatcgagaa gatcttgacg ttccgcatcc catactacgt gggcccgctg gctcgcggca 3600

actcccggtt cgcctggatg acccggaagt cggaggagac catcacaccc tggaactttg 3660

aggaggtggt cgataagggc gctagcgctc agagcttcat cgagcgcatg accaacttcg 3720

ataaaaacct gcccaatgaa aaagtcctcc ccaagcactc gctgctctac gagtacttca 3780

ccgtgtacaa cgagctcacc aaggtcaaat acgtcaccga gggcatgcgg aagccggcgt 3840

tcctgagcgg cgagcagaag aaggcgatag tggacctcct cttcaagacc aacaggaagg 3900

tgaccgtgaa gcaattaaaa gaggactact tcaagaaaat agagtgcttc gactccgtgg 3960

agatctcggg cgtggaggat cggttcaacg cctcactcgg cacgtatcac gacctcctca 4020

agatcattaa agacaaggac ttcctcgaca acgaggagaa cgaggacatc ctcgaggaca 4080

tcgtcctcac cctgaccctg ttcgaggacc gcgaaatgat cgaggagagg ctgaagacct 4140

acgcgcacct gttcgacgac aaggtcatga aacagctcaa gaggcgccgc tacactggtt 4200

ggggaaggct gtcccgcaag ctcattaatg gcatcaggga caagcagagc ggcaagacca 4260

tcctggactt cctcaagtcc gacgggttcg ccaaccgcaa cttcatgcag ctcattcacg 4320

acgactcgct cacgttcaag gaagacatcc agaaggcaca ggtgagcggg cagggtgact 4380

ccctccacga acacatcgcc aacctggccg gctcgccggc cattaaaaag ggcatcctgc 4440

agacggtcaa ggtcgtcgac gagctcgtga aggtgatggg ccggcacaag cccgaaaata 4500

tcgtcataga gatggccagg gagaaccaga ccacccaaaa agggcagaag aactcgcgcg 4560

agcggatgaa acggatcgag gagggcatta aagagctcgg gtcccagatc ctgaaggagc 4620

accccgtgga aaatacccag ctccagaatg aaaagctcta cctctactac ctgcagaacg 4680

gccgcgacat gtacgtggac caggagctgg acattaatcg gctatcggac tacgacgtcg 4740

accacatcgt gccgcagtcg ttcctcaagg acgatagcat cgacaacaag gtgctcaccc 4800

ggtcggataa aaatcggggc aagagcgaca acgtgcccag cgaggaggtc gtgaagaaga 4860

tgaaaaacta ctggcgccag ctcctcaacg cgaaactgat cacccagcgc aagttcgaca 4920

acctgacgaa ggcggaacgc ggtggcttga gcgaactcga taaggcgggc ttcataaaaa 4980

ggcagctggt cgagacgcgc cagatcacga agcatgtcgc ccagatcctg gacagccgca 5040

tgaatactaa gtacgatgaa aacgacaagc tgatccggga ggtgaaggtg atcacgctga 5100

agtccaagct cgtgtcggac ttccgcaagg acttccagtt ctacaaggtc cgcgagatca 5160

acaactacca ccacgcccac gacgcctacc tgaatgcggt ggtcgggacc gccctgatca 5220

agaagtaccc gaagctggag tcggagttcg tgtacggcga ctacaaggtc tacgacgtgc 5280

gcaaaatgat cgccaagtcc gagcaggaga tcggcaaggc cacggcaaaa tacttcttct 5340

actcgaacat catgaacttc ttcaagaccg agatcaccct cgcgaacggc gagatccgca 5400

agcgcccgct catcgaaacc aacggcgaga cgggcgagat cgtctgggat aagggccggg 5460

-continued

atttcgcgac ggtccgcaag gtgctctcca tgccgcaagt caatatcgtg aaaaagacgg 5520

aggtccagac gggcgggttc agcaaggagt ccatcctccc gaagcgcaac tccgacaagc 5580

tcatcgcgag gaagaaggat tgggacccga aaaaatatgg cggcttcgac agcccgaccg 5640

tcgcatacag cgtcctcgtc gtggcgaagg tggagaaggg caagtcaaag aagctcaagt 5700

ccgtgaagga gctgctcggg atcacgatta tggagcggtc ctccttcgag aagaacccga 5760

tcgacttcct agaggccaag ggatataagg aggtcaagaa ggacctgatt attaaactgc 5820

cgaagtactc gctcttcgag ctggaaaacg gccgcaagag gatgctcgcc tccgcaggcg 5880

agttgcagaa gggcaacgag ctcgccctcc cgagcaaata cgtcaatttc ctgtacctcg 5940

ctagccacta tgaaaagctc aagggcagcc cggaggacaa cgagcagaag cagctcttcg 6000

tggagcagca caagcattac ctggacgaga tcatcgagca gatcagcgag ttctcgaagc 6060

gggtgatcct cgccgacgcg aacctggaca aggtgctgtc ggcatataac aagcaccgcg 6120

acaaaccaat acgcgagcag gccgaaaata tcatccacct cttcaccctc accaacctcg 6180

gcgctccggc agccttcaag tacttcgaca ccacgattga ccggaagcgg tacacgagca 6240

cgaaggaggt gctcgatgcg acgctgatcc accagagcat cacagggctc tatgaaacac 6300

gcatcgacct gagccagctg ggcggagaca agagaccacg ggaccgccac gatggcgagc 6360

tgggaggccg caagcgggca aggtaggtac cgttaaccta gacttgtcca tcttctggat 6420

tggccaactt aattaatgta tgaaataaaa ggatgcacac atagtgacat gctaatcact 6480

ataatgtggg catcaaagtt gtgtgttatg tgtaattact agttatctga ataaaagaga 6540

aagagatcat ccatatttct tatcctaaat gaatgtcacg tgtctttata attctttgat 6600

gaaccagatg catttcatta accaaatcca tatacatata aatattaatc atatataatt 6660

aatatcaatt gggttagcaa aacaaatcta gtctaggtgt gttttgcgaa tgcggcc 6717

<210> SEQ ID NO 6
<211> LENGTH: 1000
<212> TYPE: DNA
<213> ORGANISM: Zea mays

<400> SEQUENCE: 6

tgagagtaca atgatgaacc tagattaatc aatgccaaag tctgaaaaat gcaccctcag 60

tctatgatcc agaaaatcaa gattgcttga ggccctgttc ggttgttccg gattagagcc 120

ccggattaat tcctagccgg attacttctc taatttatat agattttgat gagctggaat 180

gaatcctggc ttattccggt acaaccgaac aggccctgaa ggataccagt aatcgctgag 240

ctaaattggc atgctgtcag agtgtcagta ttgcagcaag gtagtgagat aaccggcatc 300

atggtgccag tttgatggca ccattagggt tagagatggt ggccatgggc gcatgtcctg 360

gccaactttg tatgatatat ggcagggtga ataggaaagt aaaattgtat tgtaaaaagg 420

gatttcttct gtttgttagc gcatgtacaa ggaatgcaag ttttgagcga gggggcatca 480

aagatctggc tgtgtttcca gctgtttttg ttagccccat cgaatccttg acataatgat 540

cccgcttaaa taagcaacct cgcttgtata gttccttgtg ctctaacaca cgatgatgat 600

aagtcgtaaa atagtggtgt ccaaagaatt tccaggccca gttgtaaaag ctaaaatgct 660

attcgaattt ctactagcag taagtcgtgt ttagaaatta tttttttata taccttttt 720

ccttctatgt acagtaggac acagtgtcag cgccgcgttg acggagaata tttgcaaaaa 780

-continued

agtaaaagag aaagtcatag cggcgtatgt gccaaaaact tcgtcacaga gagggccata 840

agaaacatgg cccacggccc aatacgaagc accgcgacga agcccaaaca gcagtccgta 900

ggtggagcaa agcgctgggt aatacgcaaa cgttttgtcc caccttgact aatcacaaga 960

gtggagcgta ccttataaac cgagccgcaa gcaccgaatt 1000

<210> SEQ ID NO 7
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: SV40 (Simian virus 40) Nuclear localization signal

<400> SEQUENCE: 7

Ser Arg Ala Asp Pro Lys Lys Lys Arg Lys Val
1 5 10

<210> SEQ ID NO 8
<211> LENGTH: 1047
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Maize optimized crRNA expression cassette containing the LIGCas-3 target sequence in the variable targeting domain

<400> SEQUENCE: 8

tgagagtaca atgatgaacc tagattaatc aatgccaaag tctgaaaaat gcaccctcag 60

tctatgatcc agaaaatcaa gattgcttga ggccctgttc ggttgttccg gattagagcc 120

ccggattaat tcctagccgg attacttctc taatttatat agattttgat gagctggaat 180

gaatcctggc ttattccggt acaaccgaac aggccctgaa ggataccagt aatcgctgag 240

ctaaattggc atgctgtcag agtgtcagta ttgcagcaag gtagtgagat aaccggcatc 300

atggtgccag tttgatggca ccattagggt tagagatggt ggccatgggc gcatgtcctg 360

gccaactttg tatgatatat ggcagggtga ataggaaagt aaaattgtat tgtaaaaagg 420

gatttcttct gtttgttagc gcatgtacaa ggaatgcaag ttttgagcga gggggcatca 480

aagatctggc tgtgtttcca gctgtttttg ttagccccat cgaatccttg acataatgat 540

cccgcttaaa taagcaacct cgcttgtata gttccttgtg ctctaacaca cgatgatgat 600

aagtcgtaaa atagtggtgt ccaaagaatt tccaggccca gttgtaaaag ctaaaatgct 660

attcgaattt ctactagcag taagtcgtgt ttagaaatta tttttttata tacctttttt 720

ccttctatgt acagtaggac acagtgtcag cgcccgcgttg acggagaata tttgcaaaaa 780

agtaaaagag aaagtcatag cggcgtatgt gccaaaaact tcgtcacaga gagggccata 840

agaaacatgg cccacggccc aatacgaagc accgcgacga agcccaaaca gcagtccgta 900

ggtggagcaa agcgctgggt aatacgcaaa cgttttgtcc caccttgact aatcacaaga 960

gtggagcgta ccttataaac cgagccgcaa gcaccgaatt gcgtacgcgt acgtgtggtt 1020

ttagagctat gctgttttgt ttttttt 1047

<210> SEQ ID NO 9
<211> LENGTH: 1087
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Maize optimized tracrRNA expression cassette

<400> SEQUENCE: 9

tgagagtaca atgatgaacc tagattaatc aatgccaaag tctgaaaaat gcaccctcag 60

tctatgatcc agaaaatcaa gattgcttga ggccctgttc ggttgttccg gattagagcc 120

ccggattaat tcctagccgg attacttctc taatttatat agattttgat gagctggaat 180

gaatcctggc ttattccggt acaaccgaac aggccctgaa ggataccagt aatcgctgag 240

ctaaattggc atgctgtcag agtgtcagta ttgcagcaag gtagtgagat aaccggcatc 300

atggtgccag tttgatggca ccattagggt tagagatggt ggccatgggc gcatgtcctg 360

gccaactttg tatgatatat ggcagggtga ataggaaagt aaaattgtat tgtaaaaagg 420

gatttcttct gtttgttagc gcatgtacaa ggaatgcaag ttttgagcga gggggcatca 480

aagatctggc tgtgtttcca gctgtttttg ttagccccat cgaatccttg acataatgat 540

cccgcttaaa taagcaacct cgcttgtata gttccttgtg ctctaacaca cgatgatgat 600

aagtcgtaaa atagtggtgt ccaaagaatt tccaggccca gttgtaaaag ctaaaatgct 660

attcgaattt ctactagcag taagtcgtgt ttagaaatta tttttttata tacctttttt 720

ccttctatgt acagtaggac acagtgtcag cgccgcgttg acggagaata tttgcaaaaa 780

agtaaaagag aaagtcatag cggcgtatgt gccaaaaact tcgtcacaga gagggccata 840

agaaacatgg cccacggccc aatacgaagc accgcgacga agcccaaaca gcagtccgta 900

ggtggagcaa agcgctgggt aatacgcaaa cgttttgtcc caccttgact aatcacaaga 960

gtggagcgta ccttataaac cgagccgcaa gcaccgaatt ggaaccattc aaaacagcat 1020

agcaagttaa aataaggcta gtccgttatc aacttgaaaa agtggcaccg agtcggtgct 1080

ttttttt 1087

<210> SEQ ID NO 10
<211> LENGTH: 39
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: crRNA containing a variable targeting domain
targeting the LIGCas-3 target sequence

<400> SEQUENCE: 10

gcguacgcgu acguguggúu uuagagcuau gcuguuuug 39

<210> SEQ ID NO 11
<211> LENGTH: 86
<212> TYPE: RNA
<213> ORGANISM: Streptococcus pyogenes
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(86)
<223> OTHER INFORMATION: tracrRNA

<400> SEQUENCE: 11

ggaaccauuc aaaacagcau agcaaguuaa aauaaggcua guccguuauc aacuugaaaa 60

aguggcaccg agucggugcu uuuuuu 86

<210> SEQ ID NO 12
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Zea mays

-continued

```
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(24)
<223> OTHER INFORMATION: Maize genomic target site LIGCas-1 plus PAM
      sequence

<400> SEQUENCE: 12

gtaccgtacg tgccccggcg gagg                                            24


<210> SEQ ID NO 13
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(24)
<223> OTHER INFORMATION: Maize genomic target site LIGCas-2 plus PAM
      sequence

<400> SEQUENCE: 13

ggaattgtac cgtacgtgcc ccgg                                            24


<210> SEQ ID NO 14
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(20)
<223> OTHER INFORMATION: Maize genomic target site LIGCas-3 plus PAM
      sequence

<400> SEQUENCE: 14

gcgtacgcgt acgtgtgagg                                                 20


<210> SEQ ID NO 15
<211> LENGTH: 63
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LIGCas-1 forward primer for primary PCR
      (crRNA-tracrRNA system)

<400> SEQUENCE: 15

ctacactctt tccctacacg acgctcttcc gatcttcctc tgtaacgatt tacgcacctg     60

ctg                                                                   63


<210> SEQ ID NO 16
<211> LENGTH: 58
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LIGCas-1 and LIGCas-2 reverse primer for
      primary PCR (crRNA-tracrRNA system)

<400> SEQUENCE: 16

caagcagaag acggcatacg agctcttccg atctgcaaat gagtagcagc gcacgtat       58


<210> SEQ ID NO 17
<211> LENGTH: 63
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LIGCas-2 forward primer for primary PCR
      (crRNA-tracrRNA system)
```

-continued

```
<400> SEQUENCE: 17

ctacactctt tccctacacg acgctcttcc gatctgaagc tgtaacgatt tacgcacctg     60

ctg                                                                   63


<210> SEQ ID NO 18
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LIGCas-3 forward primer for primary PCR
      (crRNA-tracrRNA system)

<400> SEQUENCE: 18

ctacactctt tccctacacg acgctcttcc gatctaaggc gcaaatgagt agcagcgcac     60


<210> SEQ ID NO 19
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LIGCas-3 reverse primer for primary PCR
      (crRNA-tracrRNA and Long guide RNA system)

<400> SEQUENCE: 19

caagcagaag acggcatacg agctcttccg atctcacctg ctgggaattg taccgta        57


<210> SEQ ID NO 20
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: LIGCas-3 forward primer for primary PCR
      (Long guide RNA system)

<400> SEQUENCE: 20

ctacactctt tccctacacg acgctcttcc gatctttccc gcaaatgagt agcagcgcac     60


<210> SEQ ID NO 21
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Forward primer for secondary PCR

<400> SEQUENCE: 21

aatgatacgg cgaccaccga gatctacact ctttccctac acg                       43


<210> SEQ ID NO 22
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Reverse primer for secondary PCR

<400> SEQUENCE: 22

caagcagaag acggcata                                                   18


<210> SEQ ID NO 23
<211> LENGTH: 93
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
```

-continued

<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(93)
<223> OTHER INFORMATION: Unmodified reference sequence for LIGCas-1
and LIGCas-2 locus

<400> SEQUENCE: 23

ctgtaacgat ttacgcacct gctgggaatt gtaccgtacg tgccccggcg gaggatatat 60

atacctcaca cgtacgcgta cgcgtatata tac 93

<210> SEQ ID NO 24
<211> LENGTH: 98
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(98)
<223> OTHER INFORMATION: Mutation 1 for LIGCas-1 locus

<400> SEQUENCE: 24

tcctctgtaa cgatttacgc acctgctggg aattgtaccg tacgtgcccc ggtcggagga 60

tatatatacc tcacacgtac gcgtacgcgt atatatac 98

<210> SEQ ID NO 25
<211> LENGTH: 98
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(98)
<223> OTHER INFORMATION: Mutation 2 for LIGCas-1

<400> SEQUENCE: 25

tcctctgtaa cgatttacgc acctgctggg aattgtaccg tacgtgcccc ggacggagga 60

tatatatacc tcacacgtac gcgtacgcgt atatatac 98

<210> SEQ ID NO 26
<211> LENGTH: 98
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(98)
<223> OTHER INFORMATION: Mutation 3 for LIGCas-1

<400> SEQUENCE: 26

tcctctgtaa cgatttacgc acctgctggg aattgtaccg tacgtgcccc gggcggagga 60

tatatatacc tcacacgtac gcgtacgcgt atatatac 98

<210> SEQ ID NO 27
<211> LENGTH: 98
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(98)
<223> OTHER INFORMATION: Mutation 4 for LIGCas-1

<400> SEQUENCE: 27

tcctctgtaa cgatttacgc acctgctggg aattgtaccg tacgtgcccc ggccggagga 60

tatatatacc tcacacgtac gcgtacgcgt atatatac 98

<210> SEQ ID NO 28
<211> LENGTH: 99
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(99)
<223> OTHER INFORMATION: Mutation 5 for LIGCas-1

<400> SEQUENCE: 28

tcctctgtaa cgatttacgc acctgctggg aattgtaccg tacgtgcccc ggatcggagg   60

atatatatac ctcacacgta cgcgtacgcg tatatatac   99

<210> SEQ ID NO 29
<211> LENGTH: 94
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(94)
<223> OTHER INFORMATION: Mutation 6 for LIGCas-1

<400> SEQUENCE: 29

tcctctgtaa cgatttacgc acctgctggg aattgtaccg tacgtgcccc ggaggatata   60

tatacctcac acgtacgcgt acgcgtatat atac   94

<210> SEQ ID NO 30
<211> LENGTH: 81
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(81)
<223> OTHER INFORMATION: Mutation 7 for LIGCas-1

<400> SEQUENCE: 30

tcctctgtaa cgatttacgc acctgctggg aattgtaccg tacgtgcccc ggttcacacg   60

tacgcgtacg cgtatatata c   81

<210> SEQ ID NO 31
<211> LENGTH: 65
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(65)
<223> OTHER INFORMATION: Mutation 8 for LIGCas-1

<400> SEQUENCE: 31

tcctctgtaa cgatttacgc acctgctggg aattgtaccg tacgtacgcg tacgcgtata   60

tatac   65

<210> SEQ ID NO 32
<211> LENGTH: 99
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(99)
<223> OTHER INFORMATION: Mutation 9 for LIGCas-1

<400> SEQUENCE: 32

tcctctgtaa cgatttacgc acctgctggg aattgtaccg tacgtgcccc ggttcggagg   60

-continued

atatatatac ctcacacgta cgcgtacgcg tatatatac 99

<210> SEQ ID NO 33
<211> LENGTH: 95
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(95)
<223> OTHER INFORMATION: Mutation 10 for LIGCas-1

<400> SEQUENCE: 33

tcctctgtaa cgatttacgc acctgctggg aattgtaccg tacgtgcccc cggaggatat 60

atatacctca cacgtacgcg tacgcgtata tatac 95

<210> SEQ ID NO 34
<211> LENGTH: 98
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(98)
<223> OTHER INFORMATION: Mutation 1 for LIGCas-2

<400> SEQUENCE: 34

gaagctgtaa cgatttacgc acctgctggg aattgtaccg tacgtgaccc cggcggagga 60

tatatatacc tcacacgtac gcgtacgcgt atatatac 98

<210> SEQ ID NO 35
<211> LENGTH: 98
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(98)
<223> OTHER INFORMATION: Mutation 2 for LIGCas-2

<400> SEQUENCE: 35

gaagctgtaa cgatttacgc acctgctggg aattgtaccg tacgtgtccc cggcggagga 60

tatatatacc tcacacgtac gcgtacgcgt atatatac 98

<210> SEQ ID NO 36
<211> LENGTH: 96
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(96)
<223> OTHER INFORMATION: Mutation 3 for LIGCas-2

<400> SEQUENCE: 36

gaagctgtaa cgatttacgc acctgctggg aattgtaccg tacgtccccg gcggaggata 60

tatatacctc acacgtacgc gtacgcgtat atatac 96

<210> SEQ ID NO 37
<211> LENGTH: 98
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(98)
<223> OTHER INFORMATION: Mutation 4 for LIGCas-2

-continued

<400> SEQUENCE: 37

gaagctgtaa cgatttacgc acctgctggg aattgtaccg tacgtggccc cggcggagga   60

tatatatacc tcacacgtac gcgtacgcgt atatatac   98

<210> SEQ ID NO 38
<211> LENGTH: 99
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(98)
<223> OTHER INFORMATION: Mutation 5 for LIGCas-2

<400> SEQUENCE: 38

gaagctgtaa cgatttacgc acctgctggg aattgtaccg tacgtgcacc ccggcggagg   60

atatatatac ctcacacgta cgcgtacgcg tatatatac   99

<210> SEQ ID NO 39
<211> LENGTH: 87
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(87)
<223> OTHER INFORMATION: Mutation 6 for LIGCas-2

<400> SEQUENCE: 39

gaagctgtaa cgatttacgc acctgctggg aattgtaccc ggcggaggat atatatacct   60

cacacgtacg cgtacgcgta tatatac   87

<210> SEQ ID NO 40
<211> LENGTH: 92
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(92)
<223> OTHER INFORMATION: Mutation 7 for LIGCas-2

<400> SEQUENCE: 40

gaagctgtaa cgatttacgc acctgctggg aattgtaccg tccccggcgg aggatatata   60

tacctcacac gtacgcgtac gcgtatatat ac   92

<210> SEQ ID NO 41
<211> LENGTH: 94
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(94)
<223> OTHER INFORMATION: Mutation 8 for LIGCas-2

<400> SEQUENCE: 41

gaagctgtaa cgatttacgc acctgctggg aattgtaccg tacccccggc ggaggatata   60

tatacctcac acgtacgcgt acgcgtatat atac   94

<210> SEQ ID NO 42
<211> LENGTH: 95
<212> TYPE: DNA
<213> ORGANISM: Zea mays

-continued

```
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(95)
<223> OTHER INFORMATION: Mutation 9 for LIGCas-2

<400> SEQUENCE: 42

gaagctgtaa cgatttacgc acctgctggg aattgtaccg tacgccccgg cggaggatat     60

atatacctca cacgtacgcg tacgcgtata tatac                                95


<210> SEQ ID NO 43
<211> LENGTH: 88
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(88)
<223> OTHER INFORMATION: Mutation 10 for LIGCas-2

<400> SEQUENCE: 43

gaagctgtaa cgatttacgc acctgctggg aattgtaccc cggcggagga tatatatacc     60

tcacacgtac gcgtacgcgt atatatac                                        88


<210> SEQ ID NO 44
<211> LENGTH: 93
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(93)
<223> OTHER INFORMATION: Unmodified reference sequence for LIGCas-3
      locus

<400> SEQUENCE: 44

cgcaaatgag tagcagcgca cgtatatata cgcgtacgcg tacgtgtgag gtatatatat     60

cctccgccgg ggcacgtacg gtacaattcc cag                                  93


<210> SEQ ID NO 45
<211> LENGTH: 98
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(98)
<223> OTHER INFORMATION: Mutation 1 for LIGCas-3 locus

<400> SEQUENCE: 45

aaggcgcaaa tgagtagcag cgcacgtata tatacgcgta cgcgtacgtt gtgaggtata     60

tatatcctcc gccggggcac gtacggtaca attcccag                             98


<210> SEQ ID NO 46
<211> LENGTH: 96
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(96)
<223> OTHER INFORMATION: Mutation 2 for LIGCas-3 locus

<400> SEQUENCE: 46

aaggcgcaaa tgagtagcag cgcacgtata tatacgcgta cgcgtacggt gaggtatata     60

tatcctccgc cggggcacgt acggtacaat tcccag                               96
```

```
<210> SEQ ID NO 47
<211> LENGTH: 95
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(95)
<223> OTHER INFORMATION: Mutation 3 for LIGCas-3 locus

<400> SEQUENCE: 47

aaggcgcaaa tgagtagcag cgcacgtata tatacgcgta cgcgtacgtg aggtatatat     60

atcctccgcc ggggcacgta cggtacaatt cccag                                95


<210> SEQ ID NO 48
<211> LENGTH: 96
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(96)
<223> OTHER INFORMATION: Mutation 4 for LIGCas-3 locus

<400> SEQUENCE: 48

aaggcgcaaa tgagtagcag cgcacgtata tatacgcgta cgcgtactgt gaggtatata     60

tatcctccgc cggggcacgt acggtacaat tcccag                               96


<210> SEQ ID NO 49
<211> LENGTH: 68
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(68)
<223> OTHER INFORMATION: Mutation 5 for LIGCas-3 locus

<400> SEQUENCE: 49

aaggcgcaaa tgagtagcag cgcacgtata tatatcctcc gccggggcac gtacggtaca     60

attcccag                                                              68


<210> SEQ ID NO 50
<211> LENGTH: 93
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(93)
<223> OTHER INFORMATION: Mutation 6 for LIGCas-3 locus

<400> SEQUENCE: 50

aaggcgcaaa tgagtagcag cgcacgtata tatacgcgta cgcgtgtgag gtatatatat     60

cctccgccgg ggcacgtacg gtacaattcc cag                                  93


<210> SEQ ID NO 51
<211> LENGTH: 89
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(89)
<223> OTHER INFORMATION: Mutation 7 for LIGCas-3 locus

<400> SEQUENCE: 51
```

aaggcgcaaa tgagtagcag cgcacgtata tatacgcgta cgtgaggtat atatatcctc 60

cgccggggca cgtacggtac aattcccag 89

<210> SEQ ID NO 52
<211> LENGTH: 89
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(89)
<223> OTHER INFORMATION: Mutation 8 for LIGCas-3 locus

<400> SEQUENCE: 52

aaggcgcaaa tgagtagcag cgcacgtata tatacgcgta cgcgtactat atatatcctc 60

cgccggggca cgtacggtac aattcccag 89

<210> SEQ ID NO 53
<211> LENGTH: 94
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(94)
<223> OTHER INFORMATION: Mutation 9 for LIGCas-3 locus

<400> SEQUENCE: 53

aaggcgcaaa tgagtagcag cgcacgtata tatacgcgta cgcgtacgga ggtatatata 60

tcctccgccg gggcacgtac ggtacaattc ccag 94

<210> SEQ ID NO 54
<211> LENGTH: 96
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(96)
<223> OTHER INFORMATION: Mutation 10 for LIGCas-3 locus

<400> SEQUENCE: 54

aaggcgcaaa tgagtagcag cgcacgtata tatacgcgta cgcgtacgat gaggtatata 60

tatcctccgc cggggcacgt acggtacaat tcccag 96

<210> SEQ ID NO 55
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(39)
<223> OTHER INFORMATION: crDNA sequence comprised of deoxyribonucleic
      acids (crDNA) targeting

<400> SEQUENCE: 55

gcgtacgcgt acgtgtggtt ttagagctat gctgttttg 39

<210> SEQ ID NO 56
<211> LENGTH: 98
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(98)
<223> OTHER INFORMATION: Mutation 1 for LIGCas-3 locus (synthetic

-continued

crRNA plus tracrRNA and Cas9

<400> SEQUENCE: 56

ggaacgcaaa tgagtagcag cgcacgtata tatacgcgta cgcgtacgtt gtgaggtata 60

tatatcctcc gccggggcac gtacggtaca attcccag 98

<210> SEQ ID NO 57
<211> LENGTH: 94
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(94)
<223> OTHER INFORMATION: Mutation 2 for LIGCas-3 locus (synthetic crRNA
plus tracrRNA and Cas9

<400> SEQUENCE: 57

ggaacgcaaa tgagtagcag cgcacgtata tatacgcgta cgcgtagtga ggtatatata 60

tcctccgccg gggcacgtac ggtacaattc ccag 94

<210> SEQ ID NO 58
<211> LENGTH: 96
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(96)
<223> OTHER INFORMATION: Mutation 3 for LIGCas-3 locus (synthetic crRNA
plus tracrRNA and Cas9

<400> SEQUENCE: 58

ggaacgcaaa tgagtagcag cgcacgtata tatacgcgta cgcgtactgt gaggtatata 60

tatcctccgc cggggcacgt acggtacaat tcccag 96

<210> SEQ ID NO 59
<211> LENGTH: 98
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(98)
<223> OTHER INFORMATION: Mutation 1 for LIGCas-3 locus (synthetic crDNA
plus tracrRNA and Cas9

<400> SEQUENCE: 59

ccttcgcaaa tgagtagcag cgcacgtata tatacgcgta cgcgtacgtt gtgaggtata 60

tatatcctcc gccggggcac gtacggtaca attcccag 98

<210> SEQ ID NO 60
<211> LENGTH: 66
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(66)
<223> OTHER INFORMATION: Mutation 2 for LIGCas-3 locus (synthetic crDNA
plus tracrRNA and Cas9

<400> SEQUENCE: 60

ggaacgcaaa tgagtagcag cgcacgtata tatactccgc cggggcacgt acggtacaat 60

tcccag 66

```
<210> SEQ ID NO 61
<211> LENGTH: 85
<212> TYPE: DNA
<213> ORGANISM: Zea mays
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(85)
<223> OTHER INFORMATION: Mutation 3 for LIGCas-3 locus (synthetic crDNA
      plus tracrRNA and Cas9

<400> SEQUENCE: 61

ggaacgcaaa tgagtagcag cgcacgtata tatacgcgta cggtatatat atcctccgcc     60

ggggcacgta cggtacaatt cccag                                           85


<210> SEQ ID NO 62
<211> LENGTH: 17
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of unmodified crRNA VT domain

<400> SEQUENCE: 62

gcguacgcgu acgugug                                                    17


<210> SEQ ID NO 63
<211> LENGTH: 22
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of unmodified crRNA CER domain

<400> SEQUENCE: 63

guuuuagagc uaugcuguuu ug                                              22


<210> SEQ ID NO 64
<211> LENGTH: 17
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crRNA VT domain modified with
      Phosphorothioate bonds at 5 prime end
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: The first N at the 5 prime end represents a G
      ribonucleotide with a phosphorothioate bond, the second  N
      represents a C ribonucleotide with a phosphorothioate bond and
      third N represents a G ribonucleotide with a Phosphorothioate
      bond.

<400> SEQUENCE: 64

nnnuacgcgu acgugug                                                    17


<210> SEQ ID NO 65
<211> LENGTH: 22
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crRNA CER domain modified with
      Phosphorothioate bonds near 3 prime end
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(22)
<223> OTHER INFORMATION: the Ns at the nineteenth, twentieth and
      twenty-first position represent U ribonucleotides with
      phosphorothioate bonds.
```

-continued

```
<400> SEQUENCE: 65

guuuuagagc uaugcugunn ng                                              22


<210> SEQ ID NO 66
<211> LENGTH: 17
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crRNA VT domain modified with
      2-O-Methyl RNA bases at 5 prime end
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: the first N at the 5 prime end represents a G
      2-O-methyl ribonucleotide, the second N represents a C 2-O-methyl
      ribonucleotide and the third N represents a G 2-O-Methyl
      ribonucleotide.
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(3)
<223> OTHER INFORMATION: n is a, c, g, or u

<400> SEQUENCE: 66

nnnuacgcgu acgugug                                                    17


<210> SEQ ID NO 67
<211> LENGTH: 22
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crRNA CER domain modified with
      2-O-Methyl RNA bases at 3 prime end
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(22)
<223> OTHER INFORMATION: the N at the twentieth position represents a U
      2-O-Methyl ribonucleotide, the N at the twenty-first position
      represents a U 2-O-Methyl ribonucleotide and the N at the twenty-
      second position represents a G 2-O-Methyl ribonucleotide.

<400> SEQUENCE: 67

guuuuagagc uaugcuguun nn                                              22


<210> SEQ ID NO 68
<211> LENGTH: 17
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crRNA VT domain modified with
      2-O-Methyl RNA bases throughout
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: The first N at the 5 prime end represents a G
      2-O-Methyl ribonucleotide, a N at the second position represents
      a C 2-O-Methyl ribonucleotide, a N at the third position
      represents a G 2-O-Methyl ribonucleotide,
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: a N at the fourth position represents an U
      2-O-Methyl ribonucleotide, a N at the fifth position represents
      an A 2-O-Methyl ribonucleotide, a N at the sixth position
      represents a C 2-O-Methyl ribonucleotide,
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: a N at the seventh position represents a G
      2-O-Methyl ribonucleotide, a N at the eighth position represents
      a C 2'O-Methyl ribonucleotide, a N at the ninth position
```

-continued

```
      represents a G 2-O-Methyl ribonucleotide
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: a N at the 10th, 11th, 12th, 13th, 14th, 15th,
      16th, 17th, position represents an U-, A-, C-, G-,U-, G-,U-, G-
      2-O-Methyl ribonucleotide, respectively.

<400> SEQUENCE: 68

nnnnnnnnnn nnnnnnn                                                   17


<210> SEQ ID NO 69
<211> LENGTH: 22
<212> TYPE: RNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crRNA CER domain modified with
      2-O-Methyl RNA bases throughout
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(22)
<223> OTHER INFORMATION: The first N at the 5 prime end represents a G
      2-O-Methyl ribonucleotide, a N at the second position represents
      an U 2-O-Methyl ribonucleotide, a N at the third position
      represents an U 2-O-Methyl ribonucleotide,
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(22)
<223> OTHER INFORMATION: a N at the 4th, 5th, 6th, 7th, 8th, 9th, 10th,
      11th, 12th, 13th, 14th, 15th, 16th, 17th,18th, 19th, 20th, 21st,
      22nd position represents an U-, U-, A-, G-, A-, G-, C-, U-, A-,
      U-, G-, C-, U-, G, U-, U-, U-, U-, G-2-O-Methyl ribonucleotide,
      respectively.

<400> SEQUENCE: 69

nnnnnnnnnn nnnnnnnnnn nn                                             22


<210> SEQ ID NO 70
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of unmodified crDNA VT domain

<400> SEQUENCE: 70

gcgtacgcgt acgtgtg                                                   17


<210> SEQ ID NO 71
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of unmodified crDNA CER domain

<400> SEQUENCE: 71

gttttagagc tatgctgttt tg                                             22


<210> SEQ ID NO 72
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crDNA VT domain modified with one
      Locked Nucleic Acid base
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: a N at the sixteenth position represents a T
```

-continued

```
      Locked Nucleic Acid base.

<400> SEQUENCE: 72

gcgtacgcgt acgtgng                                                   17


<210> SEQ ID NO 73
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crDNA VT domain modified with three
      Locked Nucleic Acid bases
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: a  N at the twelfth position represents a C
      Locked Nucleic Acid base, a N at the fourteenth position
      represents a T Locked Nucleic Acid base and a N at the
      sixteenth position represents a T Locked Nucleic Acid base.

<400> SEQUENCE: 73

gcgtacgcgt angngng                                                   17


<210> SEQ ID NO 74
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crDNA VT domain modified with six
      Locked Nucleic Acid bases
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: a N at the 6th position represents a C Locked
      Nucleic Acid base, a N at the 8th position represents a C Locked
      Nucleic Acid base, a N at the tenth position represents a T Locked
      Nucleic Acid base, a N at the twelfth position represents a C
      Locked Nucleic
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: a N at the fourteenth position represents a T
      Locked Nucleic Acid base and a N at the sixteenth position
      represents a T Locked Nucleic Acid base.

<400> SEQUENCE: 74

gcgtangngn angngng                                                   17


<210> SEQ ID NO 75
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crDNA VT domain modified with three
      Locked Nucleic Acid bases and three phosphorothioate bonds
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: a first N at the 5 prime end represents a G
      deoxyribonucleotide with a phosphorothioate bond, a N at the
      second position  and a N at the third position represents a C-,
      and G-deoxyribonucleotide with a phosphorothioate bond,
      respectively
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: a N at the twelfth position represents a C
      Locked Nucleic Acid base, a N at the fourteenth position
      represents a T Locked Nucleic Acid base and a N at the sixteenth
      position represents a T Locked Nucleic Acid base.
```

-continued

```
<400> SEQUENCE: 75

nnntacgcgt angngng                                                   17


<210> SEQ ID NO 76
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crDNA CER domain modified with
      three phosphorothioate bonds
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(22)
<223> OTHER INFORMATION: the Ns at the nineteenth, twentieth and
      twenty-first positions represent T deoxyribonucleotides with
      Phosphorothioate bonds.

<400> SEQUENCE: 76

gttttagagc tatgctgtnn ng                                             22


<210> SEQ ID NO 77
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crDNA VT domain modified with one
      5-Methyl dC base
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: a  N at the twelfth position represents a
      5-Methyl dC deoxyribonucleotide.

<400> SEQUENCE: 77

gcgtacgcgt angtgtg                                                   17


<210> SEQ ID NO 78
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crDNA VT domain modified with three
      5-Methyl dC bases
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: Ns at the sixth, eighth and twelfth positions
      represent 5-Methyl dC deoxyribonucleotides

<400> SEQUENCE: 78

gcgtangngt angtgtg                                                   17


<210> SEQ ID NO 79
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crDNA VT domain modified with one
      2,6-Diaminopurine base
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: a N at the eleventh position represents a
      2,6-Diaminopurine deoxyribonucleotide.

<400> SEQUENCE: 79
```

```
gcgtacgcgt ncgtgtg                                                      17


<210> SEQ ID NO 80
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crDNA VT domain modified with two
      2,6-Diaminopurine base
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: a Ns at the fifth and eleventh positions
      represent 2,6-diaminopurine deoxyribonucleotides

<400> SEQUENCE: 80

gcgtncgcgt ncgtgtg                                                      17


<210> SEQ ID NO 81
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crDNA VT domain with Locked Nucleic
      Acid base modifications at the 5 prime end
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: The first N at the 5 prime end represent a G
      Locked Nucleic Acid base, second N represents a C Locked Nucleic
      Acid base and third N represents a G Locked Nucleic Acid base.

<400> SEQUENCE: 81

nnntacgcgt acgtgtg                                                      17


<210> SEQ ID NO 82
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crDNA CER domain with Locked
      Nucleic Acid base modifications at the 3 prime end
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(22)
<223> OTHER INFORMATION: A N at the twentieth position represents a T
      Locked Nucleic Acid base, a N at the twenty-first position
      represents a T Locked Nucleic Acid base and a N at the
      twenty-second position represents a G Locked Nucleic Acid base.

<400> SEQUENCE: 82

gttttagagc tatgctgttn nn                                                22


<210> SEQ ID NO 83
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crDNA VT domain with
      Phosphorothioate bond modifications at the 5 prime end
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: The first N at the 5 prime end represents a G
      deoxyribonucleotide with a phosphorothioate bond, second N
      represents a C deoxyribonucleotide with a phosphorothioate bond
      and third N represents a G deoxyribonucleotide with a
      phosphorothioate bond.
```

-continued

```
<400> SEQUENCE: 83

nnntacgcgt acgtgtg                                                  17


<210> SEQ ID NO 84
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crDNA VT domain with 2-O-Methyl RNA
      base modifications at the 5 prime end
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: The first N at the 5 prime end represents a G
      2-O-Methyl ribonucleotide, second N represents a C 2-O-Methyl
      ribonucleotide and third N represents a G 2-O-Methyl
      ribonucleotide.

<400> SEQUENCE: 84

nnntacgcgt acgtgtg                                                  17


<210> SEQ ID NO 85
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crDNA CER domain with 2-O-Methyl
      RNA base modifications at the 3 prime end
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(22)
<223> OTHER INFORMATION: a N at the twentieth position represents a U
      2-O-Methyl ribonucleotide, a N at the twenty-first position
      represents a U 2-O-Methyl ribonucleotide and a N at the
      twenty-second position represent G 2-O-Methyl ribonucleotide.

<400> SEQUENCE: 85

gttttagagc tatgctgttn nn                                            22


<210> SEQ ID NO 86
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of crDNA VT domain with 2-O-Methyl
      RNA base modifications at each position except T
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: A first N at 5 prime end represents a G
      2-O-Methyl ribonucleotide, a N at the second position represents
      a C 2-O-Methyl ribonucleotide, a N at the third position
      represents a G 2-O-Methyl ribonucleotide,
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(17)
<223> OTHER INFORMATION: a N at the  5th, 6th, 7th, 8th, 9th, 11th,
      12th, 13th, 15th, 17th position represents an A-, C-, G-, C-, G-,
      A-, C-, G-, G-, G- 2'-O-Methyl ribonucleotide., respectively.

<400> SEQUENCE: 86

nnntnnnnnt nnntntn                                                  17


<210> SEQ ID NO 87
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
```

-continued

```
<223> OTHER INFORMATION: Sequence of crDNA CER domain with 2-O-Methyl
      RNA base modifications at each position except T
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(22)
<223> OTHER INFORMATION: a N at the sixth position represents an A
      2-O-Methyl ribonucleotide, a N at the seventh position represents
      a G 2-O-Methyl ribonucleotide, a N at the eighth position
      represents an A 2-O-Methyl ribonucleotide,
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(22)
<223> OTHER INFORMATION: a N at the ninth position represents a G
      2-O-Methyl ribonucleotide, a N at the tenth position represents
      a C 2-O-Methyl ribonucleotide, a N at the twelfth position
      represents an A 2-O-Methyl ribonucleotide,
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(22)
<223> OTHER INFORMATION: a N at the fourteenth position represents a G
      2-O-Methyl ribonucleotide, a N at the fifteenth position
      represents a C 2-O-Methyl ribonucleotide, a N at the seventeenth
      position represents a G 2-O-Methyl ribonucleotide
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(22)
<223> OTHER INFORMATION: and a N at the twenty-second position
      represents a G 2-O-Methyl ribonucleotide.

<400> SEQUENCE: 87

nttttnnnnn tntnntnttt tn                                              22


<210> SEQ ID NO 88
<211> LENGTH: 4104
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Saccharomyces cerevisiae codon optimized Cas9
      coding sequence

<400> SEQUENCE: 88

atggataaga agtactccat tggcctagac atcggcacca attccgtggg ttgggccgtg      60

atcaccgacg agtacaaggt tccctccaag aagttcaagg tcttaggcaa taccgacagg     120

cactctatca agaagaatct gatcggtgct ttactgtttg actctggcga gaccgccgag     180

gccaccaggt tgaaacgtac cgctagaagg aggtacacca ggaggaagaa ccgtatctgc     240

tacctacaag agatcttctc caatgagatg gccaaggtcg acgactcctt cttccacagg     300

cttgaggagt ccttcctggt ggaggaggac aagaagcatg aaaggcaccc tatctttggc     360

aacatcgtcg acgaggttgc ctaccacgag aagtacccaa ccatctacca tctgaggaag     420

aaactggtag actccaccga caaggccgac ctgcgtctga tctacttagc cttagcccac     480

atgatcaagt ttagaggcca cttcctgatc gagggcgacc tgaatcccga taactccgac     540

gtcgataagc tgttcatcca gctggtccag acttacaatc agctgttcga ggagaacccc     600

atcaacgcct ctggcgtcga cgccaaagct atcctatcag cccgtctttc caagtctagg     660

aggctagaga acttgatcgc ccagcttccc ggcgagaaga agaacggcct gtttggcaat     720

ctgatcgccc tgtccttagg cttaactccc aacttcaaat ccaacttcga cctggccgaa     780

gatgctaagt tgcagttatc caaggacact tacgacgacg accttgataa cctgctggcc     840

caaatcggcg accagtacgc cgacctgttc ctagccgcca agaacttatc tgatgccatc     900

ttactgagtg acattctgag ggtcaacacc gagatcacca aagctccctt gagtgcctct     960
```

-continued

```
atgatcaaac gttacgacga acatcaccag gaccttaccc tgctaaaggc cctggtgagg   1020

caacaactgc ccgagaagta caaggagatc ttctttgacc agtccaagaa cggctacgcc   1080

ggctacatcg atggcggcgc ctctcaggag gagttctaca agttcatcaa gccaatctta   1140

gagaagatgg acggtaccga ggagctgctg gtaaagctga atagggagga cctgttaaga   1200

aaacagagga cctttgacaa cggctccatc cctcatcaga tccacctggg cgagctacac   1260

gccatcctgc gtcgtcagga ggacttctat cccttcctga aggacaacag agagaagatc   1320

gagaagattc tgacctttcg tataccctac tacgtgggcc ctctggccag aggcaattcc   1380

aggtttgcct ggatgaccag aaagtccgag gagaccatca caccctggaa ctttgaagag   1440

gttgtggaca agggcgcttc tgctcagtcc ttcatcgaga ggatgacaaa cttcgacaaa   1500

aacctgccca acgagaaggt cttacccaag cacagtctgc tgtacgagta cttcaccgta   1560

tacaacgagc tgaccaaggt taagtacgtc accgagggca tgagaaagcc tgccttcctg   1620

tccggagagc agaagaaggc cattgtggac ttgctgttca agaccaacag gaaagtgacc   1680

gtgaagcagc tgaaagagga ctacttcaag aagatcgagt gcttcgactc cgtggagatc   1740

agtggcgtgg aggataggtt caacgcctcc ctaggcacat accacgatct actaaagatt   1800

atcaaggaca aggacttcct tgacaacgag gagaacgagg acatcctaga ggacatcgtc   1860

ttaaccttaa ccctattcga ggacagggag atgatcgagg agaggctgaa gacctacgcc   1920

cacctgttcg acgacaaggt gatgaaacag ttgaagagga ggagatacac cggctggggc   1980

aggttatccc gtaagcttat caacggcatc cgtgacaagc agtccggcaa gaccatctta   2040

gacttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac   2100

tccctgactt tcaaggagga tatccagaag gctcaggtga gtggccaggg cgactcctta   2160

catgaacaca tcgcaaacct ggccggaagt cccgccatca agaagggcat cttgcaaact   2220

gtgaaagtgg ttgacgagct ggtgaaggtg atgggcaggc ataagcccga gaacatcgtt   2280

attgagatgg cccgtgagaa ccagacaacc cagaagggcc agaagaacag tcgtgaaagg   2340

atgaagagga tcgaggaagg catcaaggag ttaggctccc agatcctaaa ggagcacccc   2400

gtcgagaaca cccaactgca gaacgagaag ctgtacttgt actatctgca gaacggtagg   2460

gacatgtacg tcgaccagga gctggacatc aacaggctga gtgactacga cgtagaccac   2520

atcgtgccac agagtttcct gaaggacgac tctatcgaca acaaggtgct taccaggtcc   2580

gacaagaacc gtggcaagag tgacaatgtg ccctccgaag aggtggtcaa gaagatgaag   2640

aactactgga ggcagctgtt gaacgccaag ctgatcaccc agaggaagtt cgacaactta   2700

accaaagctg agagaggagg cctatctgaa ctagacaagg ccggtttcat caagaggcag   2760

ctggtcgaga ccaggcaaat caccaagcat gttgcccaga tcctagactc ccgtatgaac   2820

accaagtacg acgagaacga caagctaatc agagaggtga aggtgatcac cctgaagtcc   2880

aagctggtct ccgacttccg taaggacttc cagttctaca aggtcaggga gatcaacaac   2940

taccatcatg ctcatgacgc ctaccttaac gctgtggttg gcaccgcact gatcaagaag   3000

taccctaagt tggagagtga gttcgtatac ggcgactaca aggtctacga cgtgcgtaag   3060

atgatcgcca agtcagaaca ggaaatcggc aaggccaccg ccaagtactt cttctactct   3120

aacatcatga acttcttcaa aaccgagatc accctggcca acggcgagat ccgtaagagg   3180

cctcttatcg agaccaacgg cgaaaccggt gagatagtgt gggacaaggg cagagacttc   3240
```

-continued

gccaccgtga gaaaggtctt atccatgcct caagtcaaca ttgtcaagaa gaccgaggtc 3300

cagaccggcg gcttctctaa ggagagtatc ctgcccaagc gtaactctga caagctaatc 3360

gccagaaaga aggactggga ccctaagaag tatggaggct tcgactctcc cacagtagcc 3420

tactccgtgc tggtggtcgc caaggtcgag aagggcaaat ccaagaagct gaagtctgtg 3480

aaggagttgc taggcatcac catcatggag aggtcctcct tcgagaagaa tcccatcgac 3540

ttcttggagg ccaaaggtta caaggaggtc aagaaggacc tgatcatcaa gctgcccaag 3600

tactccttgt tcgagttaga gaacggcagg aagaggatgc tggcctccgc aggcgagtta 3660

cagaagggta acgagctggc cttaccctcc aagtacgtaa actttctgta cctggcttcc 3720

cattacgaga agctaaaggg cagtcccgag gacaacgagc agaagcagct gttcgtggag 3780

caacacaagc attacctgga cgagatcatc gagcagatct ccgagttcag taagcgtgtg 3840

atcctggccg acgccaatct ggacaaggtc ctgtccgcct acaacaagca tcgtgacaaa 3900

ccaatccgtg agcaggctga gaacatcatc catctattca ccctgaccaa cttaggcgcc 3960

cccgccgcct tcaagtactt tgataccact atcgacagga agaggtacac ctccaccaag 4020

gaagtactag acgccaccct gatccaccaa tccatcacag gcctgtacga gactaggatc 4080

gacttatccc agctaggtgg cgat 4104

<210> SEQ ID NO 89
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Bacteriophage T7
<220> FEATURE:
<221> NAME/KEY: T7 promoter sequence
<222> LOCATION: (1)..(20)

<400> SEQUENCE: 89

taatacgact cactataggg 20

<210> SEQ ID NO 90
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Saccharomyces cerevisiae
<220> FEATURE:
<221> NAME/KEY: ADE-URA3-DE2 target sequence minus PAM sequence
<222> LOCATION: (1)..(20)

<400> SEQUENCE: 90

gcagacatta cgaatgcaca 20

<210> SEQ ID NO 91
<211> LENGTH: 3387
<212> TYPE: DNA
<213> ORGANISM: S. thermophilus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(3387)
<223> OTHER INFORMATION: Cas9 endonuclease, genbank CS571758.1

<400> SEQUENCE: 91

atgagtgact tagttttagg acttgatatc ggtataggtt ctgttggtgt aggtatcctt 60

aacaaagtga caggagaaat tatccataaa aactcacgca tcttcccagc agctcaagca 120

gaaaataacc tagtacgtag aacgaatcgt caaggaagac gcttgacacg acgtaaaaaa 180

catcgtatag ttcgtttaaa tcgtctattt gaggaaagtg gattaatcac cgattttacg 240

-continued

```
aagatttcaa ttaatcttaa cccatatcaa ttacgagtta agggcttgac cgatgaattg    300

tctaatgaag aactgtttat cgctcttaaa aatatggtga aacaccgtgg gattagttac    360

ctcgatgatg ctagtgatga cggaaattca tcagtaggag actatgcaca aattgttaag    420

gaaaatagta aacaattaga aactaagaca ccgggacaga tacagttgga acgctaccaa    480

acatatggtc aattacgtgg tgattttact gttgagaaag atggcaaaaa acatcgcttg    540

attaatgtct ttccaacatc agcttatcgt tcagaagcct taaggatact gcaaactcaa    600

caagaattta atccacagat tacagatgaa tttattaatc gttatctcga aattttaact    660

ggaaaacgga aatattatca tggacccgga aatgaaaagt cacggactga ttatggtcgt    720

tacagaacga gtggagaaac tttagacaat atttttggaa ttctaattgg gaaatgtaca    780

ttttatccag aagagtttag agcagcaaaa gcttcctaca cggctcaaga attcaatttg    840

ctaaatgatt tgaacaatct aacagttcct actgaaacca aaaagttgag caaagaacag    900

aagaatcaaa tcattaatta tgtcaaaaat gaaaaggcaa tggggccagc gaaacttttt    960

aaatatatcg ctaagttact ttcttgtgat gttgcagata tcaagggata ccgtatcgac   1020

aaatcaggta aggctgagat tcatactttc gaagcctatc gaaaaatgaa aacgcttgaa   1080

accttagata ttgaacaaat ggatagagaa acgcttgata aattagccta tgtcttaaca   1140

ttaaacactg agagggaagg tattcaagaa gccttagaac atgaatttgc tgatggtagc   1200

tttagccaga agcaagttga cgaattggtt caattccgca aagcaaatag ttccattttt   1260

ggaaaaggat ggcataattt ttctgtcaaa ctgatgatgg agttaattcc agaattgtat   1320

gagacgtcag aagagcaaat gactatcctg acacgacttg gaaaacaaaa acgacttcgt   1380

cttcaaataa aacaaaatat ttcaaataaa acaaaatata tagatgagaa actattaact   1440

gaagaaatct ataatcctgt tgttgctaag tctgttcgcc aggctataaa aatcgtaaat   1500

gcggcgatta aagaatacgg agactttgac aatattgtca tcgaaatggc tcgtgaaaca   1560

aatgaagatg atgaaaagaa agctattcaa aagattcaaa aagccaacaa agatgaaaaa   1620

gatgcagcaa tgcttaaggc tgctaaccaa tataatggaa aggctgaatt accacatagt   1680

gttttccacg gtcataagca attagcgact aaaatccgcc tttggcatca gcaaggagaa   1740

cgttgccttt atactggtaa gacaatctca atccatgatt tgataaataa tcctaatcag   1800

tttgaagtag atcatatttt acctctttct atcacattcg atgatagcct tgcaaataag   1860

gttttggttt atgcaactgc taaccaagaa aaaggacaac gaacacctta tcaggcttta   1920

gatagtatgg atgatgcgtg gtctttccgt gaattaaaag cttttgtacg tgagtcaaaa   1980

acactttcaa acaagaaaaa agaatacctc cttacagaag aagatatttc aaagtttgat   2040

gttcgaaaga aatttattga acgaaatctt gtagatacaa gatacgcttc aagagttgtc   2100

ctcaatgccc ttcaagaaca ctttagagct cacaagattg atacaaaagt ttccgtggtt   2160

cgtggccaat ttacatctca attgagacgc cattggggaa ttgagaagac tcgtgatact   2220

tatcatcacc atgctgtcga tgcattgatt attgccgcct caagtcagtt gaatttgtgg   2280

aaaaaacaaa agaataccct tgtaagttat tcagaagaac aactccttga tattgaaaca   2340

ggtgaactta ttagtgatga tgagtacaag gaatctgtgt tcaaagcccc ttatcaacat   2400

tttgttgata cattgaagag taaagaattt gaagacagta tcttattctc atatcaagtg   2460

gattctaagt ttaatcgtaa aatatcagat gccactattt atgcgacaag acaggctaaa   2520
```

-continued

---

gtgggaaaag ataagaagga tgaaacttat gtcttaggga aaatcaaaga tatctatact 2580

caggatggtt atgatgcctt tatgaagatt tataagaagg ataagtcaaa attcctcatg 2640

tatcgtcacg acccacaaac ctttgagaaa gttatcgagc caattttaga gaactatcct 2700

aataagcaaa tgaatgaaaa aggaaaagag gtaccatgta atcctttcct aaaatataaa 2760

gaagaacatg gctatattcg taaatatagt aaaaaaggca atggtcctga aatcaagagt 2820

cttaaatact atgatagtaa gcttttaggt aatcctattg atattactcc agagaatagt 2880

aaaaataaag ttgtcttaca gtcattaaaa ccttggagaa cagatgtcta tttcaataag 2940

gctactggaa aatacgaaat ccttggatta aaatatgctg atctacaatt tgagaaaggg 3000

acaggaacat ataagatttc ccaggaaaaa tacaatgaca ttaagaaaaa agagggtgta 3060

gattctgatt cagaattcaa gtttacactt tataaaaatg atttgttact cgttaaagat 3120

acagaaacaa aagaacaaca gcttttccgt tttctttctc gaactttacc taaacaaaag 3180

cattatgttg aattaaaacc ttatgataaa cagaaatttg aaggaggtga ggcgttaatt 3240

aaagtgttgg gtaacgttgc taatggtggt caatgcataa aaggactagc aaaatcaaat 3300

atttctattt ataaagtaag aacagatgtc ctaggaaatc agcatatcat caaaaatgag 3360

ggtgataagc ctaagctaga tttttaa 3387

<210> SEQ ID NO 92
<211> LENGTH: 3369
<212> TYPE: DNA
<213> ORGANISM: S. thermophilus
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(3369)
<223> OTHER INFORMATION: Cas9 endonuclease, genbank CS571770.1

<400> SEQUENCE: 92

atgagtgact tagttttagg acttgatatc ggtataggtt ctgttggtgt aggtatcctt 60

aacaaagtga caggagaaat tatccataaa aactcacgca tcttcccagc agctcaagca 120

gaaaataacc tagtacgtag aacgaatcgt caaggaagac gcttgacacg acgtaaaaaa 180

catcgtatag ttcgtttaaa tcgtctattt gaggaaagtg gattaatcac cgattttacg 240

aagatttcaa ttaatcttaa cccatatcaa ttacgagtta agggcttgac cgatgaattg 300

tctaatgaag aactgtttat cgctcttaaa aatatggtga aacaccgtgg gattagttac 360

ctcgatgatg ctagtgatga cggaaattca tcagtaggag actatgcaca aattgttaag 420

gaaaatagta aacaattaga aactaagaca ccgggacaga tacagttgga acgctaccaa 480

acatatggtc aattacgtgg tgattttact gttgagaaag atggcaaaaa acatcgcttg 540

attaatgtct ttccaacatc agcttatcgt tcagaagcct taaggatact gcaaactcaa 600

caagaattta attcacagat tacagatgaa tttattaatc gttatctcga aattttaact 660

ggaaaacgga aatattatca tggacccgga aatgaaaagt cacggactga ttatggtcgt 720

tacagaacga atggagaaac tttagacaat atttttggaa ttctaattgg gaaatgtaca 780

ttttatccag acgagtttag agcagcaaaa gcttcctaca cggctcaaga attcaatttg 840

ctaaatgatt tgaacaatct aacagttcct actgaaacca aaaagttgag caaagaacag 900

aagaatcaaa tcattaatta tgtcaaaaat gaaaaggtaa tggggccagc gaaacttttt 960

aaatatatcg ctaaattact ttcttgtgat gttgcagata tcaagggaca ccgtatcgac 1020

-continued

```
aaatcaggta aggctgagat tcatactttc gaagcctatc gaaaaatgaa aacgcttgaa   1080

accttagata ttgagcaaat ggatagagaa acgcttgata aattagccta tgtcttaaca   1140

ttaaacactg agagggaagg tattcaagaa gctttagaac atgaatttgc tgatggtagc   1200

tttagccaga agcaagttga cgaattggtt caattccgca aagcaaatag ttccattttt   1260

ggaaaaggat ggcataattt ttctgtcaaa ctgatgatgg agttaattcc agaattgtat   1320

gagacgtcag aagagcaaat gactatcctg acacgacttg gaaaacaaaa aacaacttcg   1380

tcttcaaata aaacaaaata tatagatgag aaactattaa ctgaagaaat ctataatcct   1440

gttgttgcta agtctgttcg ccaggctata aaaatcgtaa atgcggcgat taaagaatac   1500

ggagactttg acaatattgt catcgaaatg gctcgtgaaa caaatgaaga tgatgaaaag   1560

aaagctattc aaaagattca aaaagccaac aaagatgaaa aagatgcagc aatgcttaag   1620

gctgctaacc aatataatgg aaaggctgaa ttaccacata gtgttttcca cggtcataag   1680

caattagcga ctaaaatccg cctttggcat cagcaaggag aacgttgcct ttatactggt   1740

aagacaatct caatccatga tttgataaat aatcctaatc agtttgaagt agatcatatt   1800

ttacctcttt ctatcacatt cgatgatagc cttgcaaata aggttttggt ttatgcaact   1860

gctaaccaag aaaaaggaca acgaacacct tatcaggctt tagatagtat ggatgatgcg   1920

tggtctttcc gtgaattaaa agcttttgta cgtgagtcaa aaacactttc aaacaagaaa   1980

aaagaatacc tccttacaga agaagatatt tcaaagtttg atgttcgaaa gaaatttatt   2040

gaacgaaatc ttgtagatac aagatacgct tcaagagttg tcctcaatgc ccttcaagaa   2100

cactttagag ctcacaagat tgatacaaaa gtttccgtgg ttcgtggcca atttacatct   2160

caattgagac gccattgggg aattgagaag actcgtgata cttatcatca ccatgctgtc   2220

gatgcattga ttattgccgc ctcaagtcag ttgaatttgt ggaaaaaaca aaagaatacc   2280

cttgtaagtt attcagaaga acaactcctt gatattgaaa caggtgaact tattagtgat   2340

gatgagtaca aggaatctgt gttcaaagcc ccttatcaac attttgttga tacattgaag   2400

agtaaagaat ttgaagacag tatcttattc tcatatcaag tggattctaa gtttaatcgt   2460

aaaatatcag atgccactat ttatgcgaca agacaggcta aagtgggaaa agataagaag   2520

gatgaaactt atgtcttagg gaaaatcaaa gatatctata ctcaggatgg ttatgatgcc   2580

tttatgaaga tttataagaa ggataagtca aaattcctca tgtatcgtca cgacccacaa   2640

acctttgaga aagttatcga gccaatttta gagaactatc ctaataagga aatgaatgaa   2700

aaagggaaag aagtaccatg taatcctttc ctaaaatata aagaagaaca tggctatatt   2760

cgtaaatata gtaaaaaagg caatggtcct gaaatcaaga gtcttaaata ctatgatagt   2820

aagcttttag gtaatcctat tgatattact ccagagaata gtaaaaataa agttgtctta   2880

cagtcattaa aaccttggag aacagatgtc tatttcaata aaaatactgg taaatatgaa   2940

attttaggac tgaaatatgc tgatttacaa tttgaaaaga agacaggaac atataagatt   3000

tcccaggaaa aatacaatgg cattatgaaa gaagagggtg tagattctga ttcagaattc   3060

aagtttacac tttataaaaa tgatttgtta ctcgttaaag atacagaaac aaaagaacaa   3120

cagcttttcc gttttctttc tcgaactatg cctaatgtga aatattatgt agagttaaag   3180

ccttattcaa aagataaatt tgagaagaat gagtcactta ttgaaatttt aggttctgca   3240

gataagtcag gacgatgtat aaaagggcta ggaaaatcaa atatttctat ttataaggta   3300
```

-continued

agaacagatg tcctaggaaa tcagcatatc atcaaaaatg agggtgataa gcctaagcta 3360

gatttttaa 3369

<210> SEQ ID NO 93
<211> LENGTH: 4113
<212> TYPE: DNA
<213> ORGANISM: S. agalactiae
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(4113)
<223> OTHER INFORMATION: Cas9 endonuclease, genbank CS571785.1

<400> SEQUENCE: 93

atgaataagc catattcaat aggccttgac atcggtacta attccgtcgg atggagcatt 60

attacagatg attataaagt acctgctaag aagatgagag ttttagggaa cactgataaa 120

gaatatatta agaagaatct cataggtgct ctgctttttg atggcgggaa tactgctgca 180

gatagacgct tgaagcgaac tgctcgtcgt cgttatacac gtcgtagaaa tcgtattcta 240

tatttacaag aaatttttgc agaggaaatg agtaaagttg atgatagttt ctttcatcga 300

ttagaggatt cttttctagt tgaggaagat aagagaggga gcaagtatcc tatctttgca 360

acattgcagg aagagaaaga ttatcatgaa aaattttcga caatctatca tttgagaaaa 420

gaattagctg acaagaaaga aaaagcagac cttcgtctta tttatattgc tctagctcat 480

atcattaaat ttagagggca tttcctaatt gaggatgata gctttgatgt caggaataca 540

gacatttcaa aacaatatca agattttta gaaatcttta atacaacttt tgaaaataat 600

gatttgttat ctcaaaacgt tgacgtagag gcaatactaa cagataagat tagcaagtct 660

gcgaagaaag atcgtatttt agcgcagtat cctaaccaaa aatctactgg catttttgca 720

gaatttttga aattgattgt cggaaatcaa gctgacttca agaaatattt caatttggag 780

gataaaacgc cgcttcaatt cgctaaggat agctacgatg aagatttaga aaatcttctt 840

ggacagattg gtgatgaatt tgcagactta ttctcagcag cgaaaaagtt atatgatagt 900

gtccttttgt ctggcattct tacagtaatc gacctcagta ccaaggcgcc actttcagct 960

tctatgattc agcgttatga tgaacataga gaggacttga aacagttaaa acaattcgta 1020

aaagcttcat tgccggaaaa atatcaagaa atatttgctg attcatcaaa agatggctac 1080

gctggttata ttgaaggtaa aactaatcaa gaagcttttt ataaatacct gtcaaaattg 1140

ttgaccaagc aagaagatag cgagaatttt cttgaaaaaa tcaagaatga agatttcttg 1200

agaaaacaaa ggacctttga taatggctca attccacacc aagtccattt gacagagctg 1260

aaagctatta tccgccgtca atcagaatac tatcccttct tgaaagagaa tcaagatagg 1320

attgaaaaaa tccttacctt tagaattcct tattatatcg ggccactagc acgtgagaag 1380

agtgattttg catggatgac tcgcaaaaca gatgacagta ttcgaccttg gaattttgaa 1440

gacttggttg ataaagaaaa atctgcggaa gcttttatcc atcgtatgac caacaatgat 1500

ttttatcttc ctgaagaaaa agttttacca aagcatagtc ttatttatga aaaatttacg 1560

gtctataatg agttgactaa ggttagatat aaaaatgagc aaggtgagac ttattttttt 1620

gatagcaata ttaaacaaga aatctttgat ggagtattca aggaacatcg taaggtatcc 1680

aagaagaagt tgctagattt tctggctaaa gaatatgagg agtttaggat agtagatgtt 1740

attggtctag ataaagaaaa taaagctttc aacgcctcat tgggaactta ccacgatctc 1800

-continued

---

gaaaaaatac tagacaaaga ttttctagat aatccagata atgagtctat tctggaagat 1860

atcgtccaaa ctctaacatt atttgaagac agagaaatga ttaagaagcg tcttgaaaac 1920

tataaagatc tttttacaga gtcacaacta aaaaaactct atcgtcgtca ctatactggc 1980

tggggacgat tgtctgctaa gttaatcaat ggtattcgag ataaagagag tcaaaaaaca 2040

atcttggact atcttattga tgatggtaga tctaatcgca actttatgca gttgataaat 2100

gatgatggtc tatctttcaa atcaattatc agtaaggcac aggctggtag tcattcagat 2160

aatctaaaag aagttgtagg tgagcttgca ggtagccctg ctattaaaaa gggaattcta 2220

caaagtttga aaattgttga tgagcttgtt aaagtcatgg gatacgaacc tgaacaaatt 2280

gtggttgaga tggcgcgtga gaatcaaaca acaaatcaag gtcgtcgtaa ctctcgacaa 2340

cgctataaac ttcttgatga tggcgttaag aatctagcta gtgacttgaa tggcaatatt 2400

ttgaaagaat atcctacgga taatcaagcg ttgcaaaatg aaagactttt cctttactac 2460

ttacaaaacg gaagagatat gtatacaggg gaagctctag atattgacaa tttaagtcaa 2520

tatgatattg accacattat tcctcaagct ttcataaaag atgattctat tgataatcgt 2580

gttttggtat catctgctaa aaatcgtgga aagtcagatg atgttcctag ccttgaaatt 2640

gtaaaagatt gtaaagtttt ctggaaaaaa ttacttgatg ctaagttaat gagtcagcgt 2700

aagtatgata atttgactaa ggcagagcgc ggaggcctaa cttccgatga taaggcaaga 2760

tttatccaac gtcagttggt tgagacacga caaattacca agcatgttgc ccgtatcttg 2820

gatgaacgct ttaataatga gcttgatagt aaaggtagaa ggatccgcaa agttaaaatt 2880

gtaaccttga agtcaaattt ggtttcaaat ttccgaaaag aatttggatt ctataaaatt 2940

cgtgaagtta acaattatca ccatgcacat gatgcctatc ttaatgcagt agttgctaaa 3000

gctattctaa ccaaatatcc tcagttagag ccagaatttg tctacggcga ctatccaaaa 3060

tataatagtt acaaaacgcg taaatccgct acagaaaagc tatttttcta ttcaaatatt 3120

atgaacttct ttaaaactaa ggtaacttta gcggatggaa ccgttgttgt aaaagatgat 3180

attgaagtta ataatgatac gggtgaaatt gtttgggata aaaagaaaca ctttgcgaca 3240

gttagaaaag tcttgtcata ccctcagaac aatatcgtga agaagacaga gattcagaca 3300

ggtggtttct ctaaggaatc aatcttggcg catggtaact cagataagtt gattccaaga 3360

aaaacgaagg atatttattt agatcctaag aaatatggag gttttgatag tccgatagta 3420

gcttactctg ttttagttgt agctgatatc aaaaagggta aagcacaaaa actaaaaaca 3480

gttacggaac ttttaggaat taccatcatg gagaggtcca gatttgagaa aaatccatca 3540

gctttccttg aatcaaaagg ctatttaaat attagggctg ataaactaat tattttgccc 3600

aagtatagtc tgttcgaatt agaaaatggg cgtcgtcgat tacttgctag tgctggtgaa 3660

ttacaaaaag gtaatgagct agccttacca acacaattta tgaagttctt ataccttgca 3720

agtcgttata atgagtcaaa aggtaaacca gaggagattg agaagaaaca agaatttgta 3780

aatcaacatg tctcttattt tgatgacatc cttcaattaa ttaatgattt ttcaaaacga 3840

gttattctag cagatgctaa tttagagaaa atcaataagc tttaccaaga taataaggaa 3900

aatatatcag tagatgaact tgctaataat attatcaatc tatttacttt taccagtcta 3960

ggagctccag cagcttttaa attttttgat aaaatagttg atagaaaacg ctatacatca 4020

actaaagaag tacttaattc taccctaatt catcaatcta ttactggact ttatgaaaca 4080

-continued

```
cgtattgatt tgggtaagtt aggagaagat tga                              4113


<210> SEQ ID NO 94
<211> LENGTH: 4134
<212> TYPE: DNA
<213> ORGANISM: S. agalactiae
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(4134)
<223> OTHER INFORMATION: Cas9 endonuclease, genbank CS571790.1

<400> SEQUENCE: 94

atgaataagc catattcaat aggccttgac atcggtacta attccgtcgg atggagcatt     60

attacagatg attataaagt acctgctaag aagatgagag ttttagggaa cactgataaa    120

gaatatatta agaagaatct cataggtgct ctgctttttg atggcgggaa tactgctgca    180

gatagacgct tgaagcgaac tgctcgtcgt cgttatacac gtcgtagaaa tcgtattcta    240

tatttacaag aaatttttgc agaggaaatg agtaaagttg atgatagttt ctttcatcga    300

ttagaggatt cttttctagt tgaggaagat aagagaggta gcaagtatcc tatctttgca    360

acaatgcagg aggagaaata ttatcatgaa aaatttccga caatctatca tttgagaaaa    420

gaattggctg acaagaaaga aaaagcagac cttcgtcttg tttatctggc tctagctcat    480

atcattaaat tcagagggca tttcctaatt gaggatgata gatttgatgt gaggaatacc    540

gatattcaaa aacaatatca agccttttta gaaatttttg atactacctt tgaaaataat    600

catttgttat ctcaaaatgt agatgtagaa gcaattctaa cagataagat tagcaagtct    660

gcgaagaagg atcgcatctt agcgcagtat cctaaccaaa aatctactgg tatttttgca    720

gaatttttga aattgattgt cggaaatcaa gctgacttca agaaacattt caatttggag    780

gataaaacac cgcttcaatt cgctaaggat agctacgatg aagatttaga aaatcttctt    840

ggacagattg gtgatgaatt tgcagactta ttctcagtag cgaaaaagct atatgatagt    900

gttcttttat ctggcattct tacagtaact gatctcagta ccaaggcgcc actttctgcc    960

tctatgattc agcgttatga tgaacatcat gaggacttaa agcatctaaa acaattcgta   1020

aaagcttcat tacctgaaaa ttatcgggaa gtatttgctg attcatcaaa agatggctac   1080

gctggctata ttgaaggcaa aactaatcaa gaagcttttt ataaatatct gttaaaattg   1140

ttgaccaaac aagaaggtag cgagtatttt cttgagaaaa ttaagaatga agatttttg    1200

agaaaacaga gaacctttga taatggctca atcccgcatc aagtccattt gacagaattg   1260

agggctatta ttcgacgtca atcagaatac tatccattct tgaaagagaa tcaagatagg   1320

attgaaaaaa tccttacctt tagaattcct tattatgtcg ggccactagc acgtgagaag   1380

agtgattttg catggatgac tcgcaaaaca gatgacagta ttcgaccttg gaattttgaa   1440

gacttggttg ataaagaaaa atctgcggaa gcttttatcc atcgcatgac caacaatgac   1500

ctctatcttc cagaagaaaa agttttacca aagcatagtc ttatttatga aaaatttact   1560

gtttacaatg aattaacgaa ggttagattt ttggcagaag gctttaaaga ttttcaattt   1620

ttaaatagga agcaaaaaga aactatcttt aacagcttgt ttaaggaaaa acgtaaagta   1680

actgaaaagg atattattag ttttttgaat aaagttgatg gatatgaagg aattgcaatc   1740

aaaggaattg agaaacagtt taacgctagc ctttcaacct atcatgatct taaaaaaata   1800

cttggcaagg atttccttga taatacagat aacgagctta ttttggaaga tatcgtccaa   1860
```

-continued

actctaacct tatttgaaga tagagaaatg attaagaagt gtcttgacat ctataaagat 1920

tttttacag agtcacagct taaaaagctc tatcgccgtc actatactgg ctggggacga 1980

ttgtctgcta agctaataaa tggcatccga aataaagaga atcaaaaaac aatcttggac 2040

tatcttattg atgatggaag tgcaaaccga aacttcatgc agttgataaa tgatgatgat 2100

ctatcattta aaccaattat tgacaaggca cgaactggta gtcattcgga taatctgaaa 2160

gaagttgtag gtgaacttgc tggtagccct gctattaaaa aagggattct acaaagtttg 2220

aaaatagttg atgagctggt taaagtcatg ggctatgaac ctgaacaaat cgtggttgaa 2280

atggcacgtg agaaccaaac gacagcaaaa ggattaagtc gttcacgaca acgcttgaca 2340

accttgagag aatctcttgc taatttgaag agtaatattt tggaagagaa aaagcctaag 2400

tatgtgaaag atcaagttga aaatcatcat ttatctgatg accgtctttt cctttactac 2460

ttacaaaacg gaagagatat gtatacaaaa aaggctctgg atattgataa tttaagtcaa 2520

tatgatattg accacattat tcctcaagct ttcataaaag atgattctat tgataatcgt 2580

gttttggtat catctgctaa aaatcgtgga aaatcagatg atgttcctag cattgaaatt 2640

gtaaaagctc gcaaaatgtt ctggaaaaat ttactggatg ctaagttaat gagtcagcgt 2700

aagtatgata atttgactaa ggcagagcgc ggaggcctaa cttccgatga taaggcaaga 2760

tttatccaac gtcagttggt tgagactcga caaattacca agcatgtagc tcgtatcttg 2820

gatgaacgct tcaataatga agttgataat ggtaaaaaga tttgcaaggt taaaattgta 2880

accttgaagt caaatttggt ttcaaatttc cgaaaagaat ttggattcta taaaattcgt 2940

gaagttaatg attatcacca tgcacacgat gcttatctta atgcagtagt tgccaaagct 3000

attctaacca aatatccaca gttagagcca gagtttgtct acggaatgta tagacagaaa 3060

aaactttcga aaatcgttca tgaggataag gaagaaaaat atagtgaagc aaccaggaaa 3120

atgtttttct actccaactt gatgaatatg ttcaaaagag ttgtgaggtt agcagatggt 3180

tctattgttg taagaccagt aatagaaact ggtagatata tgagaaaaac tgcatgggat 3240

aaaaagaaac actttgcgac agttagaaaa gtcttgtcat accctcagaa caatatcgtg 3300

aagaagacag agattcagac aggtggtttc tctaaggaat caatcttggc gcatggtaac 3360

tcagataagt tgattccaag aaaaacgaag gatatttatt tagatcctaa gaaatatgga 3420

ggttttgata gtccgatagt agcttactct gttttagttg tagctgatat caaaaaaggt 3480

aaagcacaaa aactaaaaac agttacggaa cttttaggaa ttaccatcat ggagaggtcc 3540

agatttgaga aaaatccatc agctttcctt gaatcaaaag gttatttaaa tattagggac 3600

gataaattaa tgattttacc gaagtatagt ctgttcgaat tagaaaatgg gcgtcgtcga 3660

ttacttgcta gtgctggtga attacaaaaa ggtaacgagc tagccttacc aacacaattt 3720

atgaagttct tatacсttgc aagtcgttat aatgagtcaa aaggtaaacc agaggagatt 3780

gagaagaaac aagaatttgt aaatcaacat gtctcttatt ttgatgacat ccttcaatta 3840

attaatgatt tttcaaaacg agttattcta gcagatgcta atttagagaa aatcaataag 3900

ctttaccagg ataataagga aaatatacca gtagatgaac ttgctaataa tattatcaat 3960

ctatttactt ttaccagtct aggagctcca gcagctttta aatttttga taaaatagtt 4020

gatagaaaac gctatacatc aactaaagaa gtacttaatt ctactctaat ccatcaatct 4080

attactggac tttatgaaac acgtattgat ttgggtaaat taggagaaga ttga 4134

-continued

<210> SEQ ID NO 95
<211> LENGTH: 4038
<212> TYPE: DNA
<213> ORGANISM: S. mutans
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(4038)
<223> OTHER INFORMATION: Cas9 endonuclease, genbank CS571790.1

<400> SEQUENCE: 95

atgaaaaaac cttactctat tggacttgat attggaacca attctgttgg ttgggctgtt 60

gtgacagatg actacaaagt tcctgctaag aagatgaagg ttctgggaaa tacagataaa 120

agtcatatcg agaaaaattt gcttggcgct ttattatttg atagcgggaa tactgcagaa 180

gacagacggt taaagagaac tgctcgccgt cgttacacac gtcgcagaaa tcgtatttta 240

tatttgcaag agatttttc agaagaaatg ggcaaggtag atgatagttt ctttcatcgt 300

ttagaggatt cttttcttgt tactgaggat aaacgaggag agcgccatcc catttttggg 360

aatcttgaag aagaagttaa gtatcatgaa aattttccaa ccatttatca tttgcggcaa 420

tatcttgcgg ataatccaga aaaagttgat ttgcgtttag tttatttggc tttggcacat 480

ataattaagt ttagaggtca ttttttaatt gaaggaaagt ttgatacacg caataatgat 540

gtacaaagac tgtttcaaga atttttagca gtctatgata atacttttga gaatagttcg 600

cttcaggagc aaaatgttca agttgaagaa attctgactg ataaaatcag taaatctgct 660

aagaaagata gagttttgaa actttttcct aatgaaaagt ctaatggccg ctttgcagaa 720

tttctaaaac taattgttgg taatcaagct gattttaaaa agcattttga attagaagag 780

aaagcaccat tgcaattttc taaagatact tatgaagaag agttagaagt actattagct 840

caaattggag ataattacgc agagctcttt ttatcagcaa agaaactgta tgatagtatc 900

cttttatcag ggattttaac agttactgat gttggtacca aagcgccttt atctgcttcg 960

atgattcagc gatataatga acatcagatg gatttagctc agcttaaaca attcattcgt 1020

cagaaattat cagataaata taacgaagtt ttttctgatg tttcaaaaga cggctatgcg 1080

ggttatattg atgggaaaac aaatcaagaa gcttttttata aataccttaa aggtctatta 1140

aataagattg agggaagtgg ctatttcctt gataaaattg agcgtgaaga ttttctaaga 1200

aagcaacgta cctttgacaa tggctctatt ccacatcaga ttcatcttca agaaatgcgt 1260

gctatcattc gtagacaggc tgaattttat ccgtttttag cagacaatca agataggatt 1320

gagaaattat tgactttccg tattccctac tatgttggtc cattagcgcg cggaaaaagt 1380

gattttgctt ggttaagtcg gaaatcggct gataaaatta caccatggaa ttttgatgaa 1440

atcgttgata aagaatcctc tgcagaagct tttatcaatc gtatgacaaa ttatgatttg 1500

tacttgccaa atcaaaaagt tcttcctaaa catagtttat tatacgaaaa atttactgtt 1560

tacaatgaat taacaaaggt taaatataaa acagagcaag gaaaaacagc atttttttgat 1620

gccaatatga agcaagaaat ctttgatggc gtatttaagg tttatcgaaa agtaactaaa 1680

gataaattaa tggatttcct tgaaaaagaa tttgatgaat ttcgtattgt tgatttaaca 1740

ggtctggata aagaaaataa agtatttaac gcttcttatg gaacttatca tgatttgtgt 1800

aaaattttag ataaagattt tctcgataat tcaaagaatg aaaagatttt agaagatatt 1860

gtgttgacct taacgttatt tgaagataga gaaatgatta gaaaacgtct agaaaattac 1920

-continued

agtgatttat tgaccaaaga acaagtgaaa aagctggaaa gacgtcatta tactggttgg 1980

ggaagattat cagctgagtt aattcatggt attcgcaata aagaaagcag aaaaacaatt 2040

cttgattatc tcattgatga tggcaatagc aatcggaact ttatgcaact gattaacgat 2100

gatgctcttt ctttcaaaga agagattgct aaggcacaag ttattggaga aacagacaat 2160

ctaaatcaag ttgttagtga tattgctggc agccctgcta ttaaaaaagg aattttacaa 2220

agcttgaaga ttgttgatga gcttgtcaaa attatgggac atcaacctga aaatatcgtc 2280

gtggagatgg cgcgtgaaaa ccagtttacc aatcagggac gacgaaattc acagcaacgt 2340

ttgaaaggtt tgacagattc tattaaagaa tttggaagtc aaattcttaa agaacatccg 2400

gttgagaatt cacagttaca aaatgataga ttgtttctat attatttaca aaacggcaga 2460

gatatgtata ctggagaaga attggatatt gattatctaa gccagtatga tatagaccat 2520

attatcccgc aagcttttat aaaggataat tctattgata atagagtatt gactagctca 2580

aaggaaaatc gtggaaaatc ggatgatgta ccaagtaaag atgttgttcg taaaatgaaa 2640

tcctattgga gtaagctact ttcggcaaag cttattacac aacgtaaatt tgataatttg 2700

acaaaagctg aacgaggtgg attgaccgac gatgataaag ctggattcat caagcgtcaa 2760

ttagtagaaa cacgacaaat taccaaacat gtagcacgta ttctggacga acgatttaat 2820

acagaaacag atgaaaacaa caagaaaatt cgtcaagtaa aaattgtgac cttgaaatca 2880

aatcttgttt ccaatttccg taaagagttt gaactctaca aagtgcgtga aattaatgac 2940

tatcatcatg cacatgatgc ctatctcaat gctgtaattg gaaaggcttt actaggtgtt 3000

tacccacaat tggaacctga atttgtttat ggtgattatc ctcattttca tggacataaa 3060

gaaaataaag caactgctaa gaaatttttc tattcaaata ttatgaactt ctttaaaaaa 3120

gatgatgtcc gtactgataa aaatggtgaa attatctgga aaaaagatga gcatatttct 3180

aatattaaaa aagtgctttc ttatccacaa gttaatattg ttaagaaagt agaggagcaa 3240

acgggaggat tttctaaaga atctatcttg ccgaaaggta attctgacaa gcttattcct 3300

cgaaaaacga agaaatttta ttgggatacc aagaaatatg gaggatttga tagcccgatt 3360

gttgcttatt ctattttagt tattgctgat attgaaaaag gtaaatctaa aaaattgaaa 3420

acagtcaaag ccttagttgg tgtcactatt atggaaaaga tgacttttga aagggatcca 3480

gttgcttttc ttgagcgaaa aggctatcga aatgttcaag aagaaaatat tataaagtta 3540

ccaaaatata gtttatttaa actagaaaac ggacgaaaaa ggctattggc aagtgctagg 3600

gaacttcaaa agggaaatga aatcgttttg ccaaatcatt taggaacctt gctttatcac 3660

gctaaaaata ttcataaagt tgatgaacca aagcatttgg actatgttga taaacataaa 3720

gatgaattta aggagttgct agatgttgtg tcaaactttt ctaaaaaata tactttagca 3780

gaaggaaatt tagaaaaaat caaagaatta tatgcacaaa ataatggtga agatcttaaa 3840

gaattagcaa gttcatttat caacttatta acatttactg ctataggagc accggctact 3900

tttaaattct ttgataaaaa tattgatcga aaacgatata cttcaactac tgaaattctc 3960

aacgctaccc tcatccacca atccatcacc ggtctttatg aaacgcggat tgatctcaat 4020

aagttaggag gagactaa 4038

<210> SEQ ID NO 96
<211> LENGTH: 94
<212> TYPE: RNA

-continued

```
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: single guide RNA targeting the LIGCas-3 target sequence

<400> SEQUENCE: 96

gcguacgcgu acguguqguu uuagagcuag aaauagcaag uuaaaauaag gcuaguccgu     60

uaucaacuug aaaaaguggc accgagucgg ugcu                                94
```

That which is claimed:

1. A guide polynucleotide comprising:

(i) a first nucleotide sequence domain that is complementary to a nucleotide sequence in a target DNA; and,

(ii) a second nucleotide sequence domain that interacts with a Cas endonuclease,

wherein the first nucleotide sequence domain and the second nucleotide sequence domain are composed of deoxyribonucleic acids (DNA), ribonucleic acids (RNA), or a combination thereof, wherein the guide polynucleotide does not solely comprise ribonucleic acids.

2. The guide polynucleotide of claim **1**, wherein the first nucleotide sequence domain and the second nucleotide sequence domain are located on a single molecule.

3. The guide polynucleotide of claim **1**, wherein the second nucleotide sequence domain comprises two separate molecules that are capable of hybridizing along a region of complementarity.

4. The guide polynucleotide of claim **1**, wherein the first nucleotide sequence domain is a DNA sequence and the second nucleotide sequence domain is selected from the group consisting of a DNA sequence, a RNA sequence, and a combination thereof.

5. The guide polynucleotide of claim **1**, wherein the first nucleotide sequence domain and the second nucleotide sequence domain are DNA sequences.

6. The guide polynucleotide of claim **1**, wherein the first nucleotide sequence domain and/or the second nucleotide sequence domain comprises at least one modification, wherein said at least one modification is selected from 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, and any combination thereof.

7. The guide polynucleotide of claim **1**, wherein the first nucleotide sequence domain and/or the second nucleotide sequence domain comprises at least one modification that provides for an additional beneficial feature, wherein said at least one modification is selected from 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, and any combination thereof.

8. The guide polynucleotide of claim **7**, wherein the additional beneficial feature is selected from the group consisting 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.

9. A plant or seed comprising the guide polynucleotide of claim **1**.

10. A guide polynucleotide/Cas endonuclease complex wherein the guide polynucleotide comprises:

(i) a first nucleotide sequence domain that is complementary to a nucleotide sequence in a target DNA; and,

(ii) a second nucleotide sequence domain that interacts with a Cas endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site.

11. The guide polynucleotide/Cas endonuclease complex of claim **10**, wherein the first nucleotide sequence domain and the second nucleotide sequence domain of the guide polynucleotide are composed of deoxyribonucleic acids (DNA), ribonucleic acids (RNA), or a combination thereof, wherein the guide polynucleotide does not solely comprise ribonucleic acids.

12. The guide polynucleotide/Cas endonuclease complex of claim **10**, wherein the first nucleotide sequence domain and/or the second nucleotide sequence domain of said guide polynucleotide comprises at least one modification that provides for an additional beneficial feature, wherein said at least one modification is selected from 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, and any combination thereof.

13. The guide polynucleotide/Cas endonuclease complex of claim **10**, wherein the Cas endonuclease is a Cas9 endonuclease.

14. A plant or seed comprising the guide polynucleotide/Cas endonuclease complex of claims **10**.

15. A method for modifying a target site in the genome of a cell, the method comprising providing a guide polynucleotide to a cell having a Cas endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site.

16. A method for modifying a target site in the genome of a cell, the method comprising providing a guide polynucleotide and a Cas endonuclease to a cell, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site.

17. The method of claim **16**, further comprising providing a donor DNA to said cell, wherein said donor DNA comprises a polynucleotide of interest.

18. The method of any one of claims **16**, further comprising identifying at least one cell that has a modification at said target, wherein the modification at said target site is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

19. A method for introducing a polynucleotide of interest into a target site in the genome of a cell, the method comprising:

a) providing a guide polynucleotide, a donor DNA and a Cas endonuclease to a cell, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site;

b) contacting the cell of (a) with a donor DNA comprising a polynucleotide of interest; and,

c) identifying at least one cell from (b) comprising in its genome the polynucleotide of interest integrated at said target site.

20. The method of claim **19**, wherein the donor DNA and Cas endonuclease are introduced into said cell using at least one recombinant DNA construct capable of expressing the donor DNA and/or the Cas endonuclease.

21. The method of claims **16**, wherein the guide polynucleotide is provided directly by particle bombardment.

22. The method of claims **16**, wherein the guide polynucleotide is provided via particle bombardment or *Agrobacterium* transformation of a recombinant DNA construct comprising a U6 polymerase III promoter.

23. The method of claims **16**, wherein the guide polynucleotide is a single guide polynucleotide comprising a variable targeting domain and a cas endonuclease recognition domain.

24. The method of claims **16**, wherein the guide polynucleotide is a duplex guide polynucleotide comprising a crNucleotide molecule and a tracrNucleotide molecule.

25. A method for modifying a target site in the genome of a cell, the method comprising:

a) providing to a cell a crNucleotide, a first recombinant DNA construct capable of expressing a tracrRNA, and a second recombinant DNA capable of expressing a Cas endonuclease, wherein said crNucleotide is a deoxyribonucleotide sequence or a combination of a deoxyribonucleotide and ribonucleotide sequence, wherein said crNucleotide, said tracrRNA and said Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site; and,

b) identifying at least one cell that has a modification at said target site, wherein the modification is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

26. A method for modifying a target site in the genome of a cell, the method comprising:

a) providing to a cell a tracrNucleotide, a first recombinant DNA construct capable of expressing a crRNA and a second recombinant DNA capable of expressing a Cas endonuclease, wherein said tracrNucleotide is selected a deoxyribonucleotide sequence or a combination of a deoxyribonucleotide and ribonucleotide sequence, wherein said tracrNucleotide, said crRNA and said Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site; and,

b) identifying at least one cell that has a modification at said target site, wherein the modification is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

27. A method for introducing a polynucleotide of interest into a target site in the genome of a cell, the method comprising:

a) providing to a cell a first recombinant DNA construct capable of expressing a guide polynucleotide, and a second recombinant DNA construct capable of expressing a Cas endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide and Cas endonuclease are capable of forming a complex that enables the Cas endonuclease to introduce a double strand break at said target site;

b) contacting the cell of (a) with a donor DNA comprising a polynucleotide of interest; and,

c) identifying at least one cell from (b) comprising in its genome the polynucleotide of interest integrated at said target site.

28. A method for editing a nucleotide sequence in the genome of a cell, the method comprising introducing a guide polynucleotide, a polynucleotide modification template and at least one Cas endonuclease into a cell, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein the Cas endonuclease introduces a double-strand break at a target site in the genome of said cell, wherein said polynucleotide modification template comprises at least one nucleotide modification of said nucleotide sequence.

29. The method of claim **16**, wherein the cell is selected from the group consisting of a non-human animal, bacterial, fungal, insect, yeast, and a plant cell.

30. The method of claim **29**, wherein the plant cell is selected from the group consisting of a monocot and dicot cell.

31. The method of claim **29**, wherein the plant cell is selected from the group consisting of maize, rice, sorghum, rye, barley, wheat, millet, oats, sugarcane, turfgrass, or switchgrass, soybean, canola, alfalfa, sunflower, cotton, tobacco, peanut, potato, tobacco, *Arabidopsis*, and safflower cell.

32. A plant or seed comprising a guide polynucleotide and a Cas9 endonuclease, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said Cas9 endonuclease and guide polynucleotide are capable of forming a complex and creating a double strand break in a genomic target site of said plant.

33. A plant or seed comprising a recombinant DNA construct and a guide polynucleotide, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein said recombinant DNA construct comprises a promoter operably linked to a nucleotide sequence encoding a plant optimized Cas endonuclease, wherein said plant optimized Cas endonuclease and guide polynucleotide are capable of forming a complex and creating a double strand break in a genomic target site of said plant.

34. The plant of claim **33**, further comprising a polynucleotide of interest integrated into said genomic target site of said plant.

35. The plant or seed of claim **33** further comprising a modification at said genomic target site, wherein the modification is selected from the group consisting of (i) a replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i)-(iii).

36. A plant or seed comprising at least one altered target sequence, wherein the at least one altered target sequence originated from a corresponding target sequence that was recognized and cleaved by a guide polynucleotide/Cas endonuclease complex, wherein the Cas endonuclease is capable of introducing a double-strand break at said target site in the plant genome, wherein said guide polynucleotide does not solely comprise ribonucleic acids.

37. A plant or seed comprising a modified nucleotide sequence, wherein the modified nucleotide sequence was produced by providing a guide polynucleotide, a polynucleotide modification template and at least one Cas endonuclease to a cell, wherein said guide polynucleotide does not solely comprise ribonucleic acids, wherein the Cas endonuclease is capable of introducing a double-strand break at a target site in the plant genome, wherein said polynucleotide modification template comprises at least one nucleotide modification of said nucleotide sequence.

38. The plant or plant cell of claim **29** wherein the at least one nucleotide modification is not a modification at said target site.

39. The plant of claim **32**, wherein the plant is a monocot or a dicot.

40. The plant of claim **39**, wherein the monocot is selected from the group consisting of maize, rice, sorghum, rye, barley, wheat, millet, oats, sugarcane, turfgrass, or switchgrass.

41. The plant of claim **39**, wherein the dicot is selected from the group consisting of soybean, canola, alfalfa, sunflower, cotton, tobacco, peanut, potato, tobacco, *Arabidopsis*, or safflower.

42. A method for selecting a plant comprising an altered target site in its plant genome, the method comprising: a) obtaining a first plant comprising at least one Cas endonuclease capable of introducing a double strand break at a target site in the plant genome; b) obtaining a second plant comprising a guide polynucleotide that is capable of forming a complex with the Cas endonuclease of (a), wherein the guide polynucleotide does not solely comprise ribonucleic acids, c) crossing the first plant of (a) with the second plant of (b); d) evaluating the progeny of (c) for an alteration in the target site and e) selecting a progeny plant that possesses the desired alteration of said target site.

43. A method for selecting a plant comprising an altered target site in its plant genome, the method comprising: a) obtaining a first plant comprising at least one Cas endonuclease capable of introducing a double strand break at a target site in the plant genome; b) obtaining a second plant comprising a guide polynucleotide and a donor DNA, wherein the guide polynucleotide does not solely comprise ribonucleic acids, wherein said guide polynucleotide is capable of forming a complex with the Cas endonuclease of (a), wherein said donor DNA comprises a polynucleotide of interest; c) crossing the first plant of (a) with the second plant of (b); d) evaluating the progeny of (c) for an alteration in the target site and e) selecting a progeny plant that comprises the polynucleotide of interest inserted at said target site.

* * * * *