



(86) Date de dépôt PCT/PCT Filing Date: 2015/03/14  
 (87) Date publication PCT/PCT Publication Date: 2015/09/17  
 (45) Date de délivrance/Issue Date: 2023/09/26  
 (85) Entrée phase nationale/National Entry: 2016/09/09  
 (86) N° demande PCT/PCT Application No.: US 2015/020622  
 (87) N° publication PCT/PCT Publication No.: 2015/139008  
 (30) Priorités/Priorities: 2014/03/14 (US61/953,333);  
 2014/09/17 (US62/051,579); 2014/11/05 (US62/075,816);  
 2014/11/05 (US62/075,811); 2015/03/13 (US62/133,129)

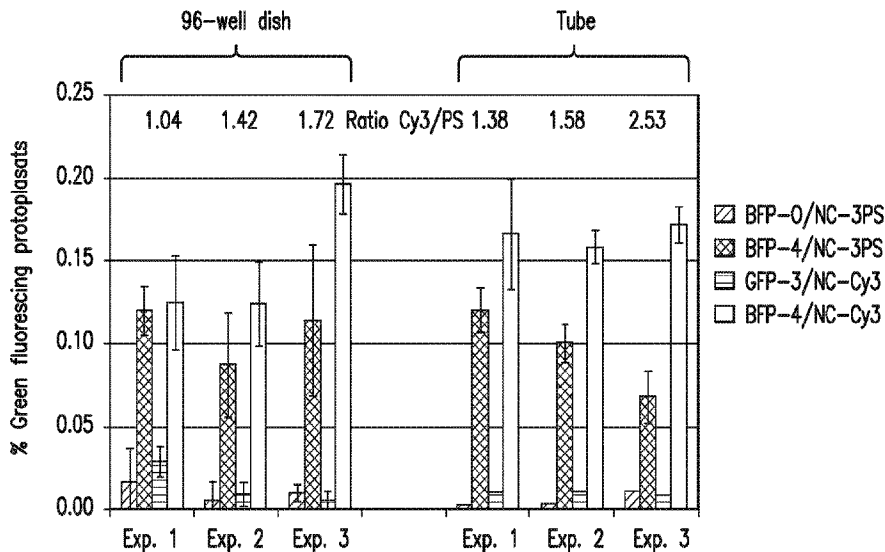
(51) Cl.Int./Int.Cl. *A01H 9/00* (2006.01),  
*C12N 15/82* (2006.01), *C12N 15/87* (2006.01),  
*C12N 5/00* (2006.01), *C12N 9/16* (2006.01)

(72) Inventeurs/Inventors:  
 BEETHAM, PETER R., US;  
 GOCAL, GREGORY F.W., US;  
 SCHOPKE, CHRISTIAN, US;  
 SAUER, NOEL, US;  
 PEARCE, JAMES, US;  
 ...

(73) Propriétaires/Owners:  
 CIBUS US LLC, US;  
 CIBUS EUROPE B.V., NL

(74) Agent: MBM INTELLECTUAL PROPERTY AGENCY

(54) Titre : PROCÉDES ET COMPOSITIONS PERMETTANT D'AMÉLIORER L'EFFICACITÉ DE MODIFICATIONS GÉNÉTIQUES CIBLÉES EN UTILISANT LA RÉPARATION DE GÈNE MÉDIÉE PAR DES OLIGONUCLÉOTIDES  
 (54) Title: METHODS AND COMPOSITIONS FOR INCREASING EFFICIENCY OF TARGETED GENE MODIFICATION USING OLIGONUCLEOTIDE-MEDIATED GENE REPAIR



(57) Abrégé/Abstract:

Provided herein include methods and compositions for making targeted changes to a DNA sequence. In various aspects and embodiments, methods and compositions for modifying a DNA sequence in a cell (such as a plant, bacterial, yeast, fungal, algal, or

(72) Inventeurs(suite)/Inventors(continued): SEGAMI, ROSA E., US; MOZORUK, JERRY, US

(57) Abrégé(suite)/Abstract(continued):

mammalian cell) are provided. In some aspects and embodiments the modification of DNA involves combining gene repair oligonucleotides with approaches that enhance the availability of components of the target cell gene repair mechanisms, such as a DNA cutter.

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

(19) World Intellectual Property

Organization

International Bureau

(43) International Publication Date

17 September 2015 (17.09.2015)



(10) International Publication Number

WO 2015/139008 A1

## (51) International Patent Classification:

A01H 9/00 (2006.01) C12N 5/00 (2006.01)

C12N 15/82 (2006.01) C12N 9/16 (2006.01)

C12N 15/87 (2006.01)

## (21) International Application Number:

PCT/US2015/020622

## (22) International Filing Date:

14 March 2015 (14.03.2015)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

61/953,333 14 March 2014 (14.03.2014) US

62/051,579 17 September 2014 (17.09.2014) US

62/075,816 5 November 2014 (05.11.2014) US

62/075,811 5 November 2014 (05.11.2014) US

62/133,129 13 March 2015 (13.03.2015) US

(71) Applicants: **CIBUS US LLC** [US/US]; 6455 Nancy Ridge Drive, San Diego, CA 92121 (US). **CIBUS EUROPE B.V.** [NL/NL]; Goessestraatweg 19, NL-4421 AD Kapelle (NL).

(72) Inventors: **BEETHAM, Peter R.**; 7101 Tanager Drive, Carlsbad, CA 92011 (US). **GOCAL, Gregory F.W.**; 13737 Bassmore Drive, San Diego, CA 92129 (US). **SCHOPKE, Christian**; 3222 Calle Vallarta, Carlsbad, CA 92009 (US). **SAUER, Noel**; 359 La Purisma Way, Ocean-side, CA 92057 (US). **PEARCE, James**; 5443 Avenida Fiesta, La Jolla, CA 92037 (US). **SEGAMI, Rosa E.**; 408

Bear Valley Parkway, Escondido, CA 92025 (US). **MO-ZORUK, Jerry**; 121 W. Glaucus Street, Unit B, Encinitas, CA 92024 (US).

(74) Agents: **WHITTAKER, Michael A.** et al.; Acuity Law Group, P.C., 12707 High Bluff Drive, Suite 200, San Diego, CA 92130 (US).

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

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

## Published:

— with international search report (Art. 21(3))

(54) Title: METHODS AND COMPOSITIONS FOR INCREASING EFFICIENCY OF TARGETED GENE MODIFICATION USING OLIGONUCLEOTIDE-MEDIATED GENE REPAIR

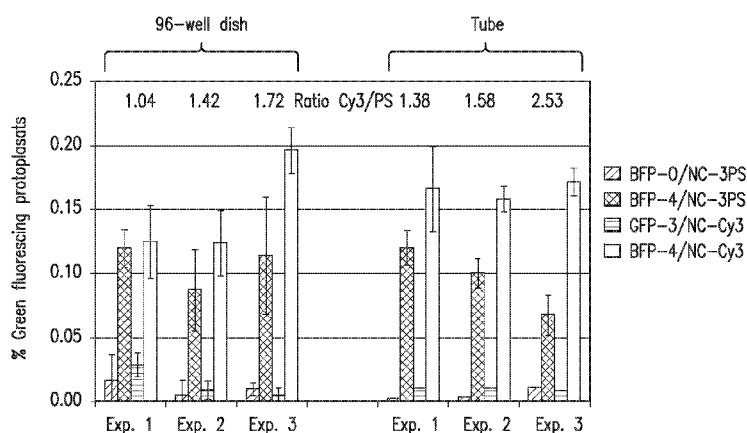


FIG. 1

(57) Abstract: Provided herein include methods and compositions for making targeted changes to a DNA sequence. In various aspects and embodiments, methods and compositions for modifying a DNA sequence in a cell (such as a plant, bacterial, yeast, fungal, algal, or mammalian cell) are provided. In some aspects and embodiments the modification of DNA involves combining gene repair oligonucleotides with approaches that enhance the availability of components of the target cell gene repair mechanisms, such as a DNA cutter.

WO 2015/139008 A1

**METHODS AND COMPOSITIONS FOR INCREASING EFFICIENCY OF  
TARGETED GENE MODIFICATION USING OLIGONUCLEOTIDE-  
MEDIATED GENE REPAIR**

[0001]

**FIELD OF THE INVENTION**

[0002] The instant disclosure relates at least in part to targeted genetic mutations and modifications, including methods and compositions for making such mutations and modifications.

**BACKGROUND**

[0003] The following discussion is merely provided to aid the reader in understanding and is not admitted to describe or constitute prior art to the present disclosure.

[0004] United States Patent No. 6,271,360 discloses methods and compositions for the introduction of predetermined genetic changes in target genes of a living cell by introducing an oligodeoxynucleotide encoding the predetermined change. The oligodeoxynucleotides are effective in mammalian, avian, plant and bacterial cells.

[0005] United States Patent No. 8,771,945 discloses vectors and vector systems, some of which encode one or more components of a CRISPR complex, as well as methods for the design and use of such vectors.

[0006] United States Patent No. 8,470,973 “refers to methods for selectively recognizing a base pair in a DNA sequence by a polypeptide, to modified polypeptides which specifically recognize one or more base pairs in a DNA sequence and, to DNA which is modified so that it can be specifically recognized by a polypeptide and to uses of the polypeptide and DNA in specific DNA targeting as well as to methods of modulating expression of target genes in a cell.”

## SUMMARY

[0007] Provided herein include methods and compositions for effecting a targeted genetic change in DNA in a cell. Certain aspects and embodiments relate to improving the efficiency of the targeting of modifications to specific locations in genomic or other nucleotide sequences. As described herein, nucleic acids which direct specific changes to the genome may be combined with various approaches to enhance the availability of components of the natural repair systems present in the cells being targeted for modification.

[0008] In a first aspect, provided are methods for introducing a gene repair oligonucleobase (GRON)-mediated mutation into a target deoxyribonucleic acid (DNA) sequence in a plant cell. In certain embodiments the methods may include, *inter alia*, culturing the plant cell under conditions that increase one or more cellular DNA repair processes prior to, and/or coincident with, delivery of a GRON into the plant cell; and/or delivery of a GRON into the plant cell greater than 15 bases in length, the GRON optionally comprising one or more; or two or more; mutation sites for introduction into the target DNA.

[0009] A “gene repair oligonucleotide” or “GRON” as used herein means an oligonucleobase (e.g., mixed duplex oligonucleotides, non-nucleotide containing molecules, single stranded oligodeoxynucleotides, double stranded oligodeoxynucleotides and other gene repair molecules) that can under certain conditions direct single, or in some embodiments multiple, nucleotide deletions, insertions or substitutions in a DNA sequence. This oligonucleotide-mediated gene repair editing of the genome may comprise both non-homology based repair systems (e.g., non-homologous end joining) and homology-based repair systems (e.g., homology-directed repair). The GRON is typically designed to align in register with a genomic target except for the designed mismatch(es). These mismatches can be recognized and corrected by harnessing one or more of the cell's endogenous DNA repair systems. In some embodiments a GRON or oligonucleotide can be designed to contain multiple differences when compared to the organisms target sequence. These differences may not all affect the protein sequence translated from said target sequence and in one or more cases be known as silent changes. Numerous variations of GRON structure, chemistry and function are described elsewhere herein. In various embodiments, a GRON as used herein may have one or more modifications. For example, a GRON as used herein may have one or more

modifications that attract DNA repair machinery to the targeted (mismatch) site and/or that prevent recombination of part or all of the GRON (other than the desired targeted deletion(s), insertion(s), substitution(s) or the like) into the genomic DNA of the target DNA sequence and/or that increase the stability of the GRON.

**[0010]** In various embodiments, a GRON may have both RNA and DNA nucleotides and/or other types of nucleobases. In some embodiments, one or more of the DNA or RNA nucleotides comprise a modification.

**[0011]** In one aspect, provided is a method of causing a genetic change in a plant cell, wherein the method involves exposing the cell to a DNA cutter and a GRON, for example a GRON that is modified as contemplated herein. In some embodiments the GRON may be modified such as with a Cy3 group, 3PS group, a 2'O-methyl group or other modification such as contemplated herein. In another aspect, provided is a plant cell that includes a DNA cutter and a GRON, for example where the GRON is modified such as with a Cy3 group, 3PS group, a 2'O-methyl group or other modification. In some embodiments, the DNA cutter is one or more selected from a CRISPR, a TALEN, a zinc finger, meganuclease, and a DNA-cutting antibiotic. In some embodiments, the DNA cutter is a CRISPR. In some embodiments, the DNA cutter is a TALEN. In some embodiments, the GRON is between 15 and 60 nucleobases in length; or between 30 and 40 nucleobases in length; or between 35 and 45 nucleobases in length; or between 20 and 70 nucleobases in length; or between 20 and 200 nucleobases in length; or between 30 and 180 nucleobases in length; or between 50 and 160 nucleobases in length; or between 70 and 150 nucleobases in length; or between 70 and 210 nucleobases in length; or between 80 and 120 nucleobases in length; or between 90 and 110 nucleobases in length; or between 95 and 105 nucleobases in length; or between 80 and 300 nucleobases in length; or between 90 and 250 nucleobases in length; or between 100 and 150 nucleobases in length; or between 100 and 200 nucleobases in length; or between 100 and 210 nucleobases in length; or between 100 and 300 nucleobases in length; or between 150 and 200 nucleobases in length; or between 200 and 300 nucleobases in length; or between 250 and 350 nucleobases in length; or between 50 and 110 nucleobases in length; or between 50 and 200 nucleobases in length; or between 150 and 210 nucleobases in length; or between 20 and 1000 nucleobases in length; or between 100 and 1000 nucleobases in length; or between 200 and 1000 nucleobases in length; or between 300 and 1000 nucleobases in length; or between 400 and 1000 nucleobases in length; or

between 500 and 1000 nucleobases in length; or between 600 and 1000 nucleobases in length; or between 700 and 1000 nucleobases in length; or between 800 and 1000 nucleobases in length; or between 900 and 1000 nucleobases in length; or between 300 and 800 nucleobases in length; or between 400 and 600 nucleobases in length; or between 500 and 700 nucleobases in length; or between 600 and 800 nucleobases in length; or longer than 30 nucleobases in length; or longer than 35 nucleobases in length; or longer than 40 nucleobases in length; or longer than 50 nucleobases in length; or longer than 60 nucleobases in length; or longer than 65 nucleobases in length; or longer than 70 nucleobases in length; or longer than 75 nucleobases in length; or longer than 80 nucleobases in length; or longer than 85 nucleobases in length; or longer than 90 nucleobases in length; or longer than 95 nucleobases in length; or longer than 100 nucleobases in length; or longer than 110 nucleobases in length; or longer than 125 nucleobases in length; or longer than 150 nucleobases in length; or longer than 165 nucleobases in length; or longer than 175 nucleobases in length; or longer than 200 nucleobases in length; or longer than 250 nucleobases in length; or longer than 300 nucleobases in length; or longer than 350 nucleobases in length; or longer than 400 nucleobases in length; or longer than 450 nucleobases in length; or longer than 500 nucleobases in length; or longer than 550 nucleobases in length; or longer than 600 nucleobases in length; or longer than 700 nucleobases in length; or longer than 800 nucleobases in length; or longer than 900 nucleobases in length.

**[0012]** GRONs may be targeted at both non-coding (NC) and coding (C) regions of a target gene. By way of example, Figs. 27 and 28 respectively depict C-GRONs and NC-GRONs suitable for introducing mutations into the rice genome in order to introduce one or more of the following amino acid substitutions to the ACCase gene. The convention is to use the amino acid numbering system for the plastidal ACCase from blackgrass (*Alopecurus myosuroides*; *Am*) as the reference. The ACCase numbering used herein is based on the numbering for the blackgrass reference sequence ACCase protein (SEQ ID NO: 1) or at an analogous amino acid residue in an ACCase paralog (V=CY3; H=3'DMT dC CPG). The following table lists ACCase mutations that produce one or more of alloxydim, butoxydim, clethodim, cloproxydim, cycloxydim, sethoxydim, tepraloxym, tralkoxydim, chlorazifop, clodinafop, clofop, diclofop, fenoxaprop, fenoxaprop-P, fenthiaprop, fluazifop, fluazifop-P, haloxyfop, haloxyfop-P, isoxapyrifop, propaquizafop,

quizalofop, quizalofop-P, trifop, pinoxaden, agronomically acceptable salts and esters of any of these herbicides, and combinations thereof resistant phenotype.

<b>Amino Acid Change</b>	<b>Codon Change</b>		<b>Amino Acid Change</b>	<b>Codon Change</b>
I1781A	ATA > GCT ATA > GCC ATA > GCA ATA > GCG		C2088F	TGC > TTT TGC > TTC
I1781L	ATA > CTT ATA > CTC ATA > CTA ATA > CTG ATA > TTA ATA > TTG		C2088G	TGC > GGT TGC > GGC TGC > GGA TGC > GGG
I1781M	ATA > ATG		C2088H	TGC > CAT TGC > CAC
I1781N	ATA > AAT ATA > AAC		C2088K	TGC > AAA TGC > AAG
I1781S	ATA > TCT ATA > TCC ATA > TCA ATA > TCG		C2088L	TGC > CTT TGC > CTC TGC > CTA TGC > CTG TGC > TTA TGC > TTG
I1781T	ATA > ACT ATA > ACC ATA > ACA ATA > ACG		C2088N	TGC > AAT TGC > AAC
I1781V	ATA > GTT ATA > GTC ATA > GTA ATA > GTG		C2088P	TGC > CCT TGC > CCC TGC > CCA TGC > CCG
G1783C	GGA > TGT GGA > TGC		C2088Q	TGC > CAA TGC > CAG
A1786P	GCT > CCT GCT > CCC GCT > CCA GCT > CCG		C2088R	TGC > CGT TGC > CGC TGC > CGA TGC > CGG TGC > AGA TGC > AGG
D2078G	GAT > GGT GAT > GGC		C2088S	TGC > TCT TGC > TCC TGC > TCA

<b>Amino Acid Change</b>	<b>Codon Change</b>		<b>Amino Acid Change</b>	<b>Codon Change</b>
D2078K	GAT > GGA		C2088T	TGC > TCG
	GAT > GGG			TGC > ACT
	GAT > AAA			TGC > ACC
	GAT > AAG			TGC > ACA
D2078T	GAT > ACT		C2088V	TGC > ACG
	GAT > ACC			TGC > GTT
	GAT > ACA			TGC > GTC
	GAT > ACG			TGC > GTA
S2079F	AGC > TTT		C2088W	TGC > GTG
	AGC > TTC			TGC > TGG
K2080E	AAG > GAA			
	AAG > GAG			

[0013] Similarly, Figs. 29 and 30 respectively depict (coding) C-GRONs and (non-coding) NC-GRONs suitable for introducing mutations into the flax genome in order to introduce one or more of the following amino acid substitutions to the EPSPS gene (with all numbering relative to the amino acid sequence of the *E. coli* AroA protein (prokaryotic EPSPS equivalent) (such as those described in US Patent No. 8,268,622). (V=CY3; H=3'DMT dC CPG). The following table lists EPSPS mutations that produce glyphosate agronomically acceptable salts and esters of any of these herbicides, and combinations thereof resistant phenotype.

<b>Amino Acid Change</b>	<b>Codon Change</b>
G96A	GGA > GCT
	GGA > GCC
	GGA > GCA
	GGA > GCG
T97I	ACA > ATT
	ACA > ATC
	ACA > ATA
P101A	CCG > GCT
	CCG > GCC
	CCG > GCA
	CCG > GCG

Amino Acid Change	Codon Change
P101S	CCG > TCT
	CCG > TCC
	CCG > TCA
	CCG > TCG
P101T	CCG > ACT
	CCG > ACC
	CCG > ACA
	CCG > ACG

[0014] The term “CRISPR” as used herein refers to elements; i.e., a cas (CRISPR associated) gene, transcript (e.g., mRNA) or protein and at least one CRISPR spacer sequence (Clustered Regularly Interspaced Short Palindromic Repeats, also known as SPIDRs--SPacer Interspersed Direct Repeats); that when effectively present or expressed in a cell could effect cleavage of a target DNA sequence via CRISPR/CAS cellular machinery such as described in e.g., Cong, L. et al., *Science*, vol. 339 no 6121 pp. 819-823 (2013); Jinek et al, *Science*, vol. 337:816-821 (2013); Wang et al., *RNA*, vol. 14, pp. 903-913 (2008); Zhang et al., *Plant Physiology*, vol. 161, pp. 20-27 (2013), Zhang et al, PCT Application No. PCT/US2013/074743; and Charpentier et al., PCT Application No. PCT/US2013/032589. In some embodiments, such as for example a CRISPR for use in a eukaryotic cell, a CRISPR as contemplated herein may also include an additional element that includes a sequence for one or more functional nuclear localization signals. CRISPRs as contemplated herein can be expressed in, administered to and/or present in a cell (such as a plant cell) in any of many ways or manifestations. For example a CRISPR as contemplated herein may include or involve one or more of a CRISPR on a plasmid, a CRISPR nickase on a plasmid, a CRISPRa on a plasmid, or a CRISPRi on a plasmid as follows:

[0015] **CRISPR on a plasmid:** A recombinant expression vector comprising:

- (i) a nucleotide sequence encoding a DNA-targeting RNA (e.g., guide RNA), wherein the DNA-targeting RNA comprises:
  - a. a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA (e.g., protospacer, spacer, or crRNA); and

- b. a second segment that interacts with a site-directed modifying polypeptide (e.g., trans-activating crRNA or tracrRNA); and
- (ii) a nucleotide sequence encoding the site-directed modifying polypeptide (e.g., cas gene), wherein the site-directed polypeptide comprises:
  - a. an RNA-binding portion that interacts with the DNA-targeting RNA (e.g., REC lobe); and
  - b. an activity portion that causes double-stranded breaks within the target DNA (e.g., NUC lobe), wherein the site of the double-stranded breaks within the target DNA is determined by the DNA-targeting RNA.

**[0016] CRISPR nickase** on a plasmid. A recombinant expression vector comprising:

- (i) a nucleotide sequence encoding a DNA-targeting RNA (e.g., guide RNA), wherein the DNA-targeting RNA comprises:
  - a. a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA (e.g., protospacer, spacer, or crRNA); and
  - b. a second segment that interacts with a site-directed modifying polypeptide (e.g., trans-activating crRNA or tracrRNA); and
- (ii) a nucleotide sequence encoding the site-directed modifying polypeptide (e.g., cas gene), wherein the site-directed polypeptide comprises:
  - a. an RNA-binding portion that interacts with the DNA-targeting RNA (e.g., REC lobe); and
  - b. an activity portion that causes single-stranded breaks within the target DNA (e.g., NUC lobe), wherein the site of the single-stranded breaks within the target DNA is determined by the DNA-targeting RNA.

**[0017] CRISPRa** on a plasmid. A recombinant expression vector comprising:

- (i) a nucleotide sequence encoding a DNA-targeting RNA (e.g., guide RNA), wherein the DNA-targeting RNA comprises:
  - a. a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA (e.g., protospacer, spacer, or crRNA); and
  - b. a second segment that interacts with a site-directed modifying polypeptide (e.g., trans-activating crRNA or tracrRNA); and

- (ii) a nucleotide sequence encoding the site-directed modifying polypeptide (e.g., cas gene), wherein the site-directed polypeptide comprises:
  - a. an RNA-binding portion that interacts with the DNA-targeting RNA (e.g., REC lobe); and
  - b. an activity portion that modulates transcription (e.g., NUC lobe; in certain embodiments increases transcription) within the target DNA, wherein the site of the transcriptional modulation within the target DNA is determined by the DNA-targeting RNA.

**[0018]** CRISPRi on a plasmid. A recombinant expression vector comprising:

- (i) a nucleotide sequence encoding a DNA-targeting RNA (e.g., guide RNA), wherein the DNA-targeting RNA comprises:
  - a. a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA (e.g., protospacer, spacer, or crRNA); and
  - b. a second segment that interacts with a site-directed modifying polypeptide (e.g., trans-activating crRNA or tracrRNA); and
- (ii) a nucleotide sequence encoding the site-directed modifying polypeptide (e.g., cas gene), wherein the site-directed polypeptide comprises:
  - a. an RNA-binding portion that interacts with the DNA-targeting RNA (e.g., REC lobe); and
  - b. an activity portion that modulates transcription/translation (e.g., NUC lobe; in some embodiments decreases transcription/translation) within the target DNA, wherein the site of transcriptional/translational modulation within the target DNA is determined by the DNA-targeting RNA.

**[0019]** Each of the CRISPR on a plasmid, CRISPR nickase on a plasmid, CRISPRa on a plasmid, and CRISPRi on a plasmid may in some embodiments alternatively have one or more appropriate elements be administered, expressed or present in a cell as an RNA (e.g., mRNA) or a protein rather than on a plasmid. Delivery of protected mRNA may be as described in Kariko, et al, US Patent No. 8,278,036.

**[0020]** In some embodiments, each of the CRISPRi and CRISPRa may include a deactivated cas9 (dCas9). A deactivated cas9 still binds to target DNA, but does not have

cutting activity. Nuclease-deficient Cas9 can result from D10A and H840A point mutations which inactivates its two catalytic domains.

**[0021]** In some embodiments, a CRISPRi inhibits transcription initiation or elongation via steric hindrance of RNA Polymerase II. CRISPRi can optionally be enhanced (CRISPREi) by fusion of a strong repressor domain to the C-terminal end of a dCas9 protein. In some embodiments, a repressor domain recruits and employs chromatin modifiers. In some embodiments, the repressor domain may include, but is not limited to domains as described in Kagale, S. et al., *Epigenetics*, vol. 6 no 2 pp141-146 (2011):

1. LDLNRPPPVEN - OsERF3 repressor domain (LxLxPP motif)
2. LRLFGVNM - AtBRD repressor domain (R/KLFGV motif)
3. LKLFGVWL - AtHsfB1 repressor domain (R/KLFGV motif)
4. LDLELRLGFA - AtSUP repressor domain (EAR motif)
5. ERSNSIELRNSFYGRARTSPWSYGDYDNCQQDHDYLLGFSWPPRSYTCSF  
CKREFRSAQALGGHMNVHRRDRARLRLQQSPSSSSTPSPYPNPNYSYST  
MANSPPPHHSPLTLFPTLSPPSSPRYRAGLIRSLSPKSKHTPENACKTKKSS  
LLVEAGEATRFTSKDACKILRNDEIISLELEIGLINESEQDLDLELRLGFA\*  
full AtSUP gene containing repressor domain (EAR motif)

**[0022]** In some embodiments, a CRISPRa activation of transcription achieved by use of dCas9 protein containing a fused C-terminal end transcriptional activator. In some embodiments, an activation may include, but is not limited to VP64 (4X VP16), AtERF98 activation domain, or AtERF98x4 concatemers such as described in Cheng, AW et al., *Cell Research*, pp1-9 (2013); Perez-Pinera, P. et al., *Nature Methods*, vol. 10 pp 913-976 (2013); Maeder, ML. et al., *Nature Methods*, vol. 10 pp 977-979 (2013) and Mali, P., et al., *Nature Biotech.*, vol. 31 pp 833-838 (2013).

**[0023]** In some embodiments the CRISPR includes a nickase. In certain embodiments, two or more CRISPR nickases are used. In some embodiments, the two or more nickases cut on opposite strands of target nucleic acid. In other embodiments, the two or more nickases cut on the same strand of target nucleic acid.

**[0024]** As used herein, “repressor protein” or “repressor” refers to a protein that binds to operator of DNA or to RNA to prevent transcription or translation, respectively.

**[0025]** As used herein, “repression” refers to inhibition of transcription or translation by binding of repressor protein to specific site on DNA or mRNA. In some embodiments,

repression includes a significant change in transcription or translation level of at least 1.5 fold, in other embodiments at least two fold, and in other embodiments at least five fold.

[0026] As used herein, an “activator protein” or “activator” with regard to gene transcription and/or translation, refers to a protein that binds to operator of DNA or to RNA to enhance or increase transcription or translation, respectively.

[0027] As used herein with regard to gene transcription and/or translation, “activation” with regard to gene transcription and/or translation, refers to enhancing or increasing transcription or translation by binding of activator protein to specific site on DNA or mRNA. In some embodiments, activation includes a significant change in transcription or translation level of at least 1.5 fold, in some embodiments at least two fold, and in some embodiments at least five fold.

[0028] In certain embodiments, conditions that increase one or more cellular DNA repair processes may include one or more of: introduction of one or more sites into the GRON or into the plant cell DNA that are targets for base excision repair, introduction of one or more sites into the GRON or into the plant cell DNA that are targets for non-homologous end joining, introduction of one or more sites into the GRON or into the plant cell DNA that are targets for microhomology-mediated end joining, introduction of one or more sites into the GRON or into the plant cell DNA that are targets for homologous recombination, and introduction of one or more sites into the GRON or into the plant cell DNA that are targets for effecting repair (e.g., base-excision repair (BER); homologous recombination repair (HR); mismatch repair (MMR); non-homologous end-joining repair (NHEJ) which include classical and alternative NHEJ; and nucleotide excision repair (NER)).

[0029] As described herein, GRONs for use herein may include one or more of the following alterations from conventional RNA and DNA nucleotides:

- one or more abasic nucleotides;
- one or more 8'oxo dA and/or 8'oxo dG nucleotides;
- a reverse base at the 3' end thereof;
- one or more 2'O-methyl nucleotides;
- one or more RNA nucleotides;
- one or more RNA nucleotides at the 5' end thereof, and in some embodiments 2, 3, 4, 5, 6, 7, 8, 9, 10, or more; wherein one or more of the RNA nucleotides may

further be modified; one or more RNA nucleotides at the 3' end thereof, and in some embodiments 2, 3, 4, 5, 6, 7, 8, 9, 10, or more; wherein one or more of the RNA nucleotides may further be modified;

one or more 2'-O-methyl RNA nucleotides at the 5' end thereof, and in some embodiments 2, 3, 4, 5, 6, 7, 8, 9, 10, or more;

an intercalating dye;

a 5' terminus cap;

a backbone modification selected from the group consisting of a phosphothioate modification, a methyl phosphonate modification, a locked nucleic acid (LNA) modification, a O-(2-methoxyethyl) (MOE) modification, a di PS modification, and a peptide nucleic acid (PNA) modification;

one or more intrastrand crosslinks;

one or more fluorescent dyes conjugated thereto, and in some embodiments at the 5' or 3' end of the GRON; and

one or more bases which increase hybridization energy.

This list is not meant to be limiting.

**[0030]** The term “wobble base” as used herein refers to a change in a one or more nucleotide bases of a reference nucleotide sequence wherein the change does not change the sequence of the amino acid coded by the nucleotide relative to the reference sequence.

**[0031]** The term “non-nucleotide” or “abasic nucleotide” as use herein refers to any group or compound which can be incorporated into a nucleic acid chain in the place of one or more nucleotide units, including either sugar and/or phosphate substitutions, and allows the remaining bases to exhibit their enzymatic activity. The group or compound is abasic in that it does not contain a commonly recognized nucleotide base, such as adenosine, guanine, cytosine, uracil or thymine. It may have substitutions for a 2' or 3' H or OH as described in the art and herein.

**[0032]** As described herein, in certain embodiments GRON quality and conversion efficiency may be improved by synthesizing all or a portion of the GRON using nucleotide multimers, such as dimers, trimers, tetramers, etc. improving its purity.

**[0033]** In certain embodiments, the target deoxyribonucleic acid (DNA) sequence is within a plant cell, for example the target DNA sequence is in the plant cell genome. The plant cell may be non-transgenic or transgenic, and the target DNA sequence may be a transgene or an endogenous gene of the plant cell.

[0034] In certain embodiments, the conditions that increase one or more cellular DNA repair processes comprise introducing one or more compounds which induce single or double DNA strand breaks into the plant cell prior to, or coincident to, or after delivering the GRON into the plant cell. Exemplary compounds are described herein.

[0035] The methods and compositions described herein are applicable to plants generally. By way of example only, a plant species may be selected from the group consisting of canola, sunflower, corn, tobacco, sugar beet, cotton, maize, wheat (including but not limited to *Triticum spp.*, *Triticum aestivum*, *Triticum durum*, *Triticum timopheevii*, *Triticum monococcum*, *Triticum spelta*, *Triticum zhukovskyi* and *Triticum urartu* and hybrids thereof), barley (including but not limited to *Hordeum vulgare* L., *Hordeum comosum*, *Hordeum depressum*, *Hordeum intercedens*, *Hordeum jubatum*, *Hordeum marinum*, *Hordeum marinum*, *Hordeum parodii*, *Hordeum pusillum*, *Hordeum secalinum*, and *Hordeum spontaneum*), rice (including but not limited to *Oryza sativa* subsp. indica, *Oryza sativa* subsp. japonica, *Oryza sativa* subsp. javanica, *Oryza sativa* subsp. glutinosa (glutinous rice), *Oryza sativa* Aromatica group (e.g., basmati), and *Oryza sativa* (floating rice group)), alfalfa, barley, sorghum, tomato, mango, peach, apple, pear, strawberry, banana, melon, cassava, potato, carrot, lettuce, onion, soy bean, soya spp, sugar cane, pea, chickpea, field pea, fava bean, lentils, turnip, rutabaga, brussel sprouts, lupin, cauliflower, kale, field beans, poplar, pine, eucalyptus, grape, citrus, triticale, alfalfa, rye (including but not limited to *Secale sylvestre*, *Secale strictum*, *Secale cereale*, *Secale vavilovii*, *Secale africanum*, *Secale ciliatoglume*, *Secale ancestrale*, and *Secale montanum*), oats, turf (including but not limited to Turf grass include *Zoysia japonica*, *Agrostis palustris*, *Poa pratensis*, *Poa annua*, *Digitaria sanguinalis*, *Cyperus rotundus*, *Kyllinga brevifolia*, *Cyperus amuricus*, *Erigeron canadensis*, *Hydrocotyle sibthorpioides*, *Kummerowia striata*, *Euphorbia humifusa*, and *Viola arvensis*) and forage grasses, flax, oilseed rape, cotton, mustard, cucumber, morning glory, balsam, pepper, eggplant, marigold, lotus, cabbage, daisy, carnation, tulip, iris, lily, nut-producing plants insofar as they are not already specifically mentioned. These may also apply in whole or in part to all other biological systems including but not limited to bacteria, yeast, fungi, algae, and mammalian cells and even their organelles (e.g., mitochondria and chloroplasts). In some embodiments, the organism or cell is of a species selected from the group consisting of *Escherichia coli*, *Mycobacterium smegmatis*, *Bacillus subtilis*, *Chlorella*, *Bacillus thuringiensis*, *Saccharomyces cerevisiae*, *Yarrowia lipolytica*, *Chlamydomonas*

*rhienhardtii*, *Pichia pastoris*, *Corynebacterium*, *Aspergillus niger*, and *Neurospora crassa*. In some embodiments, the yeast is *Yarrowia lypolitica*. In other embodiments, the yeast is not *Saccharomyces cerevisiae*. In some embodiments, the plant or plant cell is of a species selected from the group consisting of *Arabidopsis thaliana*, *Solanum tuberosum*, *Solanum phureja*, *Oryza sativa*, *Glycine max*, *Amaranthus tuberculatus*, *Linum usitatissimum*, and *Zea mays*. The plant species may be selected from the group consisting of monocotyledonous plants of the grass family Poaceae. The family Poaceae may be divided into two major clades, the clade containing the subfamilies Bambusoideae, Ehrhartoideae, and Pooideae (the BEP clade) and the clade containing the subfamilies Panicoideae, Arundinoideae, Chloridoideae, Centothecoideae, Micrairoideae, Aristidoideae, and Danthonioideae (the PACCMAD clade). The subfamily Bambusoideae includes tribe *Oryzae*. The plant species may relate to plants of the BEP clade, in particular plants of the subfamilies Bambusoideae and Ehrhartoideae. The BET clade includes subfamilies Bambusoideae, Ehrhartoideae, and group Triticoideae and no other subfamily Pooideae groups. BET crop plants are plants grown for food or forage that are members of BET subclade, for example barley, corn, etc.

[0036] In certain embodiments, the methods further comprise regenerating a plant having a mutation introduced by the GRON from the plant cell, and may comprise collecting seeds from the plant.

[0037] In related aspects, the present disclosure relates to plant cells comprising a genomic modification introduced by a GRON according to the methods described herein, a plant comprising a genomic modification introduced by a GRON according to the methods described herein, or a seed comprising a genomic modification introduced by a GRON according to the methods described herein; or progeny of a seed comprising a genomic modification introduced by a GRON according to the methods described herein.

[0038] Other embodiments of the disclosure will be apparent from the following detailed description, exemplary embodiments, and claims.

### BRIEF DESCRIPTION OF THE FIGURES

[0039] Fig. 1 depicts BFP to GFP conversion mediated by phosphothioate (PS) labeled GRONs (having 3 PS moieties at each end of the GRON) and 5'Cy3/ 3'idC labeled GRONs.

[0040] Fig. 2 depicts GRONs comprising RNA/DNA, referred to herein as “2'-O-methyl GRONs.”

[0041] Fig. 3 is a schematic of the location on the *bfp* gene where the BFP5 CRISPRs target.

[0042] Fig. 4 shows the results of the effect of CRISPRs introduced with either the Cy3 or 3PS GRONs at various lengths, on the percentage of BFP to GFP conversion in a BFP transgenic *Arabidopsis thaliana* model system.

[0043] Fig. 5 shows the results of the effect of CRISPRs introduced with the 3PS GRONs at various lengths, on the percentage of BFP to GFP conversion in a BFP transgenic *Arabidopsis thaliana* model system.

[0044] Fig. 6 is a schematic of the GRON designs used in Example 9.

[0045] Fig. 7 shows the measurement of mean percentage GFP positive protoplasts from an *Arabidopsis thaliana* BFP transgenic model system as determined by flow cytometry from 71-mer GRONs.

[0046] Fig. 8 shows the measurement of GFP positive protoplasts from *Arabidopsis thaliana* BFP transgenic model system as determined by flow cytometry from 201-mer GRONs.

[0047] Fig. 9 shows the effect of CRISPRs introduced with coding and non-coding GRONs on the mean percentage of GFP positive cells in a BFP transgenic *Arabidopsis thaliana* model system.

[0048] Fig. 10 is a schematic of tethering a single stranded GRON or double stranded DNA to the CRISPR/Cas complex.

[0049] Fig. 11 shows the results of the effect of CRISPRs and GRONs in mediating BFP to GFP conversion in a BFP transgenic *Arabidopsis thaliana* model system of spacers of differing lengths.

[0050] Fig. 12 shows the results of the effect of CRISPRs and GRONs in mediating BFP to GFP conversion in a BFP transgenic *Arabidopsis thaliana* model system of spacers were encoded on a plasmid (gRNA plasmid) or used as an amplicon (gRNA amplicon).

[0051] Fig. 13 shows the results of the effect of CRISPRs and GRONs in mediating BFP to GFP conversion in a BFP transgenic *Arabidopsis thaliana* model system of unmodified vs. 3PS modified 41-mer GRONs.

[0052] Fig. 14 shows the results of next generation sequencing of 3- and 6-week old *Linum usitatissimum* (flax) microcalli derived from shoot tip protoplasts PEG treated with CRISPR plasmid at T=0.

[0053] Fig. 15 shows the results of next generation sequencing of 3- and 6-week old *Linum usitatissimum* microcalli derived from shoot tip protoplasts PEG treated with CRISPR plasmid at T=0.

[0054] Fig. 16 shows CRISPR-Cas nuclease activity and gene editing in *A. thaliana* protoplasts. 16A: Distribution of indels based on size as determined by deep sequencing in protoplasts treated with CRISPR-Cas plasmid (BC-1) at 72 h post delivery. Indels represented 0.79% of the total reads. 16B: BFP to GFP editing measured by the percentage of GFP fluorescing protoplasts identified by flow cytometry 72 h post delivery of plasmid (BC-1) and GRON (CG-6). Represented data is not normalized for transfection efficiency. Error bars are s.e.m. ( $n = 9$ ).

[0055] Fig. 17 shows gene editing in *A. thaliana* protoplasts using CRISPR-Cas combined with GRONs of differing lengths and chemical modifications. 17a: Comparison of 3PS and unmodified GRONs in BFP to GFP gene editing as measured by flow cytometry at 72 h after delivery of plasmid (BC-1) and GRONs (CG-1) or (CG-2). 17b: Comparison of GRON lengths in BFP to GFP gene editing as measured by flow cytometry at 72 hours post delivery of plasmid (BC-2) and GRONs (CG-5) or (CG-8). 17c: Comparing 3PS to 2'-O-Me GRONs for BFP to GFP gene editing as measured by flow cytometry at 72 h post delivery of plasmid (BC-1) and GRONs (CG-6), (CG-9) or (CG-10). 17d: Comparing 3PS- to Cy3GRONs in BFP to GFP gene editing as measured by flow cytometry at 72 h post delivery of plasmid (BC-3) and GRONs (CG-3) or (CG-4). Error bars are s.e.m. ( $n = 3$ ). (CG-1): BFP antisense 41 nb unmodified; (CG-2): BFP antisense 41 nb 3PS modified; (CG-3): BFP sense 41 nb 3PS modified; (CG-4): BFP sense 41 nb Cy3 modified; (CG-5): BFP sense 60 nb 3PS modified; (CG-6): BFP antisense 201 nb 3PS modified; (CG-8): BFP sense 201 nb 3PS modified; (CG-9): BFP antisense 201 nb 2'-O-Me modification on the first 5' RNA base; (CG-10): BFP antisense 201 nb 2'-O-Me modifications on the first nine 5' RNA bases.

[0056] Fig. 18 shows TALEN nuclease activity and gene editing in *A. thaliana* and *L. usitatissimum* protoplasts. 18a: Distribution of indels based on size as determined by deep sequencing in *Arabidopsis* protoplasts treated with TALEN plasmid (BT-1) at 72 h post delivery. Indels represented 0.51% of the total reads. 18b: *BFP* to *GFP* gene editing as measured by flow cytometry at 48 h post delivery of plasmid (BT-1) and GRON (CG-7). 18c: Representative distribution of indels based on bp length in *L. usitatissimum* protoplasts treated with a TALEN (LuET-1) targeting the EPSPS genes 7 d after delivery. Total frequency of indels is 0.50%. d, *L. usitatissimum* EPSPS gene editing as measured by deep sequencing at 7 d post delivery of plasmid (LuET-1) and GRON (CG-11) into protoplasts. Percentage of total reads represents the number of reads containing both T97I and P101A edits as a percentage of the total reads. Error bars are s.e.m. ( $n = 3$ ). (CG-7): *BFP* sense 201 nb 3PS modified; (CG-11): EPSPS sense 144 nb Cy3 modified.

[0057] Fig. 19 shows effects of the double strand break inducing antibiotics zeocin and phleomycin on *BFP* to *GFP* editing in transgenic *A. thaliana* protoplasts. Protoplasts were treated with zeocin or phleomycin for 90 min before PEG introduction of GRON (CG2). Successful editing resulted in *GFP* fluorescence. Green fluorescing protoplasts were quantified using an Attune Acoustic Focusing Cytometer.

[0058] Fig. 20 shows converted BFP transgenic *A. thaliana* cells five days after GRON delivery into BFP transgenic protoplasts, targeting the conversion from BFP to GFP. Green fluorescence is indicative of BFP-GFP editing. A brightfield image; B, the same field of view in blue light. Error bars are s.e.m. ( $n = 4$ ); (CG2): BFP antisense 41 nb 3PS modified; (CG12) BFP antisense 41 nb 3PS modified non-targeting. Images were acquired with an ImageXpress Micro system (Molecular Devices, Sunnyvale, CA, USA) Scale bar = 20  $\mu\text{m}$

[0059] Fig. 21 shows CRISPR-Cas and TALEN designs. 21A: Schematic of the CRISPR-Cas plasmid. The mannopine synthase (Mas) promoter is driving the transcription of the *Cas9* gene that is codon optimized for higher plants. The *Cas9* gene contains two SV40 nuclear localization signals (NLS) at either end of the gene and a 2X FLAG epitope tag. *A. thaliana* U6 promoter is driving the transcription of the gRNA scaffold and transcription is terminated using a poly(T) signal. 21B: Schematic of the TALEN plasmid. The Mas promoter is driving the transcription of the right and left tale arms linked together with a 2A ribosome skipping sequence. A FokI endonuclease is linked to the 3' end of each Tale arm. The 5' end of the left tale contains a nuclear

localization signal (NLS) and a V5 epitope tag. *rbcT* is the *Pisum sativum* RBCSE9 gene terminator.

[0060] Fig. 22 shows BFP and EPSPS nuclease target regions. a, *BFP* gene target region for the CRISPR-Cas protospacers, BC-1, BC-2 and BC-3 and the TALEN BT-1, left and right tale arms. The PAM sequence is shown in red. TALEN binding sites are bold and underlined. The site of *BFP* to *GFP* editing CAC→TAC (H66Y) is in bold green. b, *EPSPS* gene target region for the TALEN, LuET-1, left and right tale arms. The site of *EPSPS* conversions ACA>ATA and CCG>GCG (T97I and P101A) are in green.

[0061] Fig. 23 shows the amino acid sequence of *Alopecurus myosuroides* (blackgrass) ACCase gene product (SEQ ID NO:1).

[0062] Fig. 24 shows the amino acid sequence of *Escherichia coli* EPSPS gene product (SEQ ID NO:2).

[0063] Fig. 25 shows exemplary analogous EPSPS positions.

[0064] Fig. 26 shows exemplary ACCase sequences.

#### DETAILED DESCRIPTION

[0065] Targeted genetic modification mediated by oligonucleotides is a valuable technique for use in the specific alteration of short stretches of DNA to create deletions, short insertions, and point mutations. These methods involve DNA pairing/annealing, followed by a DNA repair/recombination event. First, the nucleic acid anneals with its complementary strand in the double-stranded DNA in a process mediated by cellular protein factors. This annealing creates a centrally located mismatched base pair (in the case of a point mutation), resulting in a structural perturbation that most likely stimulates the endogenous protein machinery to initiate the second step in the repair process: site-specific modification of the chromosomal sequence and/or that in organelles (e.g., mitochondria and chloroplasts). This newly introduced mismatch induces the DNA repair machinery to perform a second repair event, leading to the final revision of the target site. The present methods and compositions in various aspects and embodiments disclosed herein, may improve the methods by providing novel approaches which increase the availability of DNA repair components, thus increasing the efficiency and reproducibility of gene repair-mediated modifications to targeted nucleic acids.

**[0066]** Efficient methods for site-directed genomic modifications are desirable for research, clinical gene therapy, industrial microbiology and agriculture. One approach utilizes triplex-forming oligonucleotides (TFO) which bind as third strands to duplex DNA in a sequence-specific manner, to mediate directed mutagenesis. Such TFO can act either by delivering a tethered mutagen, such as psoralen or chlorambucil (Havre et al., Proc Nat'l Acad Sci, U.S.A. 90:7879-7883, 1993; Havre et al., J Virol 67:7323-7331, 1993; Wang et al., Mol Cell Biol 15:1759-1768, 1995; Takasugi et al., Proc Nat'l Acad Sci, U.S.A. 88:5602-5606, 1991; Belousov et al., Nucleic Acids Res 25:3440-3444, 1997), or by binding with sufficient affinity to provoke error-prone repair (Wang et al., Science 271:802-805, 1996).

**[0067]** Another strategy for genomic modification involves the induction of homologous recombination between an exogenous DNA fragment and the targeted gene. This approach has been used successfully to target and disrupt selected genes in mammalian cells and has enabled the production of transgenic mice carrying specific gene knockouts (Capecchi et al., Science 244:1288-1292, 1989; Wagner, U.S. Pat. No. 4,873,191). This approach involves the transfer of selectable markers to allow isolation of the desired recombinants. Without selection, the ratio of homologous to non-homologous integration of transfected DNA in typical gene transfer experiments is low, usually in the range of 1:1000 or less (Sedivy et al., Gene Targeting, W. H. Freeman and Co., New York, 1992). This low efficiency of homologous integration limits the utility of gene transfer for experimental use or gene therapy. The frequency of homologous recombination can be enhanced by damage to the target site from UV irradiation and selected carcinogens (Wang et al., Mol Cell Biol 8:196-202, 1988) as well as by site-specific endonucleases (Sedivy et al., Gene Targeting, W. H. Freeman and Co., New York, 1992; Rouet et al., Proc Nat'l Acad Sci, U.S.A. 91:6064-6068, 1994; Segal et al., Proc Nat'l Acad Sci, U.S.A. 92:806-810, 1995). In addition, DNA damage induced by triplex-directed psoralen photoadducts can stimulate recombination within and between extrachromosomal vectors (Segal et al., Proc Nat'l Acad Sci, U.S.A. 92:806-810, 1995; Faruqi et al., Mol Cell Biol 16:6820-6828, 1996; Glazer, U.S. Pat. No. 5,962,426).

**[0068]** Linear donor fragments are more recombinogenic than their circular counterparts (Folger et al., Mol Cell Biol 2:1372-1387, 1982). Recombination can in certain embodiments also be influenced by the length of uninterrupted homology between both the donor and target sites, with short fragments often appearing to be ineffective

substrates for recombination (Rubnitz et al., Mol Cell Biol 4:2253-2258, 1984). Nonetheless, the use of short fragments of DNA or DNA/RNA hybrids for gene correction is the focus of various strategies. (Kunzelmann et al., Gene Ther 3:859-867, 1996).

**[0069]** “Nucleic acid sequence,” “nucleotide sequence” and “polynucleotide sequence” as used herein refer to an oligonucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense or antisense strand.

**[0070]** As used herein, the terms “oligonucleotide” and “oligomer” refer to a polymer of nucleobases of at least about 10 nucleobases and as many as about 1000 nucleobases.

**[0071]** The terms “DNA-modifying molecule” and “DNA-modifying reagent” as used herein refer to a molecule which is capable of recognizing and specifically binding to a nucleic acid sequence in the genome of a cell, and which is capable of modifying a target nucleotide sequence within the genome, wherein the recognition and specific binding of the DNA-modifying molecule to the nucleic acid sequence is protein-independent. The term “protein-independent” as used herein in connection with a DNA-modifying molecule means that the DNA-modifying molecule does not require the presence and/or activity of a protein and/or enzyme for the recognition of, and/or specific binding to, a nucleic acid sequence. DNA-modifying molecules are exemplified, but not limited to triplex forming oligonucleotides, peptide nucleic acids, polyamides, and oligonucleotides which are intended to promote gene conversion. The DNA-modifying molecules of the present disclosure are in certain embodiments distinguished from the prior art's nucleic acid sequences which are used for homologous recombination (Wong & Capecchi, Molec. Cell. Biol. 7:2294-2295, 1987) in that the prior art's nucleic acid sequences which are used for homologous recombination are protein-dependent. The term “protein-dependent” as used herein in connection with a molecule means that the molecule requires the presence and/or activity of a protein and/or enzyme for the recognition of, and/or specific binding of the molecule to, a nucleic acid sequence. Methods for determining whether a DNA-modifying molecule requires the presence and/or activity of a protein and/or enzyme for the recognition of, and/or specific binding to, a nucleic acid sequence are within the skill in the art (see, e.g., Dennis et al. Nucl. Acids Res. 27:4734-4742, 1999). For example, the DNA-modifying molecule may be incubated *in vitro* with the nucleic acid sequence in the absence of any proteins and/or enzymes. The detection

of specific binding between the DNA-modifying molecule and the nucleic acid sequence demonstrates that the DNA-modifying molecule is protein-independent. On the other hand, the absence of specific binding between the DNA-modifying molecule and the nucleic acid sequence demonstrates that the DNA-modifying molecule is protein-dependent and/or requires additional factors.

**[0072]** “Triplex forming oligonucleotide” (TFO) is defined as a sequence of DNA or RNA that is capable of binding in the major groove of a duplex DNA or RNA helix to form a triple helix. Although the TFO is not limited to any particular length, a preferred length of the TFO is 250 nucleotides or less, 200 nucleotides or less, or 100 nucleotides or less, or from 5 to 50 nucleotides, or from 10 to 25 nucleotides, or from 15 to 25 nucleotides. Although a degree of sequence specificity between the TFO and the duplex DNA is necessary for formation of the triple helix, no particular degree of specificity is required, as long as the triple helix is capable of forming. Likewise, no specific degree of avidity or affinity between the TFO and the duplex helix is required as long as the triple helix is capable of forming. While not intending to limit the length of the nucleotide sequence to which the TFO specifically binds in one embodiment, the nucleotide sequence to which the TFO specifically binds is from 1 to 100, in some embodiments from 5 to 50, yet other embodiments from 10 to 25, and in other embodiments from 15 to 25, nucleotides. Additionally, “triple helix” is defined as a double-helical nucleic acid with an oligonucleotide bound to a target sequence within the double-helical nucleic acid. The “double-helical” nucleic acid can be any double-stranded nucleic acid including double-stranded DNA, double-stranded RNA and mixed duplexes of DNA and RNA. The double-stranded nucleic acid is not limited to any particular length. However, in preferred embodiments it has a length of greater than 500 bp, in some embodiments greater than 1 kb and in some embodiments greater than about 5 kb. In many applications the double-helical nucleic acid is cellular, genomic nucleic acid. The triplex forming oligonucleotide may bind to the target sequence in a parallel or anti-parallel manner.

**[0073]** “Peptide Nucleic Acids,” “polyamides” or “PNA” are nucleic acids wherein the phosphate backbone is replaced with an N-aminoethylglycine-based polyamide structure. PNAs have a higher affinity for complementary nucleic acids than their natural counterparts following the Watson-Crick base-pairing rules. PNAs can form highly stable triple helix structures with DNA of the following stoichiometry: (PNA)<sub>2</sub>.DNA. Although the peptide nucleic acids and polyamides are not limited to any particular length, a preferred

length of the peptide nucleic acids and polyamides is 200 nucleotides or less, in some embodiments 100 nucleotides or less, and in some embodiments from 5 to 50 nucleotides long. While not intending to limit the length of the nucleotide sequence to which the peptide nucleic acid and polyamide specifically binds, in one embodiment, the nucleotide sequence to which the peptide nucleic acid and polyamide specifically bind is from 1 to 100, in some embodiments from 5 to 50, yet other embodiments from 5 to 25, and other embodiments from 5 to 20, nucleotides.

**[0074]** The term “cell” refers to a single cell. The term “cells” refers to a population of cells. The population may be a pure population comprising one cell type. Likewise, the population may comprise more than one cell type. In the present disclosure, there is no limit on the number of cell types that a cell population may comprise. A cell as used herein includes without limitation plant callus cells, cells with and without cell walls, prokaryotic cells and eukaryotic cells.

**[0075]** The term “synchronize” or “synchronized,” when referring to a sample of cells, or “synchronized cells” or “synchronized cell population” refers to a plurality of cells which have been treated to cause the population of cells to be in the same phase of the cell cycle. It is not necessary that all of the cells in the sample be synchronized. A small percentage of cells may not be synchronized with the majority of the cells in the sample. A preferred range of cells that are synchronized is between 10-100%. A more preferred range is between 30-100%. Also, it is not necessary that the cells be a pure population of a single cell type. More than one cell type may be contained in the sample. In this regard, only one of cell types may be synchronized or may be in a different phase of the cell cycle as compared to another cell type in the sample.

**[0076]** The term “synchronized cell” when made in reference to a single cell means that the cell has been manipulated such that it is at a cell cycle phase which is different from the cell cycle phase of the cell prior to the manipulation. Alternatively, a “synchronized cell” refers to a cell that has been manipulated to alter (i.e., increase or decrease) the duration of the cell cycle phase at which the cell was prior to the manipulation when compared to a control cell (e.g., a cell in the absence of the manipulation).

**[0077]** The term “cell cycle” refers to the physiological and morphological progression of changes that cells undergo when dividing (i.e. proliferating). The cell cycle is generally recognized to be composed of phases termed “interphase,” “prophase,”

“metaphase,” “anaphase,” and “telophase”. Additionally, parts of the cell cycle may be termed “M (mitosis),” “S (synthesis),” “G0,” “G1 (gap 1)” and “G2 (gap2)”. Furthermore, the cell cycle includes periods of progression that are intermediate to the above named phases.

**[0078]** The term “cell cycle inhibition” refers to the cessation of cell cycle progression in a cell or population of cells. Cell cycle inhibition is usually induced by exposure of the cells to an agent (chemical, proteinaceous or otherwise) that interferes with aspects of cell physiology to prevent continuation of the cell cycle.

**[0079]** “Proliferation” or “cell growth” refers to the ability of a parent cell to divide into two daughter cells repeatably thereby resulting in a total increase of cells in the population. The cell population may be in an organism or in a culture apparatus.

**[0080]** The term “capable of modifying DNA” or “DNA modifying means” refers to procedures, as well as endogenous or exogenous agents or reagents that have the ability to induce, or can aid in the induction of, changes to the nucleotide sequence of a targeted segment of DNA. Such changes may be made by the deletion, addition or substitution of one or more bases on the targeted DNA segment. It is not necessary that the DNA sequence changes confer functional changes to any gene encoded by the targeted sequence. Furthermore, it is not necessary that changes to the DNA be made to any particular portion or percentage of the cells.

**[0081]** The term “nucleotide sequence of interest” refers to any nucleotide sequence, the manipulation of which may be deemed desirable for any reason, by one of ordinary skill in the art. Such nucleotide sequences include, but are not limited to, coding sequences of structural genes (e.g., reporter genes, selection marker genes, oncogenes, drug resistance genes, growth factors, etc.), and non-coding regulatory sequences that do not encode an mRNA or protein product (e.g., promoter sequence, enhancer sequence, polyadenylation sequence, termination sequence, regulatory RNAs such as miRNA, etc.).

**[0082]** “Amino acid sequence,” “polypeptide sequence,” “peptide sequence” and “peptide” are used interchangeably herein to refer to a sequence of amino acids.

**[0083]** “Target sequence,” as used herein, refers to a double-helical nucleic acid comprising a sequence greater than 8 nucleotides in length but less than 201 nucleotides in length. In some embodiments, the target sequence is between 8 to 30 bases. The target

sequence, in general, is defined by the nucleotide sequence on one of the strands on the double-helical nucleic acid.

**[0084]** As used herein, a “purine-rich sequence” or “polypurine sequence” when made in reference to a nucleotide sequence on one of the strands of a double-helical nucleic acid sequence is defined as a contiguous sequence of nucleotides wherein greater than 50% of the nucleotides of the target sequence contain a purine base. However, it is preferred that the purine-rich target sequence contain greater than 60% purine nucleotides, in some embodiments greater than 75% purine nucleotides, in other embodiments greater than 90% purine nucleotides and yet other embodiments 100% purine nucleotides.

**[0085]** As used herein, a “pyrimidine-rich sequence” or “polypyrimidine sequence” when made in reference to a nucleotide sequence on one of the strands of a double-helical nucleic acid sequence is defined as a contiguous sequence of nucleotides wherein greater than 50% of the nucleotides of the target sequence contain a pyrimidine base. However, it is preferred that the pyrimidine-rich target sequence contain greater than 60% pyrimidine nucleotides and in some embodiments greater than 75% pyrimidine nucleotides. In some embodiments, the sequence contains greater than 90% pyrimidine nucleotides and, in other embodiments, is 100% pyrimidine nucleotides.

**[0086]** A “variant” of a first nucleotide sequence is defined as a nucleotide sequence which differs from the first nucleotide sequence (e.g., by having one or more deletions, insertions, or substitutions that may be detected using hybridization assays or using DNA sequencing). Included within this definition is the detection of alterations or modifications to the genomic sequence of the first nucleotide sequence. For example, hybridization assays may be used to detect (1) alterations in the pattern of restriction enzyme fragments capable of hybridizing to the first nucleotide sequence when comprised in a genome (i.e., RFLP analysis), (2) the inability of a selected portion of the first nucleotide sequence to hybridize to a sample of genomic DNA which contains the first nucleotide sequence (e.g., using allele-specific oligonucleotide probes), (3) improper or unexpected hybridization, such as hybridization to a locus other than the normal chromosomal locus for the first nucleotide sequence (e.g., using fluorescent in situ hybridization (FISH) to metaphase chromosome spreads, etc.). One example of a variant is a mutated wild type sequence.

[0087] The terms “nucleic acid” and “unmodified nucleic acid” as used herein refer to any one of the known four deoxyribonucleic acid bases (i.e., guanine, adenine, cytosine, and thymine). The term “modified nucleic acid” refers to a nucleic acid whose structure is altered relative to the structure of the unmodified nucleic acid. Illustrative of such modifications would be replacement covalent modifications of the bases, such as alkylation of amino and ring nitrogens as well as saturation of double bonds.

[0088] As used herein, the terms “mutation” and “modification” and grammatical equivalents thereof when used in reference to a nucleic acid sequence are used interchangeably to refer to a deletion, insertion, substitution, strand break, and/or introduction of an adduct. A “deletion” is defined as a change in a nucleic acid sequence in which one or more nucleotides is absent. An “insertion” or “addition” is that change in a nucleic acid sequence which has resulted in the addition of one or more nucleotides. A “substitution” results from the replacement of one or more nucleotides by a molecule which is a different molecule from the replaced one or more nucleotides. For example, a nucleic acid may be replaced by a different nucleic acid as exemplified by replacement of a thymine by a cytosine, adenine, guanine, or uridine. Pyrimidine to pyrimidine (e.g. C to T or T to C nucleotide substitutions) or purine to purine (e.g. G to A or A to G nucleotide substitutions) are termed transitions, whereas pyrimidine to purine or purine to pyrimidine (e.g. G to T or G to C or A to T or A to C) are termed transversions. Alternatively, a nucleic acid may be replaced by a modified nucleic acid as exemplified by replacement of a thymine by thymine glycol. Mutations may result in a mismatch. The term “mismatch” refers to a non-covalent interaction between two nucleic acids, each nucleic acid residing on a different polynucleic acid sequence, which does not follow the base-pairing rules. For example, for the partially complementary sequences 5'-AGT-3' and 5'-AAT-3', a G-A mismatch (a transition) is present. The terms “introduction of an adduct” or “adduct formation” refer to the covalent or non-covalent linkage of a molecule to one or more nucleotides in a DNA sequence such that the linkage results in a reduction (in some embodiments from 10% to 100%, in other embodiments from 50% to 100%, and in some embodiments from 75% to 100%) in the level of DNA replication and/or transcription.

[0089] The term “DNA cutter” refers to a moiety that effects a strand break. Non-limited examples include meganucleases, TALEs/TALENs, antibiotics, zinc fingers and CRISPRs or CRISPR/cas systems.

[0090] The term “strand break” when made in reference to a double stranded nucleic acid sequence includes a single-strand break and/or a double-strand break. A single-strand break (a nick) refers to an interruption in one of the two strands of the double stranded nucleic acid sequence. This is in contrast to a double-strand break which refers to an interruption in both strands of the double stranded nucleic acid sequence, which may result in blunt or staggered ends. Strand breaks may be introduced into a double stranded nucleic acid sequence either directly (e.g., by ionizing radiation or treatment with certain chemicals) or indirectly (e.g., by enzymatic incision at a nucleic acid base).

[0091] The terms “mutant cell” and “modified cell” refer to a cell which contains at least one modification in the cell's genomic sequence.

[0092] The term “portion” when used in reference to a nucleotide sequence refers to fragments of that nucleotide sequence. The fragments may range in size from 5 nucleotide residues to the entire nucleotide sequence minus one nucleic acid residue.

[0093] DNA molecules are said to have “5' ends” and “3' ends” because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage. Therefore, an end of an oligonucleotide is referred to as the “5' end” if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring. An end of an oligonucleotide is referred to as the “3' end” if its 3' oxygen is not linked to a 5' phosphate of another mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide, also may be said to have 5' and 3' ends. In either a linear or circular DNA molecule, discrete elements are referred to as being “upstream” or 5' of the “downstream” or 3' elements. This terminology reflects that transcription proceeds in a 5' to 3' direction along the DNA strand. The promoter and enhancer elements which direct transcription of a linked gene are generally located 5' or upstream of the coding region. However, enhancer elements can exert their effect even when located 3' of the promoter element and the coding region. Transcription termination and polyadenylation signals are located 3' or downstream of the coding region.

[0094] The term “recombinant DNA molecule” as used herein refers to a DNA molecule which is comprised of segments of DNA joined together by means of molecular biological techniques.

[0095] The term “recombinant protein” or “recombinant polypeptide” as used herein refers to a protein molecule which is expressed using a recombinant DNA molecule.

[0096] As used herein, the terms “vector” and “vehicle” are used interchangeably in reference to nucleic acid molecules that transfer DNA segment(s) from one cell to another.

[0097] The terms “in operable combination,” “in operable order” and “operably linked” as used herein refer to the linkage of nucleic acid sequences in such a manner that a nucleic acid molecule capable of directing the transcription of a given gene and/or the synthesis of a desired protein molecule is produced. The terms also refer to the linkage of amino acid sequences in such a manner so that a functional protein is produced.

[0098] The term “transfection” as used herein refers to the introduction of foreign DNA into cells. Transfection may be accomplished by a variety of means known to the art including calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection, polybrene-mediated transfection, electroporation, microinjection, liposome fusion, lipofectin, protoplast fusion, retroviral infection, biolistics (i.e., particle bombardment) and the like.

[0099] As used herein, the terms “complementary” or “complementarity” are used in reference to “polynucleotides” and “oligonucleotides” (which are interchangeable terms that refer to a sequence of nucleotides) related by the base-pairing rules. For example, the sequence “5'-CAGT-3',” is complementary to the sequence “5'-ACTG-3'.” Complementarity can be “partial” or “total”. “Partial” complementarity is where one or more nucleic acid bases is not matched according to the base pairing rules. “Total” or “complete” complementarity between nucleic acids is where each and every nucleic acid base is matched with another base under the base pairing rules. The degree of complementarity between nucleic acid strands may have significant effects on the efficiency and strength of hybridization between nucleic acid strands. This may be of particular importance in amplification reactions, as well as detection methods which depend upon binding between nucleic acids. For the sake of convenience, the terms “polynucleotides” and “oligonucleotides” include molecules which include nucleosides.

[0100] The terms “homology” and “homologous” as used herein in reference to nucleotide sequences refer to a degree of complementarity with other nucleotide sequences. There may be partial homology or complete homology (i.e., identity). When

used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term “substantially homologous” refers to any nucleic acid sequence (e.g., probe) which can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described above. A nucleotide sequence which is partially complementary, i.e., “substantially homologous,” to a nucleic acid sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid sequence. The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous sequence to a target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

**[0101]** Low stringency conditions comprise conditions equivalent to binding or hybridization at 68°C. in a solution consisting of 5×SSPE (43.8 g/l NaCl, 6.9 g/l NaH<sub>2</sub>PO<sub>4</sub>•H<sub>2</sub>O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5× Denhardt's reagent (50× Denhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V; Sigma)) and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 2.0×SSPE, 0.1% SDS at room temperature when a probe of about 100 to about 1000 nucleotides in length is employed.

**[0102]** In addition, conditions which promote hybridization under conditions of high stringency (e.g., increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.) are well known in the art. High stringency conditions, when used in reference to nucleic acid hybridization, comprise conditions equivalent to binding or hybridization at 68°C. in a solution consisting of 5×SSPE, 1% SDS, 5×Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1×SSPE and 0.1% SDS at 68°C when a probe of about 100 to about 1000 nucleotides in length is employed.

[0103] It is well known in the art that numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (e.g., the presence or absence of formamide, dextran sulfate, polyethylene glycol), as well as components of the hybridization solution may be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions.

[0104] The term “equivalent” when made in reference to a hybridization condition as it relates to a hybridization condition of interest means that the hybridization condition and the hybridization condition of interest result in hybridization of nucleic acid sequences which have the same range of percent (%) homology. For example, if a hybridization condition of interest results in hybridization of a first nucleic acid sequence with other nucleic acid sequences that have from 50% to 70% homology to the first nucleic acid sequence, then another hybridization condition is said to be equivalent to the hybridization condition of interest if this other hybridization condition also results in hybridization of the first nucleic acid sequence with the other nucleic acid sequences that have from 50% to 70% homology to the first nucleic acid sequence.

[0105] As used herein, the term “hybridization” is used in reference to the pairing of complementary nucleic acids using any process by which a strand of nucleic acid joins with a complementary strand through base pairing to form a hybridization complex. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the  $T_m$  of the formed hybrid, and the G:C ratio within the nucleic acids.

[0106] As used herein the term “hybridization complex” refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary G and C bases and between complementary A and T bases; these hydrogen bonds may be further stabilized by base stacking interactions. The two complementary nucleic acid sequences hydrogen bond in an antiparallel configuration. A hybridization complex may be formed in solution (e.g., Cot or Rot analysis) or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized to a solid support (e.g., a nylon membrane or a nitrocellulose filter as

employed in Southern and Northern blotting, dot blotting or a glass slide as employed in *in situ* hybridization, including FISH (fluorescent *in situ* hybridization)).

[0107] As used herein, the term “T<sub>m</sub>” is used in reference to the “melting temperature.” The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the T<sub>m</sub> of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T<sub>m</sub> value may be calculated by the equation:  $T_m = 81.5 + 0.41(\% G+C)$ , when a nucleic acid is in aqueous solution at 1 M NaCl (see e.g., Anderson and Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization, 1985). Other references include more sophisticated computations which take structural as well as sequence characteristics into account for the calculation of T<sub>m</sub>.

[0108] As used herein the term “stringency” is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. “Stringency” typically occurs in a range from about T<sub>m</sub>-5°C (5°C below the melting temperature of the probe) to about 20°C to 25°C below T<sub>m</sub>. As will be understood by those of skill in the art, a stringent hybridization can be used to identify or detect identical polynucleotide sequences or to identify or detect similar or related polynucleotide sequences.

[0109] The terms “specific binding,” “binding specificity,” and grammatical equivalents thereof when made in reference to the binding of a first nucleotide sequence to a second nucleotide sequence, refer to the preferential interaction between the first nucleotide sequence with the second nucleotide sequence as compared to the interaction between the second nucleotide sequence with a third nucleotide sequence. Specific binding is a relative term that does not require absolute specificity of binding; in other words, the term “specific binding” does not require that the second nucleotide sequence interact with the first nucleotide sequence in the absence of an interaction between the second nucleotide sequence and the third nucleotide sequence. Rather, it is sufficient that the level of interaction between the first nucleotide sequence and the second nucleotide sequence is greater than the level of interaction between the second nucleotide sequence with the third nucleotide sequence. “Specific binding” of a first nucleotide sequence with a second nucleotide sequence also means that the interaction between the first nucleotide sequence and the second nucleotide sequence is dependent upon the presence of a particular structure on or within the first nucleotide sequence; in other words the second

nucleotide sequence is recognizing and binding to a specific structure on or within the first nucleotide sequence rather than to nucleic acids or to nucleotide sequences in general. For example, if a second nucleotide sequence is specific for structure “A” that is on or within a first nucleotide sequence, the presence of a third nucleic acid sequence containing structure A will reduce the amount of the second nucleotide sequence which is bound to the first nucleotide sequence.

**[0110]** As used herein, the term “amplifiable nucleic acid” is used in reference to nucleic acids which may be amplified by any amplification method. It is contemplated that “amplifiable nucleic acid” will usually comprise “sample template.”

**[0111]** The terms “heterologous nucleic acid sequence” or “heterologous DNA” are used interchangeably to refer to a nucleotide sequence which is ligated to a nucleic acid sequence to which it is not ligated in nature, or to which it is ligated at a different location in nature. Heterologous DNA is not endogenous to the cell into which it is introduced, but has been obtained from another cell. Generally, although not necessarily, such heterologous DNA encodes RNA and proteins that are not normally produced by the cell into which it is expressed. Examples of heterologous DNA include reporter genes, transcriptional and translational regulatory sequences, selectable marker proteins (e.g., proteins which confer drug resistance), etc.

**[0112]** “Amplification” is defined as the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction technologies well known in the art (Dieffenbach C W and G S Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y.). As used herein, the term “polymerase chain reaction” (“PCR”) refers to the method of K. B. Mullis U.S. Pat. Nos. 4,683,195, and 4,683,202, which describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. The length of the amplified segment of the desired target sequence is determined by the relative positions of two oligonucleotide primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the “polymerase chain reaction” (“PCR”). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be “PCR amplified.”

[0113] With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (e.g., hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of <sup>32</sup>P-labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any oligonucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications.

[0114] One such preferred method, particularly for commercial applications, is based on the widely used TaqMan® real-time PCR technology, and combines Allele-Specific PCR with a Blocking reagent (ASB-PCR) to suppress amplification of the wildtype allele. ASB-PCR can be used for detection of germ line or somatic mutations in either DNA or RNA extracted from any type of tissue, including formalin-fixed paraffin-embedded tumor specimens. A set of reagent design rules are developed enabling sensitive and selective detection of single point substitutions, insertions, or deletions against a background of wild-type allele in thousand-fold or greater excess. (Morlan J, Baker J, Sinicropi D Mutation Detection by Real-Time PCR: A Simple, Robust and Highly Selective Method. PLoS ONE 4(2): e4584, 2009)

[0115] The terms “reverse transcription polymerase chain reaction” and “RT-PCR” refer to a method for reverse transcription of an RNA sequence to generate a mixture of cDNA sequences, followed by increasing the concentration of a desired segment of the transcribed cDNA sequences in the mixture without cloning or purification. Typically, RNA is reverse transcribed using a single primer (e.g., an oligo-dT primer) prior to PCR amplification of the desired segment of the transcribed DNA using two primers.

[0116] As used herein, the term “primer” refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (i.e., in the presence of nucleotides and of an inducing agent such as DNA polymerase and at a suitable temperature and pH). In some embodiments, the primer is single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. In some embodiments, the primer is an

oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

[0117] As used herein, the term “probe” refers to an oligonucleotide (i.e., a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, which is capable of hybridizing to another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular gene sequences. It is contemplated that any probe used in the present disclosure will be labeled with any “reporter molecule,” so that it is detectable in any detection system, including, but not limited to enzyme (e.g., ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present disclosure be limited to any particular detection system or label.

[0118] As used herein, the terms “restriction endonucleases” and “restriction enzymes” refer to bacterial enzymes, each of which cut or nick double- or single-stranded DNA at or near a specific nucleotide sequence, for example, an endonuclease domain of a type IIS restriction endonuclease (e.g., FokI can be used, as taught by Kim et al., 1996, Proc. Nat’l. Acad. Sci. USA, 6:1 156-60).

[0119] As used herein, the term “an oligonucleotide having a nucleotide sequence encoding a gene” means a nucleic acid sequence comprising the coding region of a gene, i.e. the nucleic acid sequence which encodes a gene product. The coding region may be present in either a cDNA, genomic DNA or RNA form. When present in a DNA form, the oligonucleotide may be single-stranded (i.e., the sense strand) or double-stranded. Additionally “an oligonucleotide having a nucleotide sequence encoding a gene” may include suitable control elements such as enhancers, promoters, splice junctions, polyadenylation signals, etc. if needed to permit proper initiation of transcription and/or correct processing of the primary RNA transcript. Further still, the coding region of the present disclosure may contain endogenous enhancers, splice junctions, intervening sequences, polyadenylation signals, etc.

[0120] Transcriptional control signals in eukaryotes comprise “enhancer” elements. Enhancers consist of short arrays of DNA sequences that interact specifically with

cellular proteins involved in transcription (Maniatis, T. et al., Science 236:1237, 1987). Enhancer elements have been isolated from a variety of eukaryotic sources including genes in plant, yeast, insect and mammalian cells and viruses. The selection of a particular enhancer depends on what cell type is to be used to express the protein of interest.

[0121] The presence of “splicing signals” on an expression vector often results in higher levels of expression of the recombinant transcript. Splicing signals mediate the removal of introns from the primary RNA transcript and consist of a splice donor and acceptor site (Sambrook, J. et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, New York, pp. 16.7-16.8, 1989). A commonly used splice donor and acceptor site is the splice junction from the 16S RNA of SV40.

[0122] Efficient expression of recombinant DNA sequences in eukaryotic cells requires expression of signals directing the efficient termination and polyadenylation of the resulting transcript. Transcription termination signals are generally found downstream of the polyadenylation signal and are a few hundred nucleotides in length. The term “poly A site” or “poly A sequence” as used herein denotes a DNA sequence which directs both the termination and polyadenylation of the nascent RNA transcript. Efficient polyadenylation of the recombinant transcript is desirable as transcripts lacking a poly A tail are unstable and are rapidly degraded. The poly A signal utilized in an expression vector may be “heterologous” or “endogenous.” An endogenous poly A signal is one that is found naturally at the 3' end of the coding region of a given gene in the genome. A heterologous poly A signal is one which is isolated from one gene and placed 3' of another gene.

[0123] The term “promoter,” “promoter element” or “promoter sequence” as used herein, refers to a DNA sequence which when placed at the 5' end of (i.e., precedes) an oligonucleotide sequence is capable of controlling the transcription of the oligonucleotide sequence into mRNA. A promoter is typically located 5' (i.e., upstream) of an oligonucleotide sequence whose transcription into mRNA it controls, and provides a site for specific binding by RNA polymerase and for initiation of transcription.

[0124] The term “promoter activity” when made in reference to a nucleic acid sequence refers to the ability of the nucleic acid sequence to initiate transcription of an oligonucleotide sequence into mRNA.

[0125] The term “tissue specific” as it applies to a promoter refers to a promoter that is capable of directing selective expression of an oligonucleotide sequence to a specific type of tissue in the relative absence of expression of the same oligonucleotide in a different type of tissue. Tissue specificity of a promoter may be evaluated by, for example, operably linking a reporter gene to the promoter sequence to generate a reporter construct, introducing the reporter construct into the genome of a plant or an animal such that the reporter construct is integrated into every tissue of the resulting transgenic animal, and detecting the expression of the reporter gene (e.g., detecting mRNA, protein, or the activity of a protein encoded by the reporter gene) in different tissues of the transgenic plant or animal. Selectivity need not be absolute. The detection of a greater level of expression of the reporter gene in one or more tissues relative to the level of expression of the reporter gene in other tissues shows that the promoter is specific for the tissues in which greater levels of expression are detected.

[0126] The term “cell type specific” as applied to a promoter refers to a promoter which is capable of directing selective expression of an oligonucleotide sequence in a specific type of cell in the relative absence of expression of the same oligonucleotide sequence in a different type of cell within the same tissue. The term “cell type specific” when applied to a promoter also means a promoter capable of promoting selective expression of an oligonucleotide in a region within a single tissue. Again, selectivity need not be absolute. Cell type specificity of a promoter may be assessed using methods well known in the art, e.g., immunohistochemical staining as described herein. Briefly, tissue sections are embedded in paraffin, and paraffin sections are reacted with a primary antibody which is specific for the polypeptide product encoded by the oligonucleotide sequence whose expression is controlled by the promoter. As an alternative to paraffin sectioning, samples may be cryosectioned. For example, sections may be frozen prior to and during sectioning thus avoiding potential interference by residual paraffin. A labeled (e.g., peroxidase conjugated) secondary antibody which is specific for the primary antibody is allowed to bind to the sectioned tissue and specific binding detected (e.g., with avidin/biotin) by microscopy.

[0127] The terms “selective expression,” “selectively express” and grammatical equivalents thereof refer to a comparison of relative levels of expression in two or more regions of interest. For example, “selective expression” when used in connection with tissues refers to a substantially greater level of expression of a gene of interest in a

particular tissue, or to a substantially greater number of cells which express the gene within that tissue, as compared, respectively, to the level of expression of, and the number of cells expressing, the same gene in another tissue (i.e., selectivity need not be absolute). Selective expression does not require, although it may include, expression of a gene of interest in a particular tissue and a total absence of expression of the same gene in another tissue. Similarly, “selective expression” as used herein in reference to cell types refers to a substantially greater level of expression of, or a substantially greater number of cells which express, a gene of interest in a particular cell type, when compared, respectively, to the expression levels of the gene and to the number of cells expressing the gene in another cell type.

**[0128]** The term “contiguous” when used in reference to two or more nucleotide sequences means the nucleotide sequences are ligated in tandem either in the absence of intervening sequences, or in the presence of intervening sequences which do not comprise one or more control elements.

**[0129]** As used herein, the terms “nucleic acid molecule encoding,” “nucleotide encoding,” “DNA sequence encoding” and “DNA encoding” refer to the order or sequence of deoxyribonucleotides along a strand of deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along the polypeptide (protein) chain. The DNA sequence thus codes for the amino acid sequence.

**[0130]** The term “isolated” when used in relation to a nucleic acid, as in “an isolated oligonucleotide” refers to a nucleic acid sequence that is separated from at least one contaminant nucleic acid with which it is ordinarily associated in its natural source. Isolated nucleic acid is nucleic acid present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids are nucleic acids such as DNA and RNA which are found in the state they exist in nature. For example, a given DNA sequence (e.g., a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs which encode a multitude of proteins. However, isolated nucleic acid encoding a polypeptide of interest includes, by way of example, such nucleic acid in cells ordinarily expressing the polypeptide of interest where the nucleic acid is in a chromosomal or extrachromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid or oligonucleotide may

be present in single-stranded or double-stranded form. Isolated nucleic acid can be readily identified (if desired) by a variety of techniques (e.g., hybridization, dot blotting, etc.). When an isolated nucleic acid or oligonucleotide is to be utilized to express a protein, the oligonucleotide will contain at a minimum the sense or coding strand (i.e., the oligonucleotide may be single-stranded). Alternatively, it may contain both the sense and anti-sense strands (i.e., the oligonucleotide may be double-stranded).

**[0131]** As used herein, the term “purified” or “to purify” refers to the removal of one or more (undesired) components from a sample. For example, where recombinant polypeptides are expressed in bacterial host cells, the polypeptides are purified by the removal of host cell proteins thereby increasing the percent of recombinant polypeptides in the sample.

**[0132]** As used herein, the term “substantially purified” refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated, and are at least 60% free, in some embodiments 75% free and other embodiments 90% free from other components with which they are naturally associated. An “isolated polynucleotide” is, therefore, a substantially purified polynucleotide.

**[0133]** As used herein the term “coding region” when used in reference to a structural gene refers to the nucleotide sequences which encode the amino acids found in the nascent polypeptide as a result of translation of a mRNA molecule. The coding region is bounded, in eukaryotes, on the 5' side generally by the nucleotide triplet “ATG” which encodes the initiator methionine and on the 3' side by one of the three triplets which specify stop codons (i.e., TAA, TAG, TGA).

**[0134]** By “coding sequence” is meant a sequence of a nucleic acid or its complement, or a part thereof, that can be transcribed and/or translated to produce the mRNA for and/or the polypeptide or a fragment thereof. Coding sequences include exons in a genomic DNA or immature primary RNA transcripts, which are joined together by the cell's biochemical machinery to provide a mature mRNA. The anti-sense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

**[0135]** By “non-coding sequence” is meant a sequence of a nucleic acid or its complement, or a part thereof that is not transcribed into amino acid *in vivo*, or where tRNA does not interact to place or attempt to place an amino acid. Non-coding sequences

include both intron sequences in genomic DNA or immature primary RNA transcripts, and gene-associated sequences such as promoters, enhancers, silencers, etc.

[0136] As used herein, the term “structural gene” or “structural nucleotide sequence” refers to a DNA sequence coding for RNA or a protein which does not control the expression of other genes. In contrast, a “regulatory gene” or “regulatory sequence” is a structural gene which encodes products (e.g., transcription factors) which control the expression of other genes.

[0137] As used herein, the term “regulatory element” refers to a genetic element which controls some aspect of the expression of nucleic acid sequences. For example, a promoter is a regulatory element which facilitates the initiation of transcription of an operably linked coding region. Other regulatory elements include splicing signals, polyadenylation signals, termination signals, etc.

[0138] As used herein, the term “peptide transcription factor binding site” or “transcription factor binding site” refers to a nucleotide sequence which binds protein transcription factors and, thereby, controls some aspect of the expression of nucleic acid sequences. For example, Sp-1 and AP1 (activator protein 1) binding sites are examples of peptide transcription factor binding sites.

[0139] As used herein, the term “gene” means the deoxyribonucleotide sequences comprising the coding region of a structural gene. A “gene” may also include non-translated sequences located adjacent to the coding region on both the 5' and 3' ends such that the gene corresponds to the length of the full-length mRNA. The sequences which are located 5' of the coding region and which are present on the mRNA are referred to as 5' non-translated sequences. The sequences which are located 3' or downstream of the coding region and which are present on the mRNA are referred to as 3' non-translated sequences. The term “gene” encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed “introns” or “intervening regions” or “intervening sequences.” Introns are segments of a gene which are transcribed into heterogenous nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or “spliced out” from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide. A gene is generally a

single locus. In a normal diploid organism, a gene has two alleles. In tetraploid potato, however, each gene has 4 alleles. In sugarcane, which is dodecaploid there can be 12 alleles per gene. Particular examples include flax which has two EPSPS loci each with two alleles and rice which has a single homomeric plastidal ACCase with two alleles.

**[0140]** In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5' and 3' end of the sequences which are present on the RNA transcript. These sequences are referred to as “flanking” sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers which control or influence the transcription of the gene. The 3' flanking region may contain sequences which direct the termination of transcription, post-transcriptional cleavage and polyadenylation.

**[0141]** A “non-human animal” refers to any animal which is not a human and includes vertebrates such as rodents, non-human primates, ovines, bovines, ruminants, lagomorphs, porcines, caprines, equines, canines, felines, aves, etc. Preferred non-human animals are selected from the order Rodentia. “Non-human animal” additionally refers to amphibians (e.g. *Xenopus*), reptiles, insects (e.g. *Drosophila*) and other non-mammalian animal species.

**[0142]** As used herein, the term “transgenic” refers to an organism or cell that has DNA derived from another organism inserted into which becomes integrated into the genome either of somatic and/or germ line cells of the plant or animal. A “transgene” means a DNA sequence which is partly or entirely heterologous (i.e., not present in nature) to the plant or animal in which it is found, or which is homologous to an endogenous sequence (i.e., a sequence that is found in the animal in nature) and is inserted into the plant' or animal's genome at a location which differs from that of the naturally occurring sequence. Transgenic plants or animals which include one or more transgenes are within the scope of this disclosure. Additionally, a “transgenic” as used herein refers to an organism that has had one or more genes modified and/or “knocked out” (made non-functional or made to function at reduced level, i.e., a “knockout” mutation) by the disclosure's methods, by homologous recombination, TFO mutation or by similar processes. For example, in some embodiments, a transgenic organism or cell includes inserted DNA that includes a foreign promoter and/or coding region.

[0143] A “transformed cell” is a cell or cell line that has acquired the ability to grow in cell culture for multiple generations, the ability to grow in soft agar, and/or the ability to not have cell growth inhibited by cell-to-cell contact. In this regard, transformation refers to the introduction of foreign genetic material into a cell or organism. Transformation may be accomplished by any method known which permits the successful introduction of nucleic acids into cells and which results in the expression of the introduced nucleic acid. “Transformation” includes but is not limited to such methods as transfection, microinjection, electroporation, nucleofection and lipofection (liposome-mediated gene transfer). Transformation may be accomplished through use of any expression vector. For example, the use of baculovirus to introduce foreign nucleic acid into insect cells is contemplated. The term “transformation” also includes methods such as P-element mediated germline transformation of whole insects. Additionally, transformation refers to cells that have been transformed naturally, usually through genetic mutation.

[0144] As used herein “exogenous” means that the gene encoding the protein is not normally expressed in the cell. Additionally, “exogenous” refers to a gene transfected into a cell to augment the normal (i.e. natural) level of expression of that gene.

[0145] A peptide sequence and nucleotide sequence may be “endogenous” or “heterologous” (i.e., “foreign”). The term “endogenous” refers to a sequence which is naturally found in the cell into which it is introduced so long as it does not contain some modification relative to the naturally-occurring sequence. The term “heterologous” refers to a sequence which is not endogenous to the cell into which it is introduced. For example, heterologous DNA includes a nucleotide sequence which is ligated to, or is manipulated to become ligated to, a nucleic acid sequence to which it is not ligated in nature, or to which it is ligated at a different location in nature. Heterologous DNA also includes a nucleotide sequence which is naturally found in the cell into which it is introduced and which contains some modification relative to the naturally-occurring sequence. Generally, although not necessarily, heterologous DNA encodes heterologous RNA and heterologous proteins that are not normally produced by the cell into which it is introduced. Examples of heterologous DNA include reporter genes, transcriptional and translational regulatory sequences, DNA sequences which encode selectable marker proteins (e.g., proteins which confer drug resistance), etc.

**[0146] Constructs**

**[0147]** The nucleic acid molecules disclosed herein (e.g., site specific nucleases, or guide RNA for CRISPRs) can be used in the production of recombinant nucleic acid constructs. In one embodiment, the nucleic acid molecules of the present disclosure can be used in the preparation of nucleic acid constructs, for example, expression cassettes for expression in the plant, microorganism, or animal of interest. This expression may be transient for instance when the construct is not integrated into the host genome or maintained under the control offered by the promoter and the position of the construct within the host's genome if it becomes integrated.

**[0148]** Expression cassettes may include regulatory sequences operably linked to the site specific nuclease or guide RNA sequences disclosed herein. The cassette may additionally contain at least one additional gene to be co-transformed into the organism. Alternatively, the additional gene(s) can be provided on multiple expression cassettes.

**[0149]** The nucleic acid constructs may be provided with a plurality of restriction sites for insertion of the site specific nuclease coding sequence to be under the transcriptional regulation of the regulatory regions. The nucleic acid constructs may additionally contain nucleic acid molecules encoding for selectable marker genes.

**[0150]** Any promoter can be used in the production of the nucleic acid constructs. The promoter may be native or analogous, or foreign or heterologous, to the plant, microbial, or animal host nucleic acid sequences disclosed herein. Additionally, the promoter may be the natural sequence or alternatively a synthetic sequence. Where the promoter is "foreign" or "heterologous" to the plant, microbial, or animal host, it is intended that the promoter is not found in the native plant, microbial, or animal into which the promoter is introduced. As used herein, a chimeric gene comprises a coding sequence operably linked to a transcription initiation region that is heterologous to the coding sequence.

**[0151]** The site directed nuclease sequences disclosed herein may be expressed using heterologous promoters.

**[0152]** Any promoter can be used in the preparation of constructs to control the expression of the site directed nuclease sequences, such as promoters providing for constitutive, tissue-preferred, inducible, or other promoters for expression in plants, microbes, or animals. Constitutive promoters include, for example, the core promoter of the R<sub>syn7</sub> promoter and other constitutive promoters disclosed in WO 99/43 838 and U.S.

Patent No. 6,072,050; the core CaMV 35S promoter (Odell et al. Nature 313:810-812; 1985); rice actin (McElroy et al., Plant Cell 2:163-171, 1990); ubiquitin (Christensen et al., Plant Mol. Biol. 12:619-632, 1989 and Christensen et al., Plant Mol. Biol. 18:675-689, 1992); pEMU (Last et al., Theor. Appl. Genet. 81:581-588, 1991); MAS (Velten et al., EMBO J. 3:2723-2730, 1984); ALS promoter (U.S. Patent No. 5,659,026), and the like. Other constitutive promoters include, for example, U.S. Patent 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.

**[0153]** Tissue-preferred promoters can be utilized to direct site directed nuclease expression within a particular plant tissue. Such tissue-preferred promoters include, but are not limited to, leaf-preferred promoters, root-preferred promoters, seed-preferred promoters, and stem-preferred promoters. Tissue-preferred promoters include Yamamoto et al., Plant J. 12(2):255-265, 1997; Kawamata et al., Plant Cell Physiol. 38(7):792-803, 1997; Hansen et al., Mol. Gen Genet. 254(3):337-343, 1997; Russell et al., Transgenic Res. 6(2):157-168, 1997; Rinehart et al., Plant Physiol. 1 12(3):1331-1341, 1996; Van Camp et al., Plant Physiol. 1 12(2):525-535, 1996; Canevascini et al., Plant Physiol. 112(2): 513-524, 1996; Yamamoto et al., Plant Cell Physiol. 35(5):773-778, 1994; Lam, Results Probl. Cell Differ. 20:181-196, 1994; Orozco et al. Plant Mol Biol. 23(6):1129-1138, 1993; Matsuoka et al., Proc Nat'l. Acad. Sci. USA 90(20):9586- 9590, 1993; and Guevara-Garcia et al., Plant J. 4(3):495-505, 1993.

**[0154]** The nucleic acid constructs may also include transcription termination regions. Where transcription terminations regions are used, any termination region may be used in the preparation of the nucleic acid constructs. For example, the termination region may be derived from another source (i.e., foreign or heterologous to the promoter). Examples of termination regions that are available for use in the constructs of the present disclosure include those from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also Guerineau et al., Mol. Gen. Genet. 262:141-144, 1991; Proudfoot, Cell 64:671-674, 1991; Sanfacon et al., Genes Dev. 5:141-149, 1991; Mogen et al., Plant Cell 2:1261-1272, 1990; Munroe et al., Gene 91:151-158, 1990; Ballas et al., Nucleic Acids Res. 17:7891-7903, 1989; and Joshi et al., Nucleic Acid Res. 15:9627-9639, 1987.

**[0155]** In conjunction with any of the aspects, embodiments, methods and/or compositions disclosed herein, the nucleic acids may be optimized for increased

expression in the transformed plant. That is, the nucleic acids encoding the site directed nuclease proteins can be synthesized using plant-preferred codons for improved expression. See, for example, Campbell and Gowri, (*Plant Physiol.* 92:1-11, 1990) for a discussion of host-preferred codon usage. Methods are available in the art for synthesizing plant-preferred genes. See, for example, U.S. Patent Nos. 5,380,831, and 5,436,391, and Murray et al., *Nucleic Acids Res.* 17:477-498, 1989. See also e.g., Lanza et al., *BMC Systems Biology* 8:33-43, 2014; Burgess-Brown et al., *Protein Expr. Purif.* 59:94-102, 2008; Gustafsson et al., *Trends Biotechnol* 22:346-353, 2004.

**[0156]** In addition, other sequence modifications can be made to the nucleic acid sequences disclosed herein. For example, additional sequence modifications are known to enhance gene expression in a cellular host. These include elimination of sequences encoding spurious polyadenylation signals, exon/intron splice site signals, transposon-like repeats, and other such well-characterized sequences that may be deleterious to gene expression. The G-C content of the sequence may also be adjusted to levels average for a target cellular host, as calculated by reference to known genes expressed in the host cell. In addition, the sequence can be modified to avoid predicted hairpin secondary mRNA structures.

**[0157]** Other nucleic acid sequences may also be used in the preparation of the constructs of the present disclosure, for example to enhance the expression of the site directed nuclease coding sequence. Such nucleic acid sequences include the introns of the maize AdhI, intron1 gene (Callis et al., *Genes and Development* 1:1183-1200, 1987), and leader sequences, (W-sequence) from the Tobacco Mosaic virus (TMV), Maize Chlorotic Mottle Virus and Alfalfa Mosaic Virus (Gallie et al., *Nucleic Acid Res.* 15:8693-8711, 1987; and Skuzeski et al., *Plant Mol. Biol.* 15:65-79, 1990). The first intron from the shrunken-1 locus of maize has been shown to increase expression of genes in chimeric gene constructs. U.S. Pat. Nos. 5,424,412 and 5,593,874 disclose the use of specific introns in gene expression constructs, and Gallie et al. (*Plant Physiol.* 106:929-939, 1994) also have shown that introns are useful for regulating gene expression on a tissue specific basis. To further enhance or to optimize site directed nuclease gene expression, the plant expression vectors disclosed herein may also contain DNA sequences containing matrix attachment regions (MARs). Plant cells transformed with such modified expression systems, then, may exhibit overexpression or constitutive expression of a nucleotide sequence of the disclosure.

[0158] The expression constructs disclosed herein can also include nucleic acid sequences capable of directing the expression of the site directed nuclease sequence to the chloroplast or other organelles and structures in both prokaryotes and eukaryotes. Such nucleic acid sequences include chloroplast targeting sequences that encodes a chloroplast transit peptide to direct the gene product of interest to plant cell chloroplasts. Such transit peptides are known in the art. With respect to chloroplast-targeting sequences, “operably linked” means that the nucleic acid sequence encoding a transit peptide (i.e., the chloroplast-targeting sequence) is linked to the site directed nuclease nucleic acid molecules disclosed herein such that the two sequences are contiguous and in the same reading frame. See, for example, Von Heijne et al., *Plant Mol. Biol. Rep.* 9:104-126, 1991; Clark et al., *J. Biol. Chem.* 264:17544-17550, 1989; Della-Cioppa et al., *Plant Physiol.* 84:965-968, 1987; Romer et al., *Biochem. Biophys. Res. Commun.* 196:1414-1421, 1993; and Shah et al., *Science* 233:478-481, 1986.

[0159] Chloroplast targeting sequences are known in the art and include the chloroplast small subunit of ribulose-1,5-bisphosphate carboxylase (Rubisco) (de Castro Silva Filho et al., *Plant Mol. Biol.* 30:769-780, 1996; Schnell et al., *J. Biol. Chem.* 266(5):3335-3342, 1991); 5- (enolpyruvyl)shikimate-3-phosphate synthase (EPSPS) (Archer et al., *J. Bioenerg. Biomemb.* 22(6):789-810, 1990); tryptophan synthase (Zhao et al., *J. Biol. Chem.* 270(11):6081-6087, 1995); plastocyanin (Lawrence et al., *J. Biol. Chem.* 272(33):20357-20363, 1997); chorismate synthase (Schmidt et al., *J. Biol. Chem.* 268(36):27447-27457, 1993); and the light harvesting chlorophyll a/b binding protein (LHBP) (Lamppa et al., *J. Biol. Chem.* 263:14996-14999, 1988). See also Von Heijne et al., *Plant Mol. Biol. Rep.* 9:104-126, 1991; Clark et al., *J. Biol. Chem.* 264:17544-17550, 1989; Della-Cioppa et al., *Plant Physiol.* 84:965-968, 1987; Romer et al., *Biochem. Biophys. Res. Commun.* 196:1414-1421, 1993; and Shah et al., *Science* 233: 478-481, 1986.

[0160] In conjunction with any of the aspects, embodiments, methods and/or compositions disclosed herein, the nucleic acid constructs may be prepared to direct the expression of the mutant site directed nuclease coding sequence from the plant cell chloroplast. Methods for transformation of chloroplasts are known in the art. See, for example, Svab et al., *Proc. Nat'l. Acad. Sci. USA* 87:8526-8530, 1990; Svab and Maliga, *Proc. Nat'l. Acad. Sci. USA* 90:913-917, 1993; Svab and Maliga, *EMBO J.* 12:601-606, 1993. The method relies on particle gun delivery of DNA containing a selectable marker

and targeting of the DNA to the plastid genome through homologous recombination. Additionally, plastid transformation can be accomplished by transactivation of a silent plastid-borne transgene by tissue-preferred expression of a nuclear-encoded and plastid-directed RNA polymerase. Such a system has been reported in McBride et al. Proc. Nat'l. Acad. Sci. USA 91:7301-7305, 1994.

**[0161]** The nucleic acids of interest to be targeted to the chloroplast may be optimized for expression in the chloroplast to account for differences in codon usage between the plant nucleus and this organelle. In this manner, the nucleic acids of interest may be synthesized using chloroplast-preferred codons. See, for example, U.S. Patent No. 5,380,831.

**[0162]** The nucleic acid constructs can be used to transform plant cells and regenerate transgenic plants comprising the site directed nuclease coding sequences. Numerous plant transformation vectors and methods for transforming plants are available. See, for example, U.S. Patent No. 6,753,458, An, G. et al., *Plant Physiol.*, 81:301-305, 1986; Fry, J. et al., *Plant Cell Rep.* 6:321-325, 1987; Block, M., *Theor. Appl Genet.* 76:767-774, 1988; Hinchee et al., *Stadler. Genet. Symp.* 203212.203-212, 1990; Cousins et al., *Aust. J. Plant Physiol.* 18:481-494, 1991; Chee, P. P. and Slightom, J. L., *Gene*.118:255-260, 1992; Christou et al., *Trends. Biotechnol.* 10:239-246, 1992; D'Halluin et al., *Bio/Technol.* 10:309-314, 1992; Dhir et al., *Plant Physiol.* 99:81-88, 1992; Casas et al., *Proc. Nat'l. Acad. Sci. USA* 90:11212-11216, 1993; Christou, P., *In Vitro Cell. Dev. Biol.-Plant* 29P:119-124, 1993; Davies, et al., *Plant Cell Rep.* 12:180-183, 1993; Dong, J. A. and Mc Hughen, A., *Plant Sci.* 91:139-148, 1993; Franklin, C. I. and Trieu, T. N., *Plant. Physiol.* 102:167, 1993; Golovkin et al., *Plant Sci.* 90:41-52, 1993; Guo *Chin Sci. Bull.* 38:2072-2078; Asano, et al., *Plant Cell Rep.* 13, 1994; Ayeres N. M. and Park, W. D., *Crit. Rev. Plant. Sci.* 13:219-239, 1994; Barcelo et al., *Plant. J.* 5:583-592, 1994; Becker, et al., *Plant. J.* 5:299-307, 1994; Borkowska et al., *Acta. Physiol Plant.* 16:225-230, 1994; Christou, P., *Agro. Food. Ind. Hi Tech.* 5:17-27, 1994; Eapen et al., *Plant Cell Rep.* 13:582-586, 1994; Hartman et al., *Bio-Technology* 12:919-923, 1994; Ritala et al., *Plant. Mol. Biol.* 24:317-325, 1994; and Wan, Y. C. and Lemaux, P. G., *Plant Physiol.* 104:3748, 1994. The constructs may also be transformed into plant cells using homologous recombination.

**[0163]** The term "wild-type" when made in reference to a peptide sequence and nucleotide sequence refers to a peptide sequence and nucleotide sequence

(locus/gene/allele), respectively, which has the characteristics of that peptide sequence and nucleotide sequence when isolated from a naturally occurring source. A wild-type peptide sequence and nucleotide sequence is that which is most frequently observed in a population and is thus arbitrarily designated the “normal” or “wild-type” form of the peptide sequence and nucleotide sequence, respectively. “Wild-type” may also refer to the sequence at a specific nucleotide position or positions, or the sequence at a particular codon position or positions, or the sequence at a particular amino acid position or positions.

[0164] “Consensus sequence” is defined as a sequence of amino acids or nucleotides that contain identical amino acids or nucleotides or functionally equivalent amino acids or nucleotides for at least 25% of the sequence. The identical or functionally equivalent amino acids or nucleotides need not be contiguous.

[0165] The term “*Brassica*” as used herein refers to plants of the *Brassica* genus. Exemplary *Brassica* species include, but are not limited to, *B. carinata*, *B. elongate*, *B. fruticulosa*, *B. juncea*, *B. napus*, *B. narinosa*, *B. nigra*, *B. oleracea*, *B. perviridis*, *B. rapa* (*syn B. campestris*), *B. rupestris*, *B. septiceps*, and *B. tournefortii*.

[0166] A nucleobase is a base, which in certain preferred embodiments is a purine, pyrimidine, or a derivative or analog thereof. Nucleosides are nucleobases that contain a pentosefuranosyl moiety, e.g., an optionally substituted riboside or 2'-deoxyriboside. Nucleosides can be linked by one of several linkage moieties, which may or may not contain phosphorus. Nucleosides that are linked by unsubstituted phosphodiester linkages are termed nucleotides. The term “nucleobase” as used herein includes peptide nucleobases, the subunits of peptide nucleic acids, and morpholine nucleobases as well as nucleosides and nucleotides.

[0167] An oligonucleobase is a polymer comprising nucleobases; in some embodiments at least a portion of which can hybridize by Watson-Crick base pairing to a DNA having the complementary sequence. An oligonucleobase chain may have a single 5' and 3' terminus, which are the ultimate nucleobases of the polymer. A particular oligonucleobase chain can contain nucleobases of all types. An oligonucleobase compound is a compound comprising one or more oligonucleobase chains that may be complementary and hybridized by Watson-Crick base pairing. Ribo-type nucleobases include pentosefuranosyl containing nucleobases wherein the 2' carbon is a methylene

substituted with a hydroxyl, alkyloxy or halogen. Deoxyribo-type nucleobases are nucleobases other than ribo-type nucleobases and include all nucleobases that do not contain a pentosefuranosyl moiety.

**[0168]** In certain embodiments, an oligonucleobase strand may include both oligonucleobase chains and segments or regions of oligonucleobase chains. An oligonucleobase strand may have a 3' end and a 5' end, and when an oligonucleobase strand is coextensive with a chain, the 3' and 5' ends of the strand are also 3' and 5' termini of the chain.

**[0169]** As used herein the term "codon" refers to a sequence of three adjacent nucleotides (either RNA or DNA) constituting the genetic code that determines the insertion of a specific amino acid in a polypeptide chain during protein synthesis or the signal to stop protein synthesis. The term "codon" is also used to refer to the corresponding (and complementary) sequences of three nucleotides in the messenger RNA into which the original DNA is transcribed.

**[0170]** As used herein, the term "homology" refers to sequence similarity among proteins and DNA. The term "homology" or "homologous" refers to a degree of identity. There may be partial homology or complete homology. A partially homologous sequence is one that has less than 100% sequence identity when compared to another sequence.

**[0171]** "Heterozygous" refers to having different alleles at one or more genetic loci in homologous chromosome segments. As used herein "heterozygous" may also refer to a sample, a cell, a cell population or an organism in which different alleles at one or more genetic loci may be detected. Heterozygous samples may also be determined via methods known in the art such as, for example, nucleic acid sequencing. For example, if a sequencing electropherogram shows two peaks at a single locus and both peaks are roughly the same size, the sample may be characterized as heterozygous. Or, if one peak is smaller than another, but is at least about 25% the size of the larger peak, the sample may be characterized as heterozygous. In some embodiments, the smaller peak is at least about 15% of the larger peak. In other embodiments, the smaller peak is at least about 10% of the larger peak. In other embodiments, the smaller peak is at least about 5% of the larger peak. In other embodiments, a minimal amount of the smaller peak is detected.

**[0172]** As used herein, "homozygous" refers to having identical alleles at one or more genetic loci in homologous chromosome segments. "Homozygous" may also refer to a

sample, a cell, a cell population or an organism in which the same alleles at one or more genetic loci may be detected. Homozygous samples may be determined via methods known in the art, such as, for example, nucleic acid sequencing. For example, if a sequencing electropherogram shows a single peak at a particular locus, the sample may be termed “homozygous” with respect to that locus.

[0173] The term “hemizygous” refers to a gene or gene segment being present only once in the genotype of a cell or an organism because the second allele is deleted, or is not present on the homologous chromosome segment. As used herein “hemizygous” may also refer to a sample, a cell, a cell population or an organism in which an allele at one or more genetic loci may be detected only once in the genotype.

[0174] The term “zygosity status” as used herein refers to a sample, a cell population, or an organism as appearing heterozygous, homozygous, or hemizygous as determined by testing methods known in the art and described herein. The term “zygosity status of a nucleic acid” means determining whether the source of nucleic acid appears heterozygous, homozygous, or hemizygous. The “zygosity status” may refer to differences in at a single nucleotide position in a sequence. In some methods, the zygosity status of a sample with respect to a single mutation may be categorized as homozygous wild-type, heterozygous (i.e., one wild-type allele and one mutant allele), homozygous mutant, or hemizygous (i.e., a single copy of either the wild-type or mutant allele).

[0175] As used herein, the term “*RTDS*” refers to The Rapid Trait Development System™ (RTDS) developed by Cibus. *RTDS* is a site-specific gene modification system that is effective at making precise changes in a gene sequence without the incorporation of foreign genes or control sequences.

[0176] The term “about” as used herein means in quantitative terms plus or minus 10%. For example, “about 3%” would encompass 2.7-3.3% and “about 10%” would encompass 9-11%. Moreover, where “about” is used herein in conjunction with a quantitative term it is understood that in addition to the value plus or minus 10%, the exact value of the quantitative term is also contemplated and described. For example, the term “about 3%” expressly contemplates, describes and includes exactly 3%.

[0177] ***RTDS* and Repair Oligonucleotides (GRONs)**

[0178] This disclosure generally relates to novel methods to improve the efficiency of the targeting of modifications to specific locations in genomic or other nucleotide

sequences. Additionally, this disclosure relates to target DNA that has been modified, mutated or marked by the approaches disclosed herein. The disclosure also relates to cells, tissue, and organisms which have been modified by the disclosure's methods. The present disclosure builds on the development of compositions and methods related in part to the successful conversion system, the Rapid Trait Development System (*RTDS*<sup>™</sup>, Cibus US LLC).

[0179] *RTDS* is based on altering a targeted gene by utilizing the cell's own gene repair system to specifically modify the gene sequence *in situ* and not insert foreign DNA and gene expression control sequences. This procedure effects a precise change in the genetic sequence while the rest of the genome is left unaltered. In contrast to conventional transgenic GMOs, there is no integration of foreign genetic material, nor is any foreign genetic material left in the plant. The changes in the genetic sequence introduced by *RTDS* are not randomly inserted. Since affected genes remain in their native location, no random, uncontrolled or adverse pattern of expression occurs.

[0180] The *RTDS* that effects this change is a chemically synthesized oligonucleotide (GRON) as described herein which may be composed of both DNA and modified RNA bases as well as other chemical moieties, and is designed to hybridize at the targeted gene location to create a mismatched base-pair(s). This mismatched base-pair acts as a signal to attract the cell's own natural gene repair system to that site and correct (replace, insert or delete) the designated nucleotide(s) within the gene. Once the correction process is complete the *RTDS* molecule is degraded and the now-modified or repaired gene is expressed under that gene's normal endogenous control mechanisms.

[0181] The methods and compositions disclosed herein can be practiced or made with "gene repair oligonucleobases" (GRON) having the conformations and chemistries as described in detail herein and below. The "gene repair oligonucleobases" as contemplated herein have also been described in published scientific and patent literature using other names including "recombinagenic oligonucleobases;" "RNA/DNA chimeric oligonucleotides;" "chimeric oligonucleotides;" "mixed duplex oligonucleotides" (MDONs); "RNA DNA oligonucleotides (RDOs);" "gene targeting oligonucleotides;" "genoplasts;" "single stranded modified oligonucleotides;" "Single stranded oligodeoxynucleotide mutational vectors" (SSOMVs); "duplex mutational vectors;" and "heteroduplex mutational vectors." The gene repair oligonucleobase can be introduced into a plant cell using any method commonly used in the art, including but not limited to,

microcarriers (biolistic delivery), microfibers, polyethylene glycol (PEG)-mediated uptake, electroporation, and microinjection.

**[0182]** In one embodiment, the gene repair oligonucleobase is a mixed duplex oligonucleotides (MDON) in which the RNA-type nucleotides of the mixed duplex oligonucleotide are made RNase resistant by replacing the 2'-hydroxyl with a fluoro, chloro or bromo functionality or by placing a substituent on the 2'-O. Suitable substituents include the substituents taught by the Kmiec II. Alternative substituents include the substituents taught by U.S. Pat. No. 5,334,711 (Sproat) and the substituents taught by patent publications EP 629 387 and EP 679 657 (collectively, the Martin Applications). As used herein, a 2'-fluoro, chloro or bromo derivative of a ribonucleotide or a ribonucleotide having a T- OH substituted with a substituent described in the Martin Applications or Sproat is termed a "T- Substituted Ribonucleotide." As used herein the term "RNA-type nucleotide" means a T- hydroxyl or 2 '-Substituted Nucleotide that is linked to other nucleotides of a mixed duplex oligonucleotide by an unsubstituted phosphodiester linkage or any of the non-natural linkages taught by Kmiec I or Kmiec II. As used herein the term "deoxyribo-type nucleotide" means a nucleotide having a T-H, which can be linked to other nucleotides of a gene repair oligonucleobase by an unsubstituted phosphodiester linkage or any of the non-natural linkages taught by Kmiec I or Kmiec II.

**[0183]** In a particular embodiment of the present disclosure, the gene repair oligonucleobase is a mixed duplex oligonucleotide (MDON) that is linked solely by unsubstituted phosphodiester bonds. In alternative embodiments, the linkage is by substituted phosphodiesters, phosphodiester derivatives and non-phosphorus-based linkages as taught by Kmiec II. In yet another embodiment, each RNA-type nucleotide in the mixed duplex oligonucleotide is a 2 '-Substituted Nucleotide. Particular preferred embodiments of 2'-Substituted Ribonucleotides are 2'-fluoro, T- methoxy, 2'-propyloxy, 2'-allyloxy, 2'-hydroxyethyloxy, 2'-methoxyethyloxy, T- fluoropropyloxy and 2'-trifluoropropyloxy substituted ribonucleotides. More preferred embodiments of 2'-Substituted Ribonucleotides are 2'-fluoro, 2'-methoxy, 2'-methoxyethyloxy, and 2'-allyloxy substituted nucleotides. In another embodiment the mixed duplex oligonucleotide is linked by unsubstituted phosphodiester bonds,

**[0184]** Although mixed duplex oligonucleotides (MDONs) having only a single type of 2'- substituted RNA-type nucleotide are more conveniently synthesized, the methods of

the disclosure can be practiced with mixed duplex oligonucleotides having two or more types of RNA-type nucleotides. The function of an RNA segment may not be affected by an interruption caused by the introduction of a deoxynucleotide between two RNA-type trinucleotides, accordingly, the term RNA segment encompasses terms such as "interrupted RNA segment." An uninterrupted RNA segment is termed a contiguous RNA segment. In an alternative embodiment an RNA segment can contain alternating RNase-resistant and unsubstituted 2'-OH nucleotides. The mixed duplex oligonucleotides in some embodiments have fewer than 100 nucleotides and other embodiments fewer than 85 nucleotides, but more than 50 nucleotides. The first and second strands are Watson-Crick base paired. In one embodiment the strands of the mixed duplex oligonucleotide are covalently bonded by a linker, such as a single stranded hexa, penta or tetranucleotide so that the first and second strands are segments of a single oligonucleotide chain having a single 3' and a single 5' end. The 3' and 5' ends can be protected by the addition of a "hairpin cap" whereby the 3' and 5' terminal nucleotides are Watson-Crick paired to adjacent nucleotides. A second hairpin cap can, additionally, be placed at the junction between the first and second strands distant from the 3' and 5' ends, so that the Watson-Crick pairing between the first and second strands is stabilized.

**[0185]** The first and second strands contain two regions that are homologous with two fragments of the target gene/allele, i.e., have the same sequence as the target gene/allele. A homologous region contains the nucleotides of an RNA segment and may contain one or more DNA-type nucleotides of connecting DNA segment and may also contain DNA-type nucleotides that are not within the intervening DNA segment. The two regions of homology are separated by, and each is adjacent to, a region having a sequence that differs from the sequence of the target gene, termed a "heterologous region." The heterologous region can contain one, two or three mismatched nucleotides. The mismatched nucleotides can be contiguous or alternatively can be separated by one or two nucleotides that are homologous with the target gene/allele. Alternatively, the heterologous region can also contain an insertion or one, two, three or of five or fewer nucleotides. Alternatively, the sequence of the mixed duplex oligonucleotide may differ from the sequence of the target gene/allele only by the deletion of one, two, three, or five or fewer nucleotides from the mixed duplex oligonucleotide. The length and position of the heterologous region is, in this case, deemed to be the length of the deletion, even though no nucleotides of the mixed duplex oligonucleotide are within the heterologous

region. The distance between the fragments of the target gene that are complementary to the two homologous regions is identical to the length of the heterologous region where a substitution or substitutions is intended. When the heterologous region contains an insertion, the homologous regions are thereby separated in the mixed duplex oligonucleotide farther than their complementary homologous fragments are in the gene/allele, and the converse is applicable when the heterologous region encodes a deletion.

**[0186]** The RNA segments of the mixed duplex oligonucleotides are each a part of a homologous region, i.e., a region that is identical in sequence to a fragment of the target gene, which segments together in some embodiments contain at least 13 RNA-type nucleotides and in some embodiments from 16 to 25 RNA-type nucleotides or yet other embodiments 18-22 RNA-type nucleotides or in some embodiments 20 nucleotides. In one embodiment, RNA segments of the homology regions are separated by and adjacent to, i.e., “connected by” an intervening DNA segment. In one embodiment, each nucleotide of the heterologous region is a nucleotide of the intervening DNA segment. An intervening DNA segment that contains the heterologous region of a mixed duplex oligonucleotide is termed a “mutator segment.”

**[0187]** In another embodiment of the present disclosure, the gene repair oligonucleobase (GRON) is a single stranded oligodeoxynucleotide mutational vector (SSOMV), such as disclosed in International Patent Application PCT/USOO/23457, U.S. Pat. Nos. 6,271,360, 6,479,292, and 7,060,500. The sequence of the SSOMV is based on the same principles as the mutational vectors described in U.S. Pat. Nos. 5,756,325; 5,871,984; 5,760,012; 5,888,983; 5,795,972; 5,780,296; 5,945,339; 6,004,804; and 6,010,907 and in International Publication Nos. WO 98/49350; WO 99/07865; WO 99/58723; WO 99/58702; and WO 99/40789. The sequence of the SSOMV contains two regions that are homologous with the target sequence separated by a region that contains the desired genetic alteration termed the mutator region. The mutator region can have a sequence that is the same length as the sequence that separates the homologous regions in the target sequence, but having a different sequence. Such a mutator region can cause a substitution. Alternatively, the homologous regions in the SSOMV can be contiguous to each other, while the regions in the target gene having the same sequence are separated by one, two or more nucleotides. Such an SSOMV causes a deletion from the target gene of the nucleotides that are absent from the SSOMV. Lastly, the sequence of the target gene

nucleotides that are absent from the SSOMV. Lastly, the sequence of the target gene that is identical to the homologous regions may be adjacent in the target gene but separated by one, two, or more nucleotides in the sequence of the SSOMV. Such an SSOMV causes an insertion in the sequence of the target gene. In certain embodiments, a SSOMV does not anneal to itself.

**[0188]** The nucleotides of the SSOMV are deoxyribonucleotides that are linked by unmodified phosphodiester bonds except that the 3' terminal and/or 5' terminal internucleotide linkage or alternatively the two 3' terminal and/or 5' terminal internucleotide linkages can be a phosphorothioate or phosphoamidate. As used herein an internucleotide linkage is the linkage between nucleotides of the SSOMV and does not include the linkage between the 3' end nucleotide or 5' end nucleotide and a blocking substituent. In a specific embodiment the length of the SSOMV is between 21 and 55 deoxynucleotides and the lengths of the homology regions are, accordingly, a total length of at least 20 deoxynucleotides and at least two homology regions should each have lengths of at least 8 deoxynucleotides.

**[0189]** The SSOMV can be designed to be complementary to either the coding or the non-coding strand of the target gene. When the desired mutation is a substitution of a single base, it is preferred that both the mutator nucleotide and the targeted nucleotide be a pyrimidine. To the extent that is consistent with achieving the desired functional result, it is preferred that both the mutator nucleotide and the targeted nucleotide in the complementary strand be pyrimidines. Particularly preferred are SSOMVs that encode transversion mutations, i.e., a C or T mutator nucleotide is mismatched, respectively, with a C or T nucleotide in the complementary strand.

**[0190]** Okazaki Fragment / 2'-OME GRON Design. In various embodiments, a GRON may have both RNA and DNA nucleotides and/or other types of nucleobases. In some embodiments, one or more of the DNA or RNA nucleotides comprise a modification. In certain embodiments, the first 5' nucleotide is an RNA nucleotide and the remainder of the nucleotides are DNA. In still further embodiments, the first 5' RNA nucleotide is modified with a 2-O-Me. In other embodiments, the first two, three, four, five, six, seven, eight, nine, ten or more 5' nucleotides are an RNA nucleotide and the remainder of the nucleotides are DNA. In still further embodiments, one or more of the first two, three, four, five, six, seven, eight, nine, ten or more 5' RNA nucleotide are modified with a 2-O-Me. In plant cells, double-strand breaks in DNA are typically repaired by the NHEJ DNA

repair pathway. This pathway does not require a template to repair the DNA and is therefore error prone. The advantage of using this pathway to repair DNA for a plant cell is that it is quick, ubiquitous and most importantly can occur at times when a cell is not undergoing DNA replication. Another DNA repair pathway that functions in repairing double-strand breaks outside of the replication fork in plant cells is called homologous recombination (HR); however, unlike the NHEJ pathway this type of repair is precise and requires the use of a DNA template (GRON). Since these GRONs mimic Okazaki fragments at the DNA replication fork of targeted genes, it is not obvious to use them with a double-strand DNA cutter to those skilled in the art.

**[0191] Improving efficiency**

**[0192]** The present disclosure provides a number of approaches to increase the effectiveness of conversion of a target gene using repair oligonucleotides, and which may be used alone or in combination with one another. These include:

1. Introducing modifications to the repair oligonucleotides which attract DNA repair machinery to the targeted (mismatch) site.
  - A. Introduction of one or more abasic sites in the oligonucleotide (e.g., within 10 bases, and in some embodiments with 5 bases of the desired mismatch site) generates a lesion which is an intermediate in base excision repair (BER), and which attracts BER machinery to the vicinity of the site targeted for conversion by the repair oligonucleotide. dSpacer (abasic furan) modified oligonucleotides may be prepared as described in, for example, Takeshita et al., *J. Biol. Chem.*, 262:10171-79, 1987.
  - B. Inclusion of compounds which induce single or double strand breaks, either into the oligonucleotide or together with the oligonucleotide, generates a lesion which is repaired by NHEJ, microhomology-mediated end joining (MMEJ), and homologous recombination. By way of example, the bleomycin family of antibiotics, zinc fingers, FokI (or any type IIS class of restriction enzyme) and other nucleases may be covalently coupled to the 3' or 5' end of repair oligonucleotides, in order to introduce double strand breaks in the vicinity of the site targeted for conversion by the repair oligonucleotide. The bleomycin family of antibiotics are DNA cleaving glycopeptides which

include bleomycin, zeocin, phleomycin, tallysomycin, pepleomycin and others.

C. Introduction of one or more 8'oxo dA or dG incorporated in the oligonucleotide (e.g., within 10 bases, and in some embodiments with 5 bases of the desired mismatch site) generates a lesion which is similar to lesions created by reactive oxygen species. These lesions induce the so-called "pushing repair" system. See, e.g., Kim et al., J. Biochem. Mol. Biol. 37:657-62, 2004.

2. Increase stability of the repair oligonucleotides:

Introduction of a reverse base (idC) at the 3' end of the oligonucleotide to create a 3' blocked end on the repair oligonucleotide.

Introduction of one or more 2'O-methyl nucleotides or bases which increase hybridization energy (see, e.g., WO2007/073149) at the 5' and/or 3' of the repair oligonucleotide.

Introduction of one or a plurality of 2'O-methyl RNA nucleotides at the 5' end of the repair oligonucleotide, leading into DNA bases which provide the desired mismatch site, thereby creating an Okazaki Fragment-like nucleic acid structure.

Conjugated (5' or 3') intercalating dyes such as acridine, psoralen, ethidium bromide and Syber stains.

Introduction of a 5' terminus cap such as a T/A clamp, a cholesterol moiety, SIMA (HEX), riboC and amidite.

Backbone modifications such as phosphothioate, 2'-O methyl, methyl phosphonates, locked nucleic acid (LNA), MOE (methoxyethyl), di PS and peptide nucleic acid (PNA).

Crosslinking of the repair oligonucleotide, e.g., with intrastrand crosslinking reagents agents such as cisplatin and mitomycin C.

Conjugation with fluorescent dyes such as Cy3, DY547, Cy3.5, Cy3B, Cy5 and DY647.

3. Increase hybridization energy of the repair oligonucleotide through incorporation of bases which increase hybridization energy (see, e.g., WO2007/073149).
4. Increase the quality of repair oligonucleotide synthesis by using nucleotide multimers (dimers, trimers, tetramers, etc.) as building blocks for synthesis. This results in fewer coupling steps and easier separation of the full length products from building blocks.
5. Use of long repair oligonucleotides (i.e., greater than 55 nucleotides in length, for example such as the lengths described herein, for example having one or more mutations or two or more mutations targeted in the repair oligonucleotide).

[0193] Examples of the foregoing approaches are provided in Table 1.

[0194] **Table 1. Exemplary GRON chemistries.**

	Oligo type	Modifications
5' mods	T/A clamp	T/A clamp
Backbone modifications	Phosphothioate	PS
Intercalating dyes	5' Acridine 3'	idC Acridine, idC
2'-O-methyl		DNA/RNA
Cy3 replacements		DY547
Facilitators	2'-O-Me oligos designed 5' and 3' of the converting oligo	2'-O-Me
Abasic	Abasic site placed in various locations 5' and 3' to the converting base. 44 mer	Abasic 2
Assist	Assist approach Overlap: 2 oligos: 1 with Cy3/idC, 1 unmodified repair oligo	Cy3, idC on one, none on the other:
Assist	Assist approach No overlap: 2 oligos: 1 with Cy3/idC, 1 unmodified repair oligo	only make the unmodified oligo
Abasic	THF site placed in various locations 5' and 3' to the converting base. 44 mer	Tetrahydrofuran ( dspace)
Backbone modifications	9	2'-O-Me
Trimers		Trimer amidites, Cy3, idC
Pushing repair		8'oxo dA, 5' Cy3, idC
Pushing repair		8'oxo dA, 5' Cy3, idC

	Oligo type	Modifications
Double Strand Break		Bleomycin
Crosslinker		Cisplatin
Crosslinker		Mitomycin C
Facilitators	super bases 5' and 3' of converting oligo	2 amino dA and 2- thio T
Super oligos		2' amino d, 5' Cy3, idC
Super oligos		2-thio T, 5' Cy3, idC
Super oligos		7-deaza A, 5' Cy3, idC
Super oligos		7-deaza G, 5' Cy3, idC
Super oligos		propanyl dC, 5' Cy3, idC
Intercalating dyes	5' Psoralen/3' idC	Psoralen, idC
Intercalating dyes	5' Ethidium bromide	Ethidium bromide
Intercalating dyes	5' Syber stains	Syber stains
5' mods	5' Chol/3' idC	Cholesterol
Double mutation	Long oligo (55+ bases) w/ 2 mutation	Any modification
5' mods	5' SIMA HEX/3'idC	SIMA HEX, idC
Backbone modifications	9	Methyl phosphonates
Backbone modifications		LNA
Backbone modifications	1	MOE (methoxyethyl)
Cy3 replacements		Cy3.5
Cy3 replacements		Cy5
Backbone modifications		di PS
5' mods		riboC for branch nm
Backbone modifications		PNA
Cy3 replacements		DY647
5' mods	5' branch	symmetric branch amidite/idC

[0195] The foregoing modifications may also include known nucleotide modifications such as methylation, 5' intercalating dyes, modifications to the 5' and 3' ends, backbone modifications, crosslinkers, cyclization and 'caps' and substitution of one or more of the naturally occurring nucleotides with an analog such as inosine. Modifications of nucleotides include the addition of acridine, amine, biotin, cascade blue, cholesterol, Cy3@, Cy5@, Cy5.5@ Daboyl, digoxigenin, dinitrophenyl, Edans, 6-FAM, fluorescein, 3'- glyceryl, HEX, IRD-700, IRD-800, JOE, phosphate psoralen, rhodamine, ROX, thiol (SH), spacers, TAMRA, TET, AMCA-S", SE, BODIPY°, Marina Blue@, Pacific Blue@, Oregon Green@, Rhodamine Green@, Rhodamine Red@, Rhodol Green@ and Texas Red@. Polynucleotide backbone modifications include methylphosphonate, 2'-OMe-methylphosphonate RNA, phosphorothiorate, RNA, 2'-OMeRNA. Base modifications include 2-amino-dA, 2-aminopurine, 3'- (ddA), 3'dA (cordycepin), 7-deaza-dA, 8-Br-dA,

8- oxo-dA, N6-Me-dA, abasic site (dSpacer), biotin dT, 2'-OMe-5Me-C, 2'-OMe-propynyl-C, 3'- (5-Me-dC), 3'- (ddC), 5-Br-dC, 5-I-dC, 5-Me-dC, 5-F-dC, carboxy-dT, convertible dA, convertible dC, convertible dG, convertible dT, convertible dU, 7-deaza-dG, 8-Br-dG, 8- oxo-dG, O6-Me-dG, S6-DNP-dG, 4-methyl-indole, 5-nitroindole, 2'-OMe-inosine, 2'-dl, o6- phenyl-dl, 4-methyl-indole, 2'-deoxynebularine, 5-nitroindole, 2-aminopurine, dP (purine analogue), dK (pyrimidine analogue), 3-nitropyrrole, 2-thio-dT, 4-thio-dT, biotin-dT, carboxy-dT, 04-Me-dT, 04-triazol dT, 2'-OMe-propynyl-U, 5-Br-dU, 2'-dU, 5-F-dU, 5-I-dU, 04-triazol dU. Said terms also encompass peptide nucleic acids (PNAs), a DNA analogue in which the backbone is a pseudopeptide consisting of N- (2-aminoethyl)-glycine units rather than a sugar. PNAs mimic the behavior of DNA and bind complementary nucleic acid strands. The neutral backbone of PNA results in stronger binding and greater specificity than normally achieved. In addition, the unique chemical, physical and biological properties of PNA have been exploited to produce powerful biomolecular tools, antisense and antigene agents, molecular probes and biosensors.

**[0196]** Oligonucleobases may have nick(s), gap(s), modified nucleotides such as modified oligonucleotide backbones, abasic nucleotides, or other chemical moieties. In a further embodiment, at least one strand of the oligonucleobase includes at least one additional modified nucleotide, e.g., a 2'-O-methyl modified nucleotide such as a MOE (methoxyethyl), a nucleotide having a 5'-phosphorothioate group, a terminal nucleotide linked to a cholesteryl derivative, a 2'-deoxy-2'-fluoro modified nucleotide, a 2'-deoxy-modified nucleotide, a locked nucleotide, an abasic nucleotide (the nucleobase is missing or has a hydroxyl group in place thereof (see, e.g., Glen Research, <http://www.glenresearch.com/GlenReports/GR21-14.html>)), a 2'-amino-modified nucleotide, a 2'-alkyl-modified nucleotide, a morpholino nucleotide, a phosphoramidite, and a non-natural base comprising nucleotide. Various salts, mixed salts and free acid forms are also included.

**[0197]** Preferred modified oligonucleotide backbones include, for example, phosphorothioates, chiral phosphorothioates, phosphoro-dithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates, 5'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkyl-phosphonates,

thionoalkylphosphotriesters, selenophosphates and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein one or more internucleotide linkages is a 3' to 3', 5' to 5' or 2' to 2' linkage. Preferred oligonucleotides having inverted polarity comprise a single 3' to 3' linkage at the 3'-most internucleotide linkage i.e. a single inverted nucleoside residue which may be abasic (the nucleobase is missing or has a hydroxyl group in place thereof). The most common use of a linkage inversion is to add a 3'-3' linkage to the end of an antisense oligonucleotide with a phosphorothioate backbone. The 3'-3' linkage further stabilizes the antisense oligonucleotide to exonuclease degradation by creating an oligonucleotide with two 5'-OH ends and no 3'-OH end. Linkage inversions can be introduced into specific locations during oligonucleotide synthesis through use of "reversed phosphoramidites". These reagents have the phosphoramidite groups on the 5'-OH position and the dimethoxytrityl (DMT) protecting group on the 3'-OH position. Normally, the DMT protecting group is on the 5'-OH and the phosphoramidite is on the 3'-OH.

[0198] Examples of modified bases include, but are not limited to, 2-aminopurine, 2'-amino-butryl pyrene-uridine, 2'-aminouridine, 2'-deoxyuridine, 2'-fluoro-cytidine, 2'-fluoro-uridine, 2,6-diaminopurine, 4-thio-uridine, 5-bromo-uridine, 5-fluoro-cytidine, 5-fluorouridine, 5-indo-uridine, 5-methyl-cytidine, inosine, N3-methyl-uridine, 7-deaza-guanine, 8-aminohexyl-amino-adenine, 6-thio-guanine, 4-thio-thymine, 2-thio-thymine, 5-iodo-uridine, 5-iodo-cytidine, 8-bromo-guanine, 8-bromo-adenine, 7-deaza-adenine, 7-diaza-guanine, 8-oxo-guanine, 5,6-dihydro-uridine, and 5-hydroxymethyl-uridine. These synthetic units are commercially available; (for example, purchased from Glen Research Company) and can be incorporated into DNA by chemical synthesis.

[0199] Examples of modification of the sugar moiety are 3'-deoxylation, 2'-fluorination, and arabanosidation, however, it is not to be construed as being limited thereto. Incorporation of these into DNA is also possible by chemical synthesis.

[0200] Examples of the 5' end modification are 5'-amination, 5'-biotinylation, 5'-fluoresceinylation, 5'-tetrafluoro-fluoreceinylation, 5'-thionation, and 5'-dabsylation, however it is not to be construed as being limited thereto.

[0201] Examples of the 3' end modification are 3'-amination, 3'-biotinylation, 2,3-dideoxylation, 3'-thionation, 3'-dabsylation, 3'-carboxylation, and 3'-cholesterylation, however, it is not to be construed as being limited thereto.

[0202] In one preferred embodiment, the oligonucleobase can contain a 5' blocking substituent that is attached to the 5' terminal carbons through a linker. The chemistry of the linker is not critical other than its length, which should in some embodiments be at least 6 atoms long and that the linker should be flexible. A variety of non-toxic substituents such as biotin, cholesterol or other steroids or a non-intercalating cationic fluorescent dye can be used. Particularly preferred reagents to make oligonucleobases are the reagents sold as Cy3™ and Cy5™ by Glen Research, Sterling Va. (now GE Healthcare), which are blocked phosphoramidites that upon incorporation into an oligonucleotide yield 3,3,3',3'-tetramethyl N,N'-isopropyl substituted indomonocarbocyanine and indodicarbocyanine dyes, respectively. Cy3 is particularly preferred. When the indocarbocyanine is N-oxyalkyl substituted it can be conveniently linked to the 5' terminal of the oligodeoxynucleotide as a phosphodiester with a 5' terminal phosphate. When the commercially available Cy3 phosphoramidite is used as directed, the resulting 5' modification consists of a blocking substituent and linker together which are a N-hydroxypropyl, N'-phosphatidylpropyl 3,3,3',3'-tetramethyl indomonocarbocyanine. Other dyes contemplated include Rhodamine6G, Tetramethylrhodamine, Sulforhodamine 101, Merocyanine 540, Atto565, Atto550 26, Cy3.5, Dy547, Dy548, Dy549, Dy554, Dy555, Dy556, Dy560, mStrawberry and mCherry.

[0203] In a preferred embodiment the indocarbocyanine dye is tetra substituted at the 3 and 3' positions of the indole rings. Without limitations as to theory these substitutions prevent the dye from being an intercalating dye. The identity of the substituents at these positions is not critical.

[0204] The oligo designs herein described might also be used as more efficient donor templates in combination with other DNA editing or recombination technologies including, but not limited to, gene targeting using site-specific homologous recombination by zinc finger nucleases, Transcription Activator-Like Effector Nucleases (TALENs) or Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs).

[0205] The present disclosure in certain aspects and embodiments generally relates to methods for the efficient modification of genomic cellular DNA and/or recombination of DNA into the genomic DNA of cells. Although not limited to any particular use, some methods provided herein may in certain embodiments be useful in, for example, introducing a modification into the genome of a cell for the purpose of determining the

effect of the modification on the cell. For example, a modification may be introduced into the nucleotide sequence which encodes an enzyme to determine whether the modification alters the enzymatic activity of the enzyme, and/or determine the location of the enzyme's catalytic region. Alternatively, the modification may be introduced into the coding sequence of a DNA-binding protein to determine whether the DNA binding activity of the protein is altered, and thus to delineate the particular DNA-binding region within the protein. Yet another alternative is to introduce a modification into a non-coding regulatory sequence (e.g., promoter, enhancer, regulatory RNA sequence (miRNA), etc.) in order to determine the effect of the modification on the level of expression of a second sequence which is operably linked to the non-coding regulatory sequence. This may be desirable to, for example, define the particular sequence which possesses regulatory activity.

#### **[0206] DNA Cutters**

**[0207]** One strategy for producing targeted gene disruption is through the generation of single strand or double strand DNA breaks using a DNA cutter such as a site-specific endonuclease. Endonucleases are most often used for targeted gene disruption in organisms that have traditionally been refractive to more conventional gene targeting methods, such as algae, plants, and large animal models, including humans. For example, there are currently human clinical trials underway involving zinc finger nucleases for the treatment and prevention of HIV infection. Additionally, endonuclease engineering is currently being used in attempts to disrupt genes that produce undesirable phenotypes in crops.

#### **[0208] Zinc Fingers**

**[0209]** One class of artificial endonucleases is the zinc finger endonucleases. Zinc finger endonucleases combine a non-specific cleavage domain, typically that of FokI endonuclease, with zinc finger protein domains that are engineered to bind to specific DNA sequences. The modular structure of the zinc finger endonucleases makes them a versatile platform for delivering site-specific double-strand breaks to the genome. As FokI endonuclease cleaves as a dimer, one strategy to prevent off-target cleavage events has been to design zinc finger domains that bind at adjacent 9 base pair sites. See also U.S. Pat. Nos. 7,285,416; 7,521,241; 7,361,635; 7,273,923; 7,262,054; 7,220,719;

7,070,934; 7,013,219; 6,979,539; 6,933,113; 6,824,978.

**[0210] TALENs**

**[0211]** TALENs are targetable nucleases are used to induce single- and double-strand breaks into specific DNA sites, which are then repaired by mechanisms that can be exploited to create sequence alterations at the cleavage site.

**[0212]** The fundamental building block that is used to engineer the DNA-binding region of TALENs is a highly conserved repeat domain derived from naturally occurring TALEs encoded by *Xanthomonas spp.* proteobacteria. DNA binding by a TALEN is mediated by arrays of highly conserved 33–35 amino acid repeats that are flanked by additional TALE-derived domains at the amino-terminal and carboxy-terminal ends of the repeats.

**[0213]** These TALE repeats specifically bind to a single base of DNA, the identity of which is determined by two hypervariable residues typically found at positions 12 and 13 of the repeat, with the number of repeats in an array corresponded to the length of the desired target nucleic acid, the identity of the repeat selected to match the target nucleic acid sequence. In some embodiments, the target nucleic acid is between 15 and 20 base pairs in order to maximize selectivity of the target site. Cleavage of the target nucleic acid typically occurs within 50 base pairs of TALEN binding. Computer programs for TALEN recognition site design have been described in the art. See, e.g., Cermak et al., *Nucleic Acids Res.* 2011 July; 39(12): e82.

**[0214]** Once designed to match the desired target sequence, TALENs can be expressed recombinantly and introduced into protoplasts as exogenous proteins, or expressed from a plasmid within the protoplast or administered as mRNA.

**[0215] Meganucleases**

**[0216]** The homing endonucleases, also known as meganucleases, are sequence specific endonucleases that generate double strand breaks in genomic DNA with a high degree of specificity due to their large (e.g., >14 bp) cleavage sites. While the specificity of the homing endonucleases for their target sites allows for precise targeting of the induced DNA breaks, homing endonuclease cleavage sites are rare and the probability of finding a naturally occurring cleavage site in a targeted gene is low.

[0217] Another class of artificial endonucleases is the engineered meganucleases. Engineered homing endonucleases are generated by modifying the specificity of existing homing endonucleases. In one approach, variations are introduced in the amino acid sequence of naturally occurring homing endonucleases and then the resultant engineered homing endonucleases are screened to select functional proteins which cleave a targeted binding site. In another approach, chimeric homing endonucleases are engineered by combining the recognition sites of two different homing endonucleases to create a new recognition site composed of a half- site of each homing endonuclease. See e.g., US Patent Nos. 8,338,157.

**[0218] CRISPRs or CRISPR/cas Systems**

[0219] CRISPR-Cas system contains three basic design components: 1) Cas gene, transcript (e.g., mRNA) or protein; 2) guide RNA (gRNA); and 3) crRNAs (CRISPR RNA) are RNA segments processed from RNA transcripts encoding the CRISPR repeat arrays, which harbor a “protospacer” region that are complementary to a foreign DNA site (e.g., endogenous DNA target region) and a part of the CRISPR repeat. See e.g., PCT Application No.s WO/2014/093661 and WO/2013/176772.

**[0220] Cas (CRISPR Associated) Gene, Transcript (e.g., mRNA) or Protein**

[0221] Transient Cas expression from a plasmid vector, direct delivery of Cas protein and or direct delivery of Cas mRNA into plant cells. Cas genes are codon optimized for expression in higher plants, algae or yeast and are driven by either a constitutive, inducible, tissue-specific or species-specific promoter when applicable. Cas transcript termination and polyadenylation signals are either NosT, RBCT, HSP18.2T or other gene specific or species-specific terminators. Cas gene cassettes or mRNA may contain introns, either native or in combination with gene-specific promoters and or synthetic promoters. Cas protein may contain one or more nuclear localization signal sequences (NLS), mutations, deletions, alterations or truncations. In transient expression systems, Cas gene cassettes may be combined with other components of the CRISPR-Cas system such as gRNA cassettes on the same transient expression vector. Alternatively, Cas gene cassettes may be located and expressed from constructs independent of gRNA cassettes or from other components of the CRISPR-Cas system. CRISPR associated (Cas) gene - encode for proteins with a variety of predicted nucleic acid-manipulating activities such as nucleases, helicases and polymerase. Cas genes include cas9. Cas9 is a gene encoding

a large protein containing a predicted RuvC-like and HNH endonuclease domains and is associated with the CRISPR adaptive immunity system that is present in most archaea and many bacteria. Cas9 protein consists of two lobes:

- 1) Recognition (REC) lobe- consists of three domains:
  - a) BH (bridge helix)
  - b) REC1- facilitates RNA-guided DNA targeting
  - c) REC2- facilitates RNA-guided DNA targeting
- 2) Nuclease (NUC) lobe- consists of three domains:
  - a) RuvC- facilitates RNA-guided DNA targeting; endonuclease activity
  - b) HNH - endonuclease activity
  - c) PI- PAM interacting

[0222] In other embodiments, the cas gene may be a homolog of cas9 in which the RuvC, HNH, REC and BH domains are highly conserved. In some embodiments, cas genes are those from the following species.

**[0223] Guide RNA (gRNA)**

[0224] gRNA or sgRNA (single guide RNA) is engineered as a fusion between a crRNA and part of the transactivating CRISPR RNA (tracrRNA) sequence, which guides the Cas9 to a specific target DNA sequence that is complementary to the protospacer region. Guide RNA may include an expression cassette containing a chimeric RNA design with a long tracrRNA hybrid, short tracrRNA hybrid or a native CRISPR array + tracrRNA conformation. Chimeric gRNA combines the targeting specificity of the crRNA with the scaffolding properties of the tracrRNA into a single transcript. gRNA transient expression is controlled by species-specific higher plant RNA Polymerase III promoters such as those from the U6 or U3 snRNA gene family (Wang et al 2008). gRNA transcript termination is controlled by a 6-20 nucleotide tract of poly dT as per Wang et al 2008. gRNA expression cassettes are located on the same or different transient expression vectors from other components of the CRISPR-Cas system. gRNA transcripts may be synthesized *in-vitro* and delivered directly into plant cells, independent of or in combination with gRNA transient expression vectors.

**[0225] Target region**

**[0226]** Guide RNAs contain two components that define specificity to a DNA target region, a proto-spacer and a proto-spacer adjacent motif (PAM). Proto-spacer sequence, typically 20 nucleotides but can vary based on the DNA target, provides DNA sequence specificity for the CRISPR-Cas complex. DNA targets also contain a NNG or NAG tri-nucleotide sequence (PAM) where N denotes any nucleotide, immediately 3' or downstream of the proto-spacer.

**[0227] One component approach**

**[0228]** Similar to Le Cong *et al.* (2013) and others, a simplified “one component approach” to CRISPR-Cas gene editing wherein a single transient expression construct contains all components of the CRISPR-Cas complex, i.e. both the gRNA and the Cas expressions cassettes. This allows for an easy modular design for targeting single or multiple loci in any given plant or crop. Targeting multiple loci can be achieved by simply swapping in the target- specific gRNA cassettes. Additionally, species specific promoters, terminators or other expressing enhancing elements can easily be shuttled in and out of “one component approach” transient vectors allowing for optimal expression of both gRNA and Cas protein in a species specific manner.

**[0229] Two Component Approach**

**[0230]** In the two component approach, Cas and gRNA expression cassettes are located on different transient expression vectors. This allows for delivery of a CRISPR-Cas editing components separately, allowing for different ratios of gRNA to Cas within the same cell. Similar to the one component approach, the two component approach also allows for promoters, terminators or other elements affecting expression of CRISPR-Cas components to be easily altered and allow targeting of DNA in a species-specific manner.

**[0231] Antibiotics**

**[0232]** Another class of endonucleases are antibiotics which are DNA cleaving glycopeptides such as the bleomycin family of antibiotics are DNA cleaving glycopeptides which include bleomycin, zeocin, phleomycin, tallysomycin, pepleomycin and others which are further described herein.

**[0233]** Other DNA-modifying molecules may be used in targeted gene recombination. For example, peptide nucleic acids may be used to induce modifications to the genome of

the target cell or cells (see, e.g., Ecker, U.S. Pat. No. 5,986,053). In brief, synthetic nucleotides comprising, at least, a partial peptide backbone are used to target a homologous genomic nucleotide sequence. Upon binding to the double-helical DNA, or through a mutagen ligated to the peptide nucleic acid, modification of the target DNA sequence and/or recombination is induced to take place. Targeting specificity is determined by the degree of sequence homology between the targeting sequence and the genomic sequence.

**[0234]** Furthermore, the present disclosure is not limited to the particular methods which are used herein to execute modification of genomic sequences. Indeed, a number of methods are contemplated. For example, genes may be targeted using triple helix forming oligonucleotides (TFO). TFOs may be generated synthetically, for example, by PCR or by use of a gene synthesizer apparatus. Additionally, TFOs may be isolated from genomic DNA if suitable natural sequences are found. TFOs may be used in a number of ways, including, for example, by tethering to a mutagen such as, but not limited to, psoralen or chlorambucil (see, e.g., Havre et al., Proc Nat'l Acad Sci, U.S.A. 90:7879-7883, 1993; Havre et al., J Virol 67:7323-7331, 1993; Wang et al., Mol Cell Biol 15:1759-1768, 1995; Takasugi et al., Proc Nat'l Acad Sci, U.S.A. 88:5602-5606, 1991; Belousov et al., Nucleic Acids Res 25:3440-3444, 1997). Furthermore, for example, TFOs may be tethered to donor duplex DNA (see, e.g., Chan et al., J Biol Chem 272:11541-11548, 1999). TFOs can also act by binding with sufficient affinity to provoke error-prone repair (Wang et al., Science 271:802-805, 1996).

**[0235]** The methods disclosed herein are not necessarily limited to the nature or type of DNA-modifying reagent which is used. For example, such DNA-modifying reagents release radicals which result in DNA strand breakage. Alternatively, the reagents alkylate DNA to form adducts which would block replication and transcription. In another alternative, the reagents generate crosslinks or molecules that inhibit cellular enzymes leading to strand breaks. Examples of DNA-modifying reagents which have been linked to oligonucleotides to form TFOs include, but are not limited to, indolocarbazoles, naphthalene diimide (NDI), transplatin, bleomycin, analogues of cyclopropapyrroloindole, and phenanthrohydrodioxins. In particular, indolocarbazoles are topoisomerase I inhibitors. Inhibition of these enzymes results in strand breaks and DNA protein adduct formation (Arimondo et al., Bioorganic and Medicinal Chem. 8, 777, 2000). NDI is a photooxidant that can oxidize guanines which could cause mutations at sites of guanine

residues (Nunez, et al., *Biochemistry*, 39, 6190, 2000). Transplatin has been shown to react with DNA in a triplex target when the TFO is linked to the reagent. This reaction causes the formation of DNA adducts which would be mutagenic (Columbier, et al., *Nucleic Acids Research*, 24: 4519, 1996). Bleomycin is a DNA breaker, widely used as a radiation mimetic. It has been linked to oligonucleotides and shown to be active as a breaker in that format (Sergeyev, *Nucleic Acids Research* 23, 4400, 1995; Kane, et al., *Biochemistry*, 34, 16715, 1995). Analogues of cyclopropapyrroloindole have been linked to TFOs and shown to alkylate DNA in a triplex target sequence. The alkylated DNA would then contain chemical adducts which would be mutagenic (Lukhtanov, et al., *Nucleic Acids Research*, 25, 5077, 1997). Phenanthrodi-hydrodioxins are masked quinones that release radical species upon photoactivation. They have been linked to TFOs and have been shown to introduce breaks into duplex DNA on photoactivation (Bendinskas et al., *Bioconjugate Chem.* 9, 555, 1998).

**[0236]** Other methods of inducing modifications and/or recombination are contemplated by the present disclosure. For example, another embodiment involves the induction of homologous recombination between an exogenous DNA fragment and the targeted gene (see e.g., Capecchi et al., *Science* 244:1288-1292, 1989) or by using peptide nucleic acids (PNA) with affinity for the targeted site. Still other methods include sequence specific DNA recognition and targeting by polyamides (see e.g., Dervan et al., *Curr Opin Chem Biol* 3:688-693, 1999; *Biochemistry* 38:2143-2151, 1999) and the use nucleases with site specific activity (e.g., zinc finger proteins, TALENs, Meganucleases and/or CRISPRs).

**[0237]** The present disclosure is not limited to any particular frequency of modification and/or recombination. In some embodiments the methods disclosed herein result in a frequency of modification in the target nucleotide sequence of from 0.2% to 3%. Nonetheless, any frequency (i.e., between 0% and 100%) of modification and/or recombination is contemplated to be within the scope of the present disclosure. The frequency of modification and/or recombination is dependent on the method used to induce the modification and/or recombination, the cell type used, the specific gene targeted and the DNA mutating reagent used, if any. Additionally, the method used to detect the modification and/or recombination, due to limitations in the detection method, may not detect all occurrences of modification and/or recombination. Furthermore, some modification and/or recombination events may be silent, giving no detectable indication

that the modification and/or recombination has taken place. The inability to detect silent modification and/or recombination events gives an artificially low estimate of modification and/or recombination. Because of these reasons, and others, the disclosure is not necessarily limited to any particular modification and/or recombination frequency. In one embodiment, the frequency of modification and/or recombination is between 0.01% and 100%. In another embodiment, the frequency of modification and/or recombination is between 0.01% and 50%. In yet another embodiment, the frequency of modification and/or recombination is between 0.1% and 10%. In still yet another embodiment, the frequency of modification and/or recombination is between 0.1% and 5%.

**[0238]** The term “frequency of mutation” as used herein in reference to a population of cells which are treated with a DNA-modifying molecule that is capable of introducing a mutation into a target site in the cells' genome, refers to the number of cells in the treated population which contain the mutation at the target site as compared to the total number of cells which are treated with the DNA-modifying molecule. For example, with respect to a population of cells which is treated with the DNA-modifying molecule TFO tethered to psoralen which is designed to introduce a mutation at a target site in the cells' genome, a frequency of mutation of 5% means that of a total of 100 cells which are treated with TFO-psoralen, 5 cells contain a mutation at the target site.

**[0239]** Although the present disclosure is not necessarily limited to any degree of precision in the modification and/or recombination of DNA in the cell, it is contemplated that some embodiments of the present disclosure require higher degrees of precision, depending on the desired result. For example, the specific sequence changes required for gene repair (e.g., particular base changes) require a higher degree of precision as compared to producing a gene knockout wherein only the disruption of the gene is necessary. With the methods of the present disclosure, achievement of higher levels of precision in modification and/or homologous recombination techniques is greater than with prior art methods.

**[0240] Delivery of Gene Repair Oligonucleobases into Plant Cells**

**[0241]** Any commonly known method used to transform a plant cell can be used for delivering the gene repair oligonucleobases. Illustrative methods are listed below. The methods and compositions herein may involve any of many methods to transfect the cells with the DNA-modifying reagent or reagents. Methods for the introduction of DNA

modifying reagents into a cell or cells are well known in the art and include, but are not limited to, microinjection, electroporation, passive adsorption, calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection, polybrene-mediated transfection, liposome fusion, lipofectin, nucleofection, protoplast fusion, retroviral infection, biolistics (i.e., particle bombardment) and the like.

**[0242]** The use of metallic microcarriers (microspheres) for introducing large fragments of DNA into plant cells having cellulose cell walls by projectile penetration is well known to those skilled in the relevant art (henceforth biolistic delivery). U.S. Pat. Nos. 4,945,050; 5,100,792 and 5,204,253 describe general techniques for selecting microcarriers and devices for projecting them.

**[0243]** Specific conditions for using microcarriers in the methods disclosed herein may include the conditions described in International Publication WO 99/07865. In an illustrative technique, ice cold microcarriers (60 mg/mL), mixed duplex oligonucleotide (60 mg/mL) 2.5 M CaCl<sub>2</sub> and 0.1 M spermidine are added in that order; the mixture gently agitated, e.g., by vortexing, for 10 minutes and then left at room temperature for 10 minutes, whereupon the microcarriers are diluted in 5 volumes of ethanol, centrifuged and resuspended in 100% ethanol. Good results can be obtained with a concentration in the adhering solution of 8-10 µg/µL microcarriers, 14-17 µg/mL mixed duplex oligonucleotide, 1.1-1.4 M CaCl<sub>2</sub> and 18-22 mM spermidine. Optimal results were observed under the conditions of 8 µg/µL microcarriers, 16.5 µg/mL mixed duplex oligonucleotide, 1.3 M CaCl<sub>2</sub> and 21 mM spermidine.

**[0244]** Gene repair oligonucleobases can also be introduced into plant cells using microfibers to penetrate the cell wall and cell membrane. U.S. Pat. No. 5,302,523 to Coffee et al describes the use of silicon carbide fibers to facilitate transformation of suspension maize cultures of Black Mexican Sweet. Any mechanical technique that can be used to introduce DNA for transformation of a plant cell using microfibers can be used to deliver gene repair oligonucleobases for transmutation.

**[0245]** An illustrative technique for microfiber delivery of a gene repair oligonucleobase is as follows: Sterile microfibers (2 µg) are suspended in 150 µL of plant culture medium containing about 10 µg of a mixed duplex oligonucleotide. A suspension culture is allowed to settle and equal volumes of packed cells and the sterile fiber/nucleotide suspension are vortexed for 10 minutes and plated. Selective media are

applied immediately or with a delay of up to about 120 h as is appropriate for the particular trait.

**[0246]** In an alternative embodiment, the gene repair oligonucleobases can be delivered to the plant cell by electroporation of a protoplast derived from a plant part. The protoplasts are formed by enzymatic treatment of a plant part, particularly a leaf, according to techniques well known to those skilled in the art. See, e.g., Gallois et al., 1996, in *Methods in Molecular Biology* 55:89-107, Humana Press, Totowa, N.J.; Kipp et al., 1999, in *Methods in Molecular Biology* 133:213-221, Humana Press, Totowa, NJ. The protoplasts need not be cultured in growth media prior to electroporation. Illustrative conditions for electroporation are 300,000 protoplasts in a total volume of 0.3 mL with a concentration of gene repair oligonucleobase of between 0.6-4  $\mu\text{g/mL}$ .

**[0247]** In an alternative embodiment, nucleic acids are taken up by plant protoplasts in the presence of the membrane-modifying agent polyethylene glycol, according to techniques well known to those skilled in the art. In another alternative embodiment, the gene repair oligonucleobases can be delivered by injecting it with a microcapillary into plant cells or into protoplasts.

**[0248]** In an alternative embodiment, nucleic acids are embedded in microbeads composed of calcium alginate and taken up by plant protoplasts in the presence of the membrane-modifying agent polyethylene glycol (see, e.g., Sone et al., 2002, Liu et al., 2004).

**[0249]** In an alternative embodiment, nucleic acids frozen in water and introduced into plant cells by bombardment in the form of microparticles (see, e.g., Gilmore, 1991, U.S. Patent 5,219,746; Brinegar et al.).

**[0250]** In an alternative embodiment, nucleic acids attached to nanoparticles are introduced into intact plant cells by incubation of the cells in a suspension containing the nanoparticle (see, e.g., Pasupathy et al., 2008) or by delivering them into intact cells through particle bombardment or into protoplasts by co-incubation (see, e.g., Torney et al., 2007).

**[0251]** In an alternative embodiment, nucleic acids complexed with penetrating peptides and delivered into cells by co-incubation (see, e.g., Chugh et al., 2008, WO 2008148223 A1; Eudes and Chugh).

[0252] In an alternative embodiment, nucleic acids are introduced into intact cells through electroporation (see, e.g., He et al., 1998, US 2003/0115641 A1, Dobres et al.).

[0253] In an alternative embodiment, nucleic acids are delivered into cells of dry embryos by soaking them in a solution with nucleic acids (see, e.g., Töpfer et al., 1989, Senaratna et al., 1991) or in other embodiments are introduced by Cellsqueeze (SQZ Biotech), .

**[0254] Selection of Plants**

[0255] In various embodiments, plants as disclosed herein can be of any species of dicotyledonous, monocotyledonous or gymnospermous plant, including any woody plant species that grows as a tree or shrub, any herbaceous species, or any species that produces edible fruits, seeds or vegetables, or any species that produces colorful or aromatic flowers. For example, the plant maybe selected from a species of plant from the group consisting of canola, sunflower, corn, tobacco, sugar beet, cotton, maize, wheat, barley, rice, alfalfa, barley, sorghum, tomato, mango, peach, apple, pear, strawberry, banana, melon, cassava, potato, carrot, lettuce, onion, soy bean, soya spp, sugar cane, pea, chickpea, field pea, fava bean, lentils, turnip, rutabaga, brussel sprouts, lupin, cauliflower, kale, field beans, poplar, pine, eucalyptus, grape, citrus, triticale, alfalfa, rye, oats, turf and forage grasses, flax, oilseed rape, mustard, cucumber, morning glory, balsam, pepper, eggplant, marigold, lotus, cabbage, daisy, carnation, tulip, iris, lily, and nut producing plants insofar as they are not already specifically mentioned.

[0256] Plants and plant cells can be tested for resistance or tolerance to an herbicide using commonly known methods in the art, e.g., by growing the plant or plant cell in the presence of an herbicide and measuring the rate of growth as compared to the growth rate in the absence of the herbicide.

[0257] As used herein, substantially normal growth of a plant, plant organ, plant tissue or plant cell is defined as a growth rate or rate of cell division of the plant, plant organ, plant tissue, or plant cell that is at least 35%, at least 50%, at least 60%, or at least 75% of the growth rate or rate of cell division in a corresponding plant, plant organ, plant tissue or plant cell expressing the wild-type protein of interest.

[0258] As used herein, substantially normal development of a plant, plant organ, plant tissue or plant cell is defined as the occurrence of one or more development events in the plant, plant organ, plant tissue or plant cell that are substantially the same as those

occurring in a corresponding plant, plant organ, plant tissue or plant cell expressing the wild-type protein.

**[0259]** In certain embodiments plant organs provided herein include, but are not limited to, leaves, stems, roots, vegetative buds, floral buds, meristems, embryos, cotyledons, endosperm, sepals, petals, pistils, carpels, stamens, anthers, microspores, pollen, pollen tubes, ovules, ovaries and fruits, or sections, slices or discs taken therefrom. Plant tissues include, but are not limited to, callus tissues, ground tissues, vascular tissues, storage tissues, meristematic tissues, leaf tissues, shoot tissues, root tissues, gall tissues, plant tumor tissues, and reproductive tissues. Plant cells include, but are not limited to, isolated cells with cell walls, variously sized aggregates thereof, and protoplasts.

**[0260]** Plants are substantially “tolerant” to a relevant herbicide when they are subjected to it and provide a dose/response curve which is shifted to the right when compared with that provided by similarly subjected non-tolerant like plant. Such dose/response curves have “dose” plotted on the X-axis and “percentage kill”, “herbicidal effect”, etc., plotted on the y-axis. Tolerant plants will require more herbicide than non-tolerant like plants in order to produce a given herbicidal effect. Plants that are substantially “resistant” to the herbicide exhibit few, if any, necrotic, lytic, chlorotic or other lesions, when subjected to herbicide at concentrations and rates which are typically employed by the agrochemical community to kill weeds in the field. Plants which are resistant to an herbicide are also tolerant of the herbicide.

**[0261] Generation of plants**

**[0262]** Tissue culture of various tissues of plant species and regeneration of plants therefrom is known. For example, the propagation of a canola cultivar by tissue culture is described in any of the following but not limited to any of the following: Chuong et al., “A Simple Culture Method for *Brassica* hypocotyls Protoplasts,” *Plant Cell Reports* 4:4-6, 1985; Barsby, T. L., et al., “A Rapid and Efficient Alternative Procedure for the Regeneration of Plants from Hypocotyl Protoplasts of *Brassica napus*,” *Plant Cell Reports* (Spring, 1996); Kartha, K., et al., “In vitro Plant Formation from Stem Explants of Rape,” *Physiol. Plant*, 31:217-220, 1974; Narasimhulu, S., et al., “Species Specific Shoot Regeneration Response of Cotyledonary Explants of Brassicas,” *Plant Cell Reports* (Spring 1988); Swanson, E., “Microspore Culture in *Brassica*,” *Methods in Molecular Biology*, Vol. 6, Chapter 17, p. 159, 1990.

[0263] Further reproduction of the variety can occur by tissue culture and regeneration. Tissue culture of various tissues of soybeans and regeneration of plants therefrom is well known and widely published. For example, reference may be had to Komatsuda, T. et al., "Genotype X Sucrose Interactions for Somatic Embryogenesis in Soybeans," *Crop Sci.* 31:333-337, 1991; Stephens, P. A., et al., "Agronomic Evaluation of Tissue-Culture-Derived Soybean Plants," *Theor. Appl. Genet.* 82:633-635, 1991; Komatsuda, T. et al., "Maturation and Germination of Somatic Embryos as Affected by Sucrose and Plant Growth Regulators in Soybeans *Glycine gracilis* Skvortz and *Glycine max* (L.) Merr." *Plant Cell, Tissue and Organ Culture*, 28:103-113, 1992; Dhir, S. et al., "Regeneration of Fertile Plants from Protoplasts of Soybean (*Glycine max* L. Merr.); Genotypic Differences in Culture Response," *Plant Cell Reports* 11:285-289, 1992; Pandey, P. et al., "Plant Regeneration from Leaf and Hypocotyl Explants of *Glycine wightii* (W. and A.) VERDC. var. *longicauda*," *Japan J. Breed.* 42:1-5, 1992; and Shetty, K., et al., "Stimulation of *In Vitro* Shoot Organogenesis in *Glycine max* (Merrill.) by Allantoin and Amides," *Plant Science* 81:245-251, 1992. The disclosures of U.S. Pat. No. 5,024,944 issued Jun. 18, 1991 to Collins et al., and U.S. Pat. No. 5,008,200 issued Apr. 16, 1991 to Ranch et al.

**[0264] Exemplary Embodiments**

[0265] In addition to the aspects and embodiments described and provided elsewhere in this disclosure, the following non-limiting list of particular embodiments are specifically contemplated.

1. A method of causing a genetic change in a cell, said method comprising exposing said cell to a DNA cutter and a modified GRON.
2. A cell comprising a DNA cutter and a GRON.
3. The method or cell of any of the preceding embodiments, wherein said cells is one or more species of cell selected from the group consisting of plant, bacteria, yeast, fungi, algae, and mammalian.
4. The method or cell of any of the preceding embodiments, wherein said cells is one or more species of cell selected from the group consisting of *Escherichia coli*, *Mycobacterium smegmatis*, *Baccillus subtilis*, *Chlorella*, *Bacillus thuringiensis*, *Saccharomyces cerevisiae*, *Yarrowia lipolytica*, *Chlamydomonas rhienhardtii*, *Pichia pastoris*, *Corynebacterium*, *Aspergillus niger*, and *Neurospora crassa*. *Arabidopsis*

*thaliana*, *Solanum tuberosum*, *Solanum phureja*, *Oryza sativa*, *Glycine max*, *Amaranthus tuberculatus*, *Linum usitatissimum*, and *Zea mays*

5. The method or cell of any of the preceding embodiments, wherein said cell is *Yarrowia lipolytica*.
6. The method or cell of any of the preceding embodiments, wherein said cell is a yeast cell that is not *Saccharomyces cerevisiae*.
7. A method of causing a genetic change in a plant cell, said method comprising exposing said cell to a DNA cutter and a modified GRON.
8. A plant cell comprising a DNA cutter and a modified GRON.
9. A method of causing a genetic change in a plant cell, said method comprising exposing said cell to a DNA cutter and a GRON that comprises DNA and RNA.
10. A plant cell comprising a DNA cutter that comprises DNA and RNA.
11. A method of causing a genetic change in a Acetyl-Coenzyme A carboxylase (ACCase) gene in a cell, wherein said genetic change causes a change in the Acetyl-Coenzyme A carboxylase (ACCase) protein at one or more amino acid positions, said positions selected from the group consisting of 1781, 1783, 1786, 2078, 2079, 2080 and 2088 based on the numbering of the blackgrass reference sequence SEQ ID NO:1 or at an analogous amino acid residue in an ACCase paralog said method comprising exposing said cell to a modified GRON.
12. A method of causing a genetic change in a Acetyl-Coenzyme A carboxylase (ACCase) gene in a cell, wherein said genetic change causes a change in the Acetyl-Coenzyme A carboxylase (ACCase) protein at one or more amino acid positions, said positions selected from the group consisting of 1781, 1783, 1786, 2078, 2079, 2080 and 2088 based on the numbering of the blackgrass reference sequence SEQ ID NO:1 or at an analogous amino acid residue in an ACCase paralog said method comprising exposing said cell to a DNA cutter and a modified GRON.
13. A method for producing a plant or plant cell, comprising introducing into a plant cell a gene repair oligonucleobase (GRON) with a targeted mutation in an Acetyl-Coenzyme A carboxylase (ACCase) gene to produce a plant cell with an ACCase gene that expresses an ACCase protein comprising a mutation at one or more amino acid positions corresponding to a position selected from the group consisting of 1781,

1783, 1786, 2078, 2079, 2080 and 2088 based on the numbering of the blackgrass reference sequence SEQ ID NO:1 or at an analogous amino acid residue in an ACCase paralog.

14. A method for producing a plant or plant cell, comprising introducing into a plant cell a DNA cutter and a gene repair oligonucleobase (GRON) with a targeted mutation in an Acetyl-Coenzyme A carboxylase (ACCase) gene to produce a plant cell with an ACCase gene that expresses an ACCase protein comprising a mutation at one or more amino acid positions corresponding to a position selected from the group consisting of 1781, 1783, 1786, 2078, 2079, 2080 and 2088 based on the numbering of the blackgrass reference sequence SEQ ID NO:1 or at an analogous amino acid residue in an ACCase paralog.
15. A fertile plant comprising an Acetyl-Coenzyme A carboxylase (ACCase) gene that encodes a protein comprising a mutation at position 2078 based on the numbering of the blackgrass reference sequence SEQ ID NO:1 or at an analogous amino acid residue in an ACCase paralog.
16. A fertile rice plant comprising an Acetyl-Coenzyme A carboxylase (ACCase) gene that encodes a protein comprising a mutation at position 2078 based on the numbering of the blackgrass reference sequence SEQ ID NO:1 or at an analogous amino acid residue in an ACCase paralog.
17. A plant cell comprising an Acetyl-Coenzyme A carboxylase (ACCase) gene that encodes a protein comprising a mutation at position 2078 based on the numbering of the blackgrass reference sequence SEQ ID NO:1 or at an analogous amino acid residue in an ACCase paralog and that further comprises an Acetyl-Coenzyme A carboxylase (ACCase) gene that encodes a protein comprising a mutation at one or more amino acid positions, said positions selected from the group consisting of 1781, 1783, 1786, 2079, 2080 and 2088 based on the numbering of the blackgrass reference sequence SEQ ID NO:1 or at an analogous amino acid residue in an ACCase paralog.
18. A fertile plant comprising an Acetyl-Coenzyme A carboxylase (ACCase) gene that encodes a protein comprising a mutation at position 2078 based on the numbering of the blackgrass reference sequence SEQ ID NO:1 or at an analogous amino acid residue in an ACCase paralog and that further comprises an Acetyl-Coenzyme A carboxylase (ACCase) gene that encodes a protein comprising a mutation at one or

- more amino acid positions, said positions selected from the group consisting of 1781, 1783, 1786, 2079, 2080 and 2088 based on the numbering of the blackgrass reference sequence SEQ ID NO:1 or at an analogous amino acid residue in an ACCase paralog.
19. A method of causing a genetic change in a Acetyl-Coenzyme A carboxylase (ACCase) gene in a cell, wherein said genetic change causes a change in the Acetyl-Coenzyme A carboxylase (ACCase) protein at position 2078 based on the numbering of the blackgrass reference sequence SEQ ID NO:1 or at an analogous amino acid residue in an ACCase paralog said method comprising exposing said cell to a modified GRON.
20. A method of causing a genetic change in a Acetyl-Coenzyme A carboxylase (ACCase) gene in a cell, wherein said genetic change causes a change in the Acetyl-Coenzyme A carboxylase (ACCase) protein at position 2078 based on the numbering of the blackgrass reference sequence SEQ ID NO:1 or at an analogous amino acid residue in an ACCase paralog said method comprising exposing said cell to a DNA cutter and a modified GRON.
21. The method, plant or cell of any of the preceding embodiments, wherein said mutation or change in an Acetyl-Coenzyme A carboxylase (ACCase) gene, if present results in an Acetyl-Coenzyme A carboxylase (ACCase) protein comprising one or more selected from the group consisting of an isoleucine to alanine at a position corresponding to position 1781 of SEQ ID NO:1; an isoleucine to leucine at a position corresponding to position 1781 of SEQ ID NO:1; an isoleucine to methionine at a position corresponding to position 1781 of SEQ ID NO:1; an isoleucine to asparagine at a position corresponding to position 1781 of SEQ ID NO:1; an isoleucine to serine at a position corresponding to position 1781 of SEQ ID NO:1; an isoleucine to threonine at a position corresponding to position 1781 of SEQ ID NO:1; an isoleucine to valine at a position corresponding to position 1781 of SEQ ID NO:1; a glycine to cysteine at a position corresponding to position 1783 of SEQ ID NO:1; an alanine to proline at a position corresponding to position 1786 of SEQ ID NO:1; an aspartate to glycine at a position corresponding to position 2078 of SEQ ID NO:1; an aspartate to lysine at a position corresponding to position 2078 of SEQ ID NO:1; an aspartate to threonine at a position corresponding to position 2078 of SEQ ID NO:1; a serine to phenylalanine at a position corresponding to position 2079 of SEQ ID NO:1; a lysine

to glutamate at a position corresponding to position 2080 of SEQ ID NO:1; a cysteine to phenylalanine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to glycine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to histidine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to lysine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to leucine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to asparagine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to proline at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to glutamine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to arginine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to serine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to threonine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to valine at a position corresponding to position 2088 of SEQ ID NO:1; and a cysteine to a tryptophan at a position corresponding to position 2088 of SEQ ID NO:1.

22. The plant or cell of any of the preceding embodiments, or a plant or plant cell made by any of the methods of the preceding embodiments, wherein said plant or cell comprises an Acetyl-Coenzyme A carboxylase (ACCase) protein comprising one or more selected from the group consisting of an isoleucine to alanine at a position corresponding to position 1781 of SEQ ID NO:1; an isoleucine to leucine at a position corresponding to position 1781 of SEQ ID NO:1; an isoleucine to methionine at a position corresponding to position 1781 of SEQ ID NO:1; an isoleucine to asparagine at a position corresponding to position 1781 of SEQ ID NO:1; an isoleucine to serine at a position corresponding to position 1781 of SEQ ID NO:1; an isoleucine to threonine at a position corresponding to position 1781 of SEQ ID NO:1; an isoleucine to valine at a position corresponding to position 1781 of SEQ ID NO:1; a glycine to cysteine at a position corresponding to position 1783 of SEQ ID NO:1; an alanine to proline at a position corresponding to position 1786 of SEQ ID NO:1; an aspartate to glycine at a position corresponding to position 2078 of SEQ ID NO:1; an aspartate to lysine at a position corresponding to position 2078 of SEQ ID NO:1; an aspartate to threonine at a position corresponding to position 2078 of SEQ ID NO:1; a serine to phenylalanine at a position corresponding to position 2079 of SEQ ID NO:1; a lysine to glutamate at a position corresponding to position 2080 of SEQ ID NO:1; a cysteine

to phenylalanine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to glycine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to histidine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to lysine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to leucine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to asparagine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to proline at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to glutamine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to arginine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to serine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to threonine at a position corresponding to position 2088 of SEQ ID NO:1; a cysteine to valine at a position corresponding to position 2088 of SEQ ID NO:1; and a cysteine to a tryptophan at a position corresponding to position 2088 of SEQ ID NO:1.

23. The plant or cell of any of the preceding embodiments, or a plant or cell made by any of the methods of the preceding embodiments, wherein said plant or plant cell comprises an Acetyl-Coenzyme A carboxylase (ACCCase) gene that encodes a protein comprising a mutation at one or more amino acid positions, said positions selected from the group consisting of 1781, 1783, 1786, 2078, 2079, 2080 and 2088 based on the numbering of the blackgrass reference sequence SEQ ID NO:1 or at an analogous amino acid residue in an ACCCase paralog.
24. The plant or cell of any of the preceding embodiments, or a plant or cell made by any of the methods of the preceding embodiments, wherein said plant or cell comprises an Acetyl-Coenzyme A carboxylase (ACCCase) gene that encodes a protein comprising a mutation at position 2078 based on the numbering of the blackgrass reference sequence SEQ ID NO:1 or at an analogous amino acid residue in an ACCCase paralog and that further comprises an Acetyl-Coenzyme A carboxylase (ACCCase) gene that encodes a protein comprising a mutation at one or more amino acid positions, said positions selected from the group consisting of 1781, 1783, 1786, 2079, 2080 and 2088 based on the numbering of the blackgrass reference sequence SEQ ID NO:1 or at an analogous amino acid residue in an ACCCase paralog.

**[0266]** In each of the foregoing ACCCase embodiments 11-24, whether methods, plants, cells, or otherwise, the following are suitable mutations for use therein:

<b>Amino Acid Change</b>	<b>Codon Change</b>		<b>Amino Acid Change</b>	<b>Codon Change</b>
I1781A	ATA > GCT ATA > GCC ATA > GCA ATA > GCG		C2088F	TGC > TTT TGC > TTC
I1781L	ATA > CTT ATA > CTC ATA > CTA ATA > CTG ATA > TTA ATA > TTG		C2088G	TGC > GGT TGC > GGC TGC > GGA TGC > GGG
I1781M	ATA > ATG		C2088H	TGC > CAT TGC > CAC
I1781N	ATA > AAT ATA > AAC		C2088K	TGC > AAA TGC > AAG
I1781S	ATA > TCT ATA > TCC ATA > TCA ATA > TCG		C2088L	TGC > CTT TGC > CTC TGC > CTA TGC > CTG TGC > TTA TGC > TTG
I1781T	ATA > ACT ATA > ACC ATA > ACA ATA > ACG		C2088N	TGC > AAT TGC > AAC
I1781V	ATA > GTT ATA > GTC ATA > GTA ATA > GTG		C2088P	TGC > CCT TGC > CCC TGC > CCA TGC > CCG
G1783C	GGA > TGT GGA > TGC		C2088Q	TGC > CAA TGC > CAG
A1786P	GCT > CCT GCT > CCC GCT > CCA GCT > CCG		C2088R	TGC > CGT TGC > CGC TGC > CGA TGC > CGG TGC > AGA TGC > AGG
D2078G	GAT > GGT GAT > GGC GAT > GGA GAT > GGG		C2088S	TGC > TCT TGC > TCC TGC > TCA TGC > TCG

Amino Acid Change	Codon Change		Amino Acid Change	Codon Change
D2078K	GAT > AAA		C2088T	TGC > ACT
	GAT > AAG			TGC > ACC
D2078T	GAT > ACT		C2088V	TGC > ACA
	GAT > ACC			TGC > ACG
	GAT > ACA			TGC > GTT
	GAT > ACG			TGC > GTC
S2079F	AGC > TTT		C2088W	TGC > GTA
	AGC > TTC			TGC > GTG
K2080E	AAG > GAA			TGC > TGG
	AAG > GAG			

[0267] Alternative mutations include, but are not limited to, the following:

S2079A	AGC	>	GCT
	AGC	>	GCC
	AGC	>	GCA
	AGC	>	GCG
G1783A	GGA	>	GCT
	GGA	>	GCC
	GGA	>	GCA
	GGA	>	GCG
A1786G	GCT	>	GGT
	GCT	>	GGC
	GCT	>	GGA
	GCT	>	GGG

[0268] With regard to embodiments 11-24, corresponding positions to 1781m 1783, 1786, 2078, 2079, and 2080 based on the numbering of the blackgrass reference sequence are well known in the art and readily obtainable from appropriate sequence databases. By way of example, the following table shows the corresponding positions in the rice ACCase sequence:

<i>Am</i>	<i>OsI</i>	<i>OsJ</i>
I1781	I1792	I1779
G1783	G1794	G1781
A1786	A1797	A1784
D2078	D2089	D2076
S2079	S2090	S2077
K2080	K2091	K2078
C2088	C2099	C2086

*Am*: *Alopecurus myosuroides*; *OsI*: *Oryza sativa indica* variety; *OsJ*: *Oryza sativa japonica* variety

25. A method for producing a plant or plant cell with a mutated EPSPS gene, comprising introducing into a plant cell a gene repair oligonucleobase (GRON) with a targeted mutation in an 5-enol pyruvylshikimate-3-phosphate synthase (EPSPS) gene to produce a plant cell with an EPSPS gene that expresses an EPSPS protein comprising a mutation at one or more amino acid positions corresponding to a position selected from the group consisting of 96, 97 and 101 based on the numbering of the amino acid sequence for the *Escherichia coli* reference sequence SEQ ID NO:2 or at an analogous amino acid residue in an EPSPS paralog.
26. A method for producing a plant or plant cell with a mutated EPSPS gene, comprising introducing into a plant cell a DNA cutter and a gene repair oligonucleobase (GRON) with a targeted mutation in an 5-enol pyruvylshikimate-3-phosphate synthase (EPSPS) gene to produce a plant cell with an EPSPS gene that expresses an EPSPS protein comprising a mutation at one or more amino acid positions corresponding to a position selected from the group consisting of 96, 97 and 101 based on the numbering of the amino acid sequence for the *Escherichia coli* reference sequence SEQ ID NO:2 or at an analogous amino acid residue in an EPSPS paralog.
27. A plant or cell with a mutated EPSPS gene, wherein said plant or cell is made by a method introducing into a plant cell a DNA cutter and a gene repair oligonucleobase (GRON) with a targeted mutation in an 5-enol pyruvylshikimate-3-phosphate synthase (EPSPS) gene to produce a plant cell with an EPSPS gene that expresses an EPSPS protein comprising a mutation at one or more amino acid positions

corresponding to a position selected from the group consisting of 96, 97 and 101 based on the numbering of the amino acid sequence for the *Escherichia coli* reference sequence SEQ ID NO:2 or at an analogous amino acid residue in an EPSPS paralog.

28. The plant or cell of any of the preceding embodiments, or a plant or cell made by any of the methods of the preceding embodiments, wherein the plant or plant cell expresses an EPSPS protein comprising a mutation at one or more amino acid positions are selected from the group consisting of a glycine to alanine at a position corresponding to position 96 of SEQ ID NO:2; a threonine to isoleucine at a position corresponding to position 97 of SEQ ID NO:2; a proline to alanine at a position corresponding to position 101 of SEQ ID NO:2; a proline to serine at a position corresponding to position 101 of SEQ ID NO:2; and a proline to threonine at a position corresponding to position 101 of SEQ ID NO:2.
29. The plant or cell of any of the preceding embodiments, or a plant or cell made by any of the methods of the preceding embodiments, wherein the plant or plant cell expresses an EPSPS protein comprising mutation combinations selected from the group consisting of a threonine to isoleucine at a position corresponding to position 97 of SEQ ID NO:2 and a proline to alanine at a position corresponding to position 101 of SEQ ID NO:2; a threonine to isoleucine at a position corresponding to position 97 of SEQ ID NO:2 and a proline to alanine at a position corresponding to position 101 of SEQ ID NO:2; a threonine to isoleucine at a position corresponding to position 97 of SEQ ID NO:2 and a proline to serine at a position corresponding to position 101 of SEQ ID NO:2; and a threonine to isoleucine at a position corresponding to position 97 of SEQ ID NO:2 and a proline to threonine at a position corresponding to position 101 of SEQ ID NO:2.

**[0269]** With regard to embodiments 25-30, corresponding positions to 96, 97 and 101 based on the numbering of the *Escherichia coli* reference sequence SEQ ID NO:2 are well known in the art and readily obtainable from appropriate sequence databases. See e.g., US Patent No. 8,268,622. By way of example, the following table shows the corresponding positions in the flax EPSPS sequence:

<i>E. coli</i> EPSPS	Flax EPSPS	
	Gene1	Gene2
G96	G176	G177
T97	T177	T178
P101	P181	P182

[0270] *E. coli* EPSPS sequence is AroA having the sequence  
 MESLTLQPIARVDGTINLPGSKTVSNRALLLAALAHGKTVLTNLLDSDDDR  
 HMLNALTALGVSYSYTLADRTRCEIIGNGGPLHAEGAELFLGNAGTAMRPL  
 AAALCLGSNDIVLTGEPKMERPIGHLVDALRLGGAKITYLEQENYPPRLRQ  
 GGFTGGNVDDGVSQFLTALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNL  
 MKTFGVEIENQHYQQFVVKGGQSYQSPGTYLVEGDASSASYFLAAAIAKGG  
 TVKVTGIGRNSMQGDIRFADVLEKMGATICWGDDYISCTRGELNAIDMDM  
 NHIPDAAMTIATAALFAKGTTRLRNIYNWRVKETDRLFAMATELRKVGAE  
 VEEGHDIYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPVTILDPKCTA  
 KTFPDYFEQLARISQAA

[0271] Flax gene 1 sequence is lcl - g41452\_1333 having the sequence  
 MALVTKICGGANAVALPATFGTRRTKSISSVSFRSSTSPPSLKQRRRSGNVAAA  
 AAPLRVSASLTTAAEKASTVPEEVVLQPIKDISGIVTLPGSKSLSNRILLAAALSE  
 GTTVVDNLLNSDDVHYMLGALKTLGLNVEHSSEQKRAIVEGCGGVFPVGLKAK  
 NDIELFLGNAGTAMRPLTAAVTAAGGNSSYILDGVPRMRERPIGDLVVGLKQLG  
 ADVTCSSTSCPPVHVNGQGGLPGGKVKLSGSISSQYLTALLMAAPLALGDVEIEI  
 VDKLISVPYVDMTLKLMERFGVAVEHSGSWDRFFVKGQKYKSPGNAYVEGD  
 ASSASYFLAGAAITGGTITVEGCGTSSLQGDVKFAEVLEKMGAKVIWTENSVTV  
 TGPPRDASGRKHLRAVDVNMNKMPDVAMTLAVVALYADGPTAIRDVASWRV  
 KETERMIAICTELRKLKLGATVEEGPDYCIITPPEKLNIAEIDTYDDHRMAMAFSLAA  
 CADVPVTIRDPGCTKKTFFPDYFEVLERYTKH

[0272] Flax gene 2 sequence is lcl - g40547\_1271 having the sequence  
 MAQVTKICGGANAVALPATFGTRRTKSISSVSFRSSTSPPSLKQRRLLGNVAAA  
 AAAAPLRISASLATAAEKASTVPEEIVLQPIKDISGIVTLPGSKSLSNRILLAAALSE  
 GKTVDNLLNSDDVHYMLGALKTLGLNVEHSSEQKRAIVEGRGGVFPVGLKLG  
 KNDIELFLGNAGTAMRPLTAAVTAAGGNSSYILDGVPRMRERPIGDLVVGLKQL  
 GADVSCSSTSCPPVHVNAKGGPLPGGKVKLSGSISSQYLTALLMAAPLALGDVEI  
 EIVDKLISVPYVDMTLKLMERFGVAVEHSGSWDRFFVKGQKYKSPGNAYVEG  
 DASSASYFLAGAAITGGTITVEGCGTSSLQGDVKFAEVLEKMGAKVTWTETSVT  
 VTGPPRDASGKHLRAVDVNMNKMPDVAMTLAVVALYADGPTAIRDVASWR  
 VKETERMIAVCTELRKLKLGATVEEGPDYCIITPPEKLSIAEIDTYDDHRMAMAFSL  
 AACADVPTIRDPGCTKKTFFPDYFEVLERYTKH

30. The method or cell of any of the preceding embodiments, wherein said DNA cutter is one or more selected from a CRISPR, a TALEN, a zinc finger, meganuclease, and a DNA-cutting antibiotic.
31. The method or cell of any of the preceding embodiments, wherein said DNA cutter is a CRISPR or a TALEN.
32. The method or cell of any of the preceding embodiments, wherein said DNA cutter is a CRISPR.

33. The method or cell of any of the preceding embodiments, wherein said DNA cutter is a TALEN.
34. The method or cell of any of the preceding embodiments, wherein said DNA cutter is one or more DNA-cutting antibiotics selected from the group consisting of bleomycin, zeocin, phleomycin, tallysomyin and pepleomyin.
35. The method or cell of any of the preceding embodiments, wherein said DNA cutter is zeocin.
36. The method or cell of any of the preceding embodiments, wherein said GRON is single stranded.
37. The method or cell of any of the preceding embodiments, wherein the GRON is a chemically protected oligonucleotide.
38. The method or cell of any of the preceding embodiments, wherein the GRON comprises a chemically protected oligonucleotide protected at the 5' end.
39. The method or cell of any of the preceding embodiments, wherein the GRON comprises a chemically protected oligonucleotide protected at the 3' end.
40. The method or cell of any of the preceding embodiments, wherein the GRON comprises a chemically protected oligonucleotide protected at the 5' and 3' ends.
41. The method or cell of any of the preceding embodiments, wherein the GRON comprises one or more selected from a Cy3 group, a 3PS group, and a 2'-O-methyl group.
42. The method or cell of any of the preceding embodiments, wherein the GRON comprises a Cy3 group.
43. The method or cell of any of the preceding embodiments, wherein the GRON comprises a Cy3 group at the first base on the 5' end.
44. The method or cell of any of the preceding embodiments, wherein the GRON comprises a Cy3 group at the first base on the 3' end.
45. The method or cell of any of the preceding embodiments, wherein the GRON comprises a 3PS group.
46. The method or cell of any of the preceding embodiments, wherein the GRON comprises two or more 3PS groups.

47. The method or cell of any of the preceding embodiments, wherein the GRON comprises three or more 3PS groups.
48. The method or cell of any of the preceding embodiments, wherein the GRON comprises a 3PS group at the first base on the 5' end.
49. The method or cell of any of the preceding embodiments, wherein the GRON comprises a 3PS group at the first base on the 3' end.
50. The method or cell of any of the preceding embodiments, wherein the GRON comprises a 2'-O-methyl group.
51. The method or cell of any of the preceding embodiments, wherein the GRON comprises two or more 2'-O-methyl groups.
52. The method or cell of any of the preceding embodiments, wherein the GRON comprises a 2'-O-methyl group at the first base on the 5' end.
53. The method or cell of any of the preceding embodiments, wherein the GRON has a 2'-O-methyl group at the first base on the 5' end and does not have any other 2'-O-methyl groups.
54. The method or cell of any of the preceding embodiments, wherein the GRON comprises a 2'-O-methyl group on each of the first two or more bases at the 5' end.
55. The method or cell of any of the preceding embodiments, wherein the GRON comprises a 2'-O-methyl group on each of the first three or more bases at the 5' end.
56. The method or cell of any of the preceding embodiments, wherein the GRON comprises a 2'-O-methyl group on each of the first four or more bases at the 5' end.
57. The method or cell of any of the preceding embodiments, wherein the GRON comprises a 2'-O-methyl group on each of the first five or more bases at the 5' end.
58. The method or cell of any of the preceding embodiments, wherein the GRON comprises a 2'-O-methyl group on each of the first six or more bases at the 5' end.
59. The method or cell of any of the preceding embodiments, wherein the GRON comprises a 2'-O-methyl group on each of the first seven or more bases at the 5' end.
60. The method or cell of any of the preceding embodiments, wherein the GRON comprises a 2'-O-methyl group on each of the eight four or more bases at the 5' end.

61. The method or cell of any of the preceding embodiments, wherein the GRON comprises a 2'-O-methyl group on each of the first nine or more bases at the 5' end.
62. The method or cell of any of the preceding embodiments, wherein the GRON comprises a 2'-O-methyl group on each of the first ten or more bases at the 5' end.
63. The method or cell of any of the preceding embodiments, wherein the GRON comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 RNA base at the 5' end.
64. The method or cell of any of the preceding embodiments, wherein said GRON has a wobble base pair relative to the target sequence for the genetic change.
65. The method or cell of any of the preceding embodiments, wherein said GRON is between 15 and 60 nucleotides in length.
66. The method or cell of any of the preceding embodiments, wherein said GRON is 41 nucleotides in length.
67. The method or cell of any of the preceding embodiments, wherein said GRON is between 50 and 110 nucleotides in length.
68. The method or cell of any of the preceding embodiments, wherein said GRON is 101 nucleotides in length.
69. The method or cell of any of the preceding embodiments, wherein said GRON is between 150 and 210 nucleotides in length.
70. The method or cell of any of the preceding embodiments, wherein said GRON is 201 nucleotides in length.
71. The method or cell of any of the preceding embodiments, wherein said GRON is between 70 and 210 nucleotides in length.
72. The method or cell of any of the preceding emodiments, wherein said GRON is longer than 70 nucleotides in length.
73. The method or cell of any of the preceding emodiments, wherein said GRON is longer than 100 nucleotides in length.
74. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 165 nucleotides in length.

75. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 175 nucleotides in length.
76. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 185 nucleotides in length.
77. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 195 nucleotides in length.
78. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 200 nucleotides in length.
79. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 210 nucleotides in length.
80. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 220 nucleotides in length.
81. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 230 nucleotides in length.
82. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 240 nucleotides in length.
83. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 250 nucleotides in length.
84. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 260 nucleotides in length.
85. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 270 nucleotides in length.
86. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 280 nucleotides in length.
87. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 290 nucleotides in length.
88. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 300 nucleotides in length.

89. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 400 nucleotides in length.
90. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 500 nucleotides in length.
91. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 600 nucleotides in length.
92. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 700 nucleotides in length.
93. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 800 nucleotides in length.
94. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 900 nucleotides in length.
95. The method or cell of any of the preceding embodiments, wherein said GRON is longer than 1000 nucleotides in length.
96. The method or cell of any of the preceding embodiments wherein said plant is selected from the group consisting of canola, sunflower, corn, tobacco, sugar beet, cotton, maize, wheat, barley, rice, alfalfa, barley, sorghum, tomato, mango, peach, apple, pear, strawberry, banana, melon, cassava, potato, carrot, lettuce, onion, soy bean, soya spp, sugar cane, pea, chickpea, field pea, fava bean, lentils, turnip, rutabaga, brussel sprouts, lupin, cauliflower, kale, field beans, poplar, pine, eucalyptus, grape, citrus, triticale, alfalfa, rye, oats, turf and forage grasses, flax, oilseed rape, mustard, cucumber, morning glory, balsam, pepper, eggplant, marigold, lotus, cabbage, daisy, carnation, tulip, iris, and lily.
97. The method or cell of any of the preceding embodiments wherein said plant is canola.
98. The method or cell of any of the preceding embodiments wherein said plant is corn
99. The method or cell of any of the preceding embodiments wherein said plant is maize.
100. The method or cell of any of the preceding embodiments wherein said plant is rice.
101. The method or cell of any of the preceding embodiments wherein said plant is sorghum.

102. The method or cell of any of the preceding embodiments wherein said plant is potato.
103. The method or cell of any of the preceding embodiments wherein said plant is soy bean.
104. The method or cell of any of the preceding embodiments wherein said plant is flax.
105. The method or cell of any of the preceding embodiments wherein said plant is oilseed rape.
106. The method or cell of any of the preceding embodiments wherein said plant is cassava.
107. The method or cell of any of the preceding embodiments wherein said plant is sunflower.
108. A method of causing a genetic change in a plant cell, said method comprising exposing said cell to a CRISPR and a modified GRON.
109. The method or cell of any of the preceding embodiments wherein multiple genetic changes are made.
110. The method or cell of any of the preceding embodiments wherein two or more guide RNAs are used.
111. The method or cell of any of the preceding embodiments wherein each of the more than one guide RNAs is complimentary to a different target for genetic change.
112. The method or cell of any of the preceding embodiments wherein the CRISPR includes a nickase.
113. The method or cell of any of the preceding embodiments wherein the DNA cutter includes two or more nickases.
114. The method or cell of any of the preceding embodiments wherein two or more nickases cuts on opposite strands of the target nucleic acid sequence.
115. The method or cell of any of the preceding embodiments wherein two or more nickases cuts on the same strand of the target nucleic acid sequence.

116. A non-transgenic herbicide resistant or tolerant plant made by the method or from the cell of one any of the preceding embodiments.
117. The method or cell of any of the preceding embodiments, wherein said plant cell has a genetic change or mutation in Acetyl-Coenzyme A carboxylase (ACCCase) and is selected from the group consisting of barley, maize, millet, oats, rye, rice, sorghum, sugarcane, turf grasses, and wheat.
118. The method or cell of any of the preceding embodiments, wherein said plant cell has a genetic change or mutation in Acetyl-Coenzyme A carboxylase (ACCCase) and is resistant or tolerant to one or more herbicides.
119. The method or cell of any of the preceding embodiments, wherein said plant cell has a genetic change or mutation in Acetyl-Coenzyme A carboxylase (ACCCase), is resistant to one or more ACCCase-inhibiting herbicides.
120. The method or cell of any of the preceding embodiments, wherein said plant cell has a genetic change or mutation in Acetyl-Coenzyme A carboxylase (ACCCase), is resistant to one or more herbicides selected from the group consisting of alloxymid, butoxydim, clethodim, cloproxydim, cycloxydim, sethoxydim, tepraloxymid, tralkoxydim, chlorazifop, clodinafop, clofop, diclofop, fenoxaprop, fenoxaprop-P, fenthiaprop, fluazifop, fluazifop-P, haloxyfop, haloxyfop-P, isoxapyrifop, propaquizafop, quizalofop, quizalofop-P, trifop, pinoxaden, agronomically acceptable salts and esters of any of these herbicides, and combinations thereof.
121. The method or cell of any of the preceding embodiments, wherein said plant cell has a genetic change or mutation in 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), and wherein said plant cell is selected from the group consisting of corn, wheat, rice, barley, sorghum, oats, rye, sugarcane, soybean, cotton, sugarbeet, oilseed rape, canola, flax, cassava, sunflower, potato, tobacco, tomato, alfalfa, poplar, pine, eucalyptus, apple, lettuce, peas, lentils, grape and turf grasses.
122. The method or cell of any of the preceding embodiments, wherein said plant or plant cell has a genetic change or mutation in 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), and wherein plant or plant cell is resistant to at least one herbicide.
123. The method or cell of any of the preceding embodiments, wherein said plant or plant cell has a genetic change or mutation in 5-enolpyruvylshikimate-3-phosphate

- synthase (EPSPS), and wherein plant or plant cell is resistant to a herbicide of the phosphonomethylglycine family.
124. The method or cell of any of the preceding embodiments, wherein said plant or plant cell has a genetic change or mutation in 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), and wherein plant or plant cell is resistant to glyphosate.
125. The method or cell of any of the preceding embodiments, wherein said plant or plant cell has a genetic change or mutation in 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), and wherein plant or plant cell is selected from the group consisting of corn, wheat, rice, barley, sorghum, oats, rye, sugarcane, soybean, cotton, sugarbeet, oilseed rape, canola, flax, cassava, sunflower, potato, tobacco, tomato, alfalfa, poplar, pine, eucalyptus, apple, lettuce, peas, lentils, grape and turf grasses.
126. The method or cell of any of the preceding embodiments, wherein the genetic change or mutation in the cell occurs at one allele of the gene.
127. The method or cell of any of the preceding embodiments, wherein the genetic change or mutation in the cell occurs at two alleles of the gene.
128. The method or cell of any of the preceding embodiments, wherein the genetic change or mutation in the cell occurs at three alleles of the gene.
129. The method or cell of any of the preceding embodiments, wherein the genetic change or mutation in the cell occurs at four alleles of the gene.
130. The method or cell of any of the preceding embodiments, wherein the genetic change or mutation in the cell occurs at one, two, three, four, five, six, seven, eight, nine, ten, eleven, or twelve alleles of the gene.
131. The method or cell of any of the preceding embodiments, wherein the genetic change or mutation in the cell comprises a deletion or insertion resulting in a knockout of one allele of the gene.
132. The method or cell of any of the preceding embodiments, wherein the genetic change or mutation in the cell comprises a deletion or insertion resulting in a knockout of two alleles of the gene.
133. The method or cell of any of the preceding embodiments, wherein the genetic change or mutation in the cell comprises a deletion or insertion resulting in a knockout of three alleles of the gene.

134. The method or cell of any of the preceding embodiments, wherein the genetic change or mutation in the cell comprises a deletion or insertion resulting in a knockout of four alleles of the gene.
135. The method or cell of any of the preceding embodiments, wherein the genetic change or mutation in the cell comprises a deletion or insertion resulting in a knockout of one, two, three, four, five, six, seven, eight, nine, ten, eleven, or twelve alleles of the gene.
136. The method or cell of any of the preceding embodiments, wherein the genetic change or mutation in the cell occurs at one allele of the gene and a second allele of the gene comprises a deletion or insertion resulting in a knockout of said second allele.
137. The method or cell of any of the preceding embodiments, wherein the genetic change or mutation in the cell occurs at one allele of the gene and a second allele and third allele of the gene comprises a deletion or insertion resulting in a knockout of said second allele and said third allele.
138. The method or cell of any of the preceding embodiments, wherein the genetic change or mutation in the cell occurs at one allele of the gene and a second allele, third allele, and fourth allele of the gene comprises a deletion or insertion resulting in a knockout of said second allele, said third allele and said fourth allele.
139. The method or cell of any of the preceding embodiments, wherein the genetic change in the cell comprises at least one mutation at one allele and at least one knockout in another allele.
140. The method or cell of any of the preceding embodiments, wherein the genetic change in the cell comprises at least one mutation at one allele and at least one knockout in at least one other allele.
141. The method or cell of any of the preceding embodiments, wherein the genetic change in the cell comprises at least one mutation at one allele and at least one knockout in at least two other alleles.
142. The method or cell of any of the preceding embodiments, wherein the genetic change in the cell comprises at least one mutation at one allele and at least one knockout in at least three other alleles.

143. The method or cell of any of the preceding embodiments, wherein the genetic change in the cell comprises at least one mutation at one allele and a knockout in all other alleles.

### EXAMPLES

[0273] The following are examples, which illustrate procedures for practicing the methods and compositions described herein. These examples should not be construed as limiting.

**[0274] Example 1: GRON length**

[0275] Sommer et al., (Mol Biotechnol. 33:115-22, 2006) describes a reporter system for the detection of *in vivo* gene conversion which relies upon a single nucleotide change to convert between blue and green fluorescence in green fluorescent protein (GFP) variants. This reporter system was adapted for use in the following experiments using *Arabidopsis thaliana* as a model species in order to assess efficiency of GRON conversion following modification of the GRON length.

[0276] In short, for this and the subsequent examples an *Arabidopsis thaliana* line with multiple copies of a blue fluorescent protein gene was created by methods known to those skilled in the art (see, e.g., Clough and Brent, 1998). Root-derived meristematic tissue cultures were established with this line, which was used for protoplast isolation and culture (see, e.g., Mathur et al., 1995). GRON delivery into protoplasts was achieved through polyethylene glycol (PEG) mediated GRON uptake into protoplasts. A method using a 96-well format, similar to that described by similar to that described by Fujiwara and Kato (2007) was used. In the following the protocol is briefly described. The volumes given are those applied to individual wells of a 96-well dish.

1. Mix 6.25  $\mu$ l of GRON (80  $\mu$ M) with 25  $\mu$ l of *Arabidopsis thaliana* BFP transgenic root meristematic tissue-derived protoplasts at  $5 \times 10^6$  cells/ml in each well of a 96 well plate.
2. 31.25  $\mu$ l of a 40% PEG solution was added and the protoplasts were mixed.
3. Treated cells were incubated on ice for 30 min.
4. To each well 200  $\mu$ l of W5 solution was added and the cells mixed.
5. The plates were allowed to incubate on ice for 30 min allowing the protoplasts to settle to the bottom of each well.

6. 200  $\mu$ l of the medium above the settled protoplasts was removed.
7. 85  $\mu$ l of culture medium (MSAP, see Mathur et al., 1995) was added.
8. The plates were incubated at room temperature in the dark for 48 hours. The final concentration of GRON after adding culture medium is 8  $\mu$ M.

[0277] Forty eight hours after GRON delivery samples were analyzed by flow cytometry in order to detect protoplasts whose green and yellow fluorescence is different from that of control protoplasts (BFP0 indicates non-targeting GRONs with no change compared to the BFP target; C is the coding strand design and NC is the non-coding strand design). A single C to T nucleotide difference (coding strand) or G to A nucleotide targeted mutation (non-coding strand) in the center of the BFP4 molecules. The green fluorescence is caused by the introduction of a targeted mutation in the BFP gene, resulting in the synthesis of GFP. The results are shown in Figure 1.

[0278] Table 2 shows the sequences of exemplary 101-mer and 201-mer BFP4/NC 5'-3PS/ 3'-3PS GRONs designed for the conversion of a blue fluorescent protein (BFP) gene to green fluorescence. "3PS" denotes 3 phosphothioate linkages at each of the 5' and 3' oligo ends.

**[0279] Table 2: Exemplary GRON Nucleotide Sequences for BFP to GFP conversion**

GRON Name	GRON Nucleotide Sequence
BFP4/NC 101-mer	G* T*C*G TGC TGC TTC ATG TGG TCG GGG TAG CGG CTG AAG CAC TGC ACG CCG TAG GTG AAG GTG GTC ACG AGG GTG GGC CAG GGC ACG GGC AGC TTG CCG G*T*G* G
BFP0/NC 101-mer	G* T*C*G TGC TGC TTC ATG TGG TCG GGG TAG CGG CTG AAG CAC TGC ACG CCG TGG GTG AAG GTG GTC ACG AGG GTG GGC CAG GGC ACG GGC AGC TTG CCG G*T*G *G
BFP4/C 101-mer	C *C*A*C CGG CAA GCT GCC CGT GCC CTG GCC CAC CCT CGT GAC CAC CTT CAC CTA CGG CGT GCA GTG CTT CAG CCG CTA CCC CGA CCA CAT GAA GCA GCA C*G*A *C
BFP0/C 101-mer	C*C*A*CCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCAC CTTCACCCACGGCGTGACGTGCTTCAGCCGCTACCCCGACCACATG AAGCAGCAC*G*A* C
BFP4/NC 201-mer	A*A*G*ATGGTGCCTCCTGGACGTAGCCTTCGGGCATGGCGGACT TGAAGAAGTCGTGCTGCTTCATGTGGTCTGGGTAGCGGCTGAAGC ACTGCACGCCGTAGGTGAAGGTGGTCACGAGGGTGGGCCAGGGCA CGGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTTGC CG TAGGTGGCATCGCCCTCG *C*C*C
BFP0/NC	A*A*G*TGGTGCCTCCTGGACGTAGCCTTCGGGCATGGCGGACTT

GRON Name	GRON Nucleotide Sequence
201-mer	GAAGAAGTCGTGCTGCTTCATGTGGTTCGGGGTAGCGGCTGAAGCA CTGCACGCCGTGGGTGAAGGTGGTCACGAGGGTGGGCCAGGGCAC GGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTTGCC G TAGGTGGCATCGCCCTCG *C*C*C
BFP4/C 201-mer	G*G*G*CGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGA CCACCTTCACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCA CATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTA CGTCCAGGA GCGCACCAT *C*T*T
BFP0/C 201-mer	G*G*G*CGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTC ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGA CCACCTTCACCCACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCA CATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTA CGTCCAGGA GCGCACCAT *C*T*T

\* = PS linkage (phosphothioate)

**[0280] Example 2: Conversion rates using 5'Cy3/ 3'idC labeled GRONs**

[0281] The purpose of this series of examples is to compare the efficiencies of phosphothioate (PS) labeled GRONs (having 3 PS moieties at each end of the GRON) to the 5'Cy3/ 3'idC labeled GRONs. The 5'Cy3/ 3'idC labeled GRONs have a 5' Cy3 fluorophore (amidite) and a 3' idC reverse base. Efficiency was assessed using conversion of blue fluorescent protein (BFP) to green fluorescence.

[0282] In all three examples, done either by PEG delivery of GRONs into protoplasts in individual Falcon tubes (labeled "Tubes") or in 96-well plates (labeled "96-well dish"), there was no significant difference between the different GRON chemistries in BFP to GFP conversion efficiency as determined by cytometry (Figure 1).

**[0283] Example 3: Comparison between 41-mer BFP4/NC 5'-3PS/ 3'-3PS GRON and 2'-O-Me GRONs**

[0284] The purpose of this series of examples is to compare the conversion efficiencies of the phosphothioate (PS) labeled GRONs with 3PS moieties at each end of the GRON to 2'-O-Me or "Okazaki fragment GRONs" in the presence and absence of a member of the bleomycin family, Zeocin™ (1 mg/ml) to induce DNA breaks. The designs of these GRONs are depicted in Figure 2. GRONs were delivered into *Arabidopsis thaliana* BFP protoplasts by PEG treatment and BFP to GFP conversion was determined at 24 h post treatment by cytometry. Samples treated with zeocin (1 mg/ml) were incubated with zeocin for 90 min on ice prior to PEG treatment.

[0285] In general the presence of zeocin (1 mg/ml) increased BFP to GFP conversion as determined by cytometry (Table 3). In both the presence and absence of zeocin, the NC Okazaki GRON containing one 2'-O Me group on the first RNA base at the 5' end of the GRON was more efficacious at converting BFP to GFP when compared to the NC Okazaki GRON containing one 2'-O Me group on each of the first nine 5' RNA bases (Figure 2 and Table 3).

[0286] In all examples, there was no significant difference between the 41-mer BFP4/NC 5'3PS/ 3'3PS and the 71-mer Okazaki Fragment BFP4/NC GRON that contains one 5' 2'-O me group on the first 5' RNA base (denoted as BFP4 71-mer (1) NC) in BFP to GFP conversion in both the presence or absence of 1 mg/ml of zeocin as determined by cytometry (Figure 2 and Table 3). It is important to note that in the presence of zeocin (and expected for bleomycin, phleomycin, tallysomycin, pepleomycin and other members of this family of antibiotics) that conversion becomes strand independent (i.e., both coding (C) and non-coding (NC) GRONs with the designs tested in these examples display approximately equal activity).

[0287] **Table 3:** Comparison of a standard GRON design with Okazaki fragment GRON designs in the presence and absence of a glycopeptide antibiotic zeocin.

Zeocin (+)								
Exp. Name	BFP4 41-mer		BFP4 71-mer (0)		BFP4 71-mer (1)		BFP4 71-mer (9)	
	NC	C	NC	C	NC	C	NC	C
APT043	0.13	0.0875	0.2275	0.2075	0.355	0.2275	0.2325	0.195
APT066	1.9	0.713	0.762	0.683	1.318	0.7103	0.769	0.883
Mean	1.015	0.40025	0.49475	0.44525	0.8365	0.4689	0.50075	0.539
Std Dev	1.251579	0.442295	0.377949	0.336229	0.680944	0.341391	0.379363	0.486489
SE	0.885134	0.312797	0.26729	0.237786	0.481573	0.241436	0.268291	0.344052

Zeocin (-)								
Exp. Name	BFP4 41-mer		BFP4 71-mer (0)		BFP4 71-mer (1)		BFP4 71-mer (9)	
	NC	C	NC	C	NC	C	NC	C
APT043	nd	nd	0.1875	0.0175	0.21	0.025	0.1	0.0225
APT066	0.109	0.007	0.112	0.005	0.141	0.023	0.065	0.021
Mean	0.109	0.007	0.14975	0.01125	0.1755	0.024	0.0825	0.02175
Std Dev	na	na	0.053387	0.008839	0.04879	0.001414	0.024749	0.001061
SE	na	na	0.037756	0.006251	0.034505	0.001	0.017503	0.00075

BFP4 71-mer (0)		5' first 10bp are RNA and GRON has no protection
NC	C	

BFP4 71-mer (1)		5' first 10bp are RNA and first bp on the 5' end has a 2' O-Me
NC	C	

BFP4 71-mer (9)		5' first 10bp are RNA and first nine bp on the 5' end has a 2' O-Me
NC	C	

**[0288] Example 4: Comparison between 41-mer, 101-mer and 201-mer BFP4/NC 5'-3PS/ 3'-3PS GRONs**

**[0289]** The purpose of this series of examples was to compare the conversion efficiencies (in the presence and absence of zeocin) of the phosphothioate (PS) labeled GRONs with 3PS moieties at each end of the GRON of different lengths: 41-mer, 101-mer and 201-mer shown in Table 2. Again, the presence of zeocin (1 mg/ml) increased BFP to GFP conversion rates as determined by cytometry (Table 4). The overall trend in all three examples was linear with increasing NC GRON length in both the presence and absence of zeocin. Except for the BFP-4/NC/101 and BFP-4/C/101 in the presence of zeocin, this had conversion rates that were close to equal but lower than the 41-mer NC GRON. This is in contrast to all previous examples in which the BFP-4/41 coding and non-coding GRONs were used, where the non-coding was always far superior to the

coding GRON. This asymmetry in conversion frequency also applies to the BFP-4/201 GRONs used in this example series.

[0290] **Table 4:**

<b>Zeocin (+)</b>						
Exp. Name	BFP4 41-mer		BFP4 101-mer		BFP4 201-mer	
	NC	C	NC	C	NC	C
APT038	0.2425	0.1275	0.3025	0.2575	0.97	0.245
APT043	0.13	0.0875	0.185	0.2275	0.66	0.1875
APT047	0.3975	0.145	0.19	0.125	0.235	0.085
APT052	0.3275	nd	0.17	0.21	0.585	0.225
APT052	nd	nd	0.3225	0.3175	0.5075	0.3125
APT058	1.4275	nd	1.2	nd	1.9	nd
APT066	1.9	0.713	0.992	1.05	1.7	0.916
Mean	0.7375	0.26825	0.480286	0.364583	0.936786	0.3285
Std Dev	0.7382801	0.297475	0.428968	0.341634	0.630943	0.297412
SE	0.30146186	0.148738	0.162119	0.139499	0.238452	0.121442

<b>Zeocin (-)</b>						
Exp. Name	BFP4 41-mer		BFP4 101-mer		BFP4 201-mer	
	NC	C	NC	C	NC	C
APT038	0.05	0.01	0.1025	0.025	0.5725	0.025
APT066	0.109	0.007	0.214	0.047	0.566	0.035
Mean	0.0795	0.0085	0.15825	0.036	0.56925	0.03
Std Dev	0.0417193	0.002121	0.078842	0.015556	0.004596	0.007071
SE	0.02950446	0.0015	0.055758	0.011002	0.00325	0.005001

**[0291] Example 5: Delivery of Cas9 Protein into Plants**

**[0292]** This example makes use of direct delivery of recombinant Cas9 protein to plant cells as an alternative to delivery of CRISPR-Cas expression plasmids. This method employs carriers such as cell penetrating peptides (CPP), transfection liposome reagents, poly(ethylene glycol) (PEG) either alone or in combination to allow for delivery of active recombinant Cas9 protein to cells.

**[0293] Methods**

**[0294]** BFP transgenic *Arabidopsis thaliana* protoplasts derived from induced root tissue are seeded on a flat-bottom 96-well plate at 250,000 cells per well at a cell density of  $1 \times 10^7$  cells/ml. Fluorescently-tagged recombinant Cas9 protein (1  $\mu$ g) pre-coated with CPPs at 20:1, 10: or 5:1 and other CPP to cargo ratio (TAT, Penetratin, Chariot™, PEP-1 or others for example) or encapsulated with liposome reagents are then mixed with the seeded protoplasts and incubated at 23°C for 2-6 h to allow for Cas9/carrier complexes to penetrate the cells. In another series of treatments fluorescently-tagged recombinant Cas9 protein (1  $\mu$ g) either pre-coated with CPPs as described above or not coated are introduced to protoplasts using PEG methodology. Protoplasts were then analyzed by flow cytometry 24 h after treatment in order to determine the percentage of Cas9 positive protoplasts within a given treatment.

**[0295] Example 6: CRISPR with 201-mer  $\pm$  wobble base GRONs**

**[0296]** The purpose of this series of examples is to demonstrate BFP to GFP conversion in our *Arabidopsis thaliana* BFP transgenic model system using CRISPRs to create targeted double-stranded breaks in the *bfp* gene and the 201-mer GRONs to mediate conversion. The BFP CRISPR targets the coding strand of the *bfp* gene and the conversion site is 27 bp upstream of the PAM sequence (Figure 3). The GRON is used as a template to repair the double-stranded break in the *bfp* gene created by the CRISPR and along with converting the targeted gene from BFP to GFP, it introduces a wobble base in the *bfp* gene that corresponds to the PAM sequence of the BFP CRISPR as well. A wobble base in the PAM sequence of the BFP CRISPR is hypothesized to minimize re-cutting of the *bfp* gene by the CRISPRs once conversion has happened. This series of examples will help to address whether or not introducing a wobble base into the PAM

sequence of the BFP CRISPR in converted *bfp* genes will increase conversion efficiencies.

### [0297] Methods

[0298] BFP transgenic *Arabidopsis thaliana* protoplasts derived from induced root tissue were seeded on a flat-bottom 96-well plate, at 250,000 cells per well at a cell density of  $1 \times 10^7$  cells/ml. The CRISPR encoded plasmids contain the mannopine synthase (MAS) promoter driving the Cas9 coding sequence with an *rbcSE9* terminator and the *Arabidopsis* U6 promoter driving the sgRNA with a poly-T10 terminator. The CRISPR plasmids along with GRON were introduced into protoplasts by PEG mediated delivery at a final concentration of 0.05  $\mu\text{g}/\mu\text{l}$  and 0.16  $\mu\text{M}$  respectively. Protoplasts were incubated in the dark at 23°C for 72 hours, and then they were analyzed by flow cytometer in order to determine the percentage of GFP positive protoplasts within a given treatment.

[0299] The CRISPR consists of two components: the plant codon-optimized *Streptococcus pyogenes* Cas9 (SpCas9) and sgRNA both of which were expressed from the same plasmid. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA region contains the spacer sequence used to guide the Cas9 nuclease to the BFP target gene. In this example the BFP targets the *bfp* gene (Figure 3). 201-mer GRONs targeting BFP with or without wobble bases were used to determine their effect on the rate of BFP to GFP conversion. Table 5 gives a list of the GRONs and their corresponding sequences.

### [0300] Table 5. List of GRONs used in these examples

GRON Name	GRON Chemistry	GRON Sequence
BFP4/NC 201-mer	3PS	5'AAGATGGTGCCTCCTGGACGTAGCCTTCGGGCATGGCGGACTGAAGAAGTCGTGCTGCTCATGTGGTCTGGGTAGCGGCTGAAGCACTGCACCCGTAGGTGAAGGTGTCACGAGGGTGGCCAGGGCACGGGCAGCTTGCCGGTGGTGAGATGAACTCAGGGTCAAGCTTCGGTAGGTGGCATCGCCCTCGCCC3'
BFP4/C 201-mer	3PS	5'GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCATCTGCACCACGGCAAGCTGCCCGTCCCTGGCCACCCTCGTGACCACCTTACCTACGGCGTGCAAGTCTCAGCCGCTACCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCAGGAGCGCACCATCTT3'
BFP4/NC 201-mer (1 wobble)	3PS	5' AAGATGGTGCCTCCTGGACGTAGCCTTCGGGCATGGCGGACTGAAGAAGTCGTGCTGCTCATGTGGTCTGGGTAGCGGCTGAAGCACTGCACCCGTAGGTGAAGGTGTCACGAGGGTGGGCGAGGGCACGGGCAGCTTGCCGGTGGTGAGATGAACTCAGGGTCAAGCTTCGGTAGGTGGCATCGCCCTCGCCC 3'
BFP4/C 201-mer (1 wobble)	3PS	5' GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCATCTGCACCACGGCAAGCTGCCCGTCCCTGGCCACCCTCGTGACCACCTTACCTACGGCGTGCAAGTCTCAGCCGCTACCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCAGGAGCGCACCATCTT 3'

**[0301] Results**

**[0302]** Using the BFP CRISPR, the BFP4/C GRON with the wobble bases is up to a 3.5-fold more efficacious in BFP to GFP conversion when compared to the BFP4/C GRON without the wobble bases (Table 6). There is up to a 5.9-fold increase in BFP to GFP conversion when the BFP4/C GRON with the wobble base is used instead of the BFP4/NC GRON with the wobble base (Table 6). Therefore, the BFP4/C GRON with the wobble base is most efficacious in BFP to GFP conversion when used with the BFP CRISPR.

**[0303] Conclusions**

**[0304]** Including a wobble base in the GRON that changes the PAM sequence of the BFP CRISPR in the converted target gene increases BFP to GFP conversion. BFP to GFP conversion by the BFP CRISPR along with the wobble-based GRON was confirmed by Next Generation Sequencing (data not shown). Additionally, the ability of the BFP CRISPR to cleave the DNA and produce indels in the *bfp* gene was confirmed by Next Generation Sequencing (data not shown).

**[0305] Table 6.** The percentage of BFP to GFP conversion as determine by cytometry at 72 h post PEG delivery of the BFP CRISPR and GRON into protoplasts derived from the *Arabidopsis thaliana* BFP transgenic line 21-15-09.

Exp. Name	Percentage of GFP Positive Cells (std dev)			
	CRISPR: BFP5			
	BFP4/C GRON		BFP4/NC GRON	
	(-) Wobbles	(+) 1 wobbles	(-) Wobbles	(+) 1 wobbles
Exp 1	0.46(0.07)	1.59(0.06)	0.08(0.02)	0.27(0.04)
Exp 2	0.24(0.03)	0.61(0.05)	0.04(0.01)	0.16(0.04)

**[0306] Example 7: CRISPR with Cy3 modified GRONs**

**[0307]** The purpose of this series of examples is to demonstrate BFP to GFP conversion in our *Arabidopsis thaliana* BFP transgenic model system using CRISPRs to create targeted double-stranded breaks in the *bfp* gene and GRONs to mediate conversion. The BFP6 CRISPR (CR:BFP6) used in these examples targets the *bfp* gene and causes a double-stranded break in the DNA near the site of conversion. The GRONs used with the BFP6 CRISPR, contains the coding sequence of the *bfp* gene around the site of

conversion and are labeled at the 5' end with Cy3 and at the 3' end with an idC reverse base and are herein referred to as Cy3 GRONs . These GRONs are tested at three different lengths of 41-mers, 101-mers and 201-mers and they are directly compared to the 3PS GRONs which only differ from the Cy3 GRONs in that they have 3 phosphothioate linkages on both the 5' and 3' ends of the GRON. These GRONs are herein referred to as 3PS GRONs. See Table 7 for the list of GRONs used in these examples.

### [0308] Methods

[0309] BFP transgenic *Arabidopsis thaliana* protoplasts derived from induced root tissue were seeded on a flat-bottom 96-well plate, at 250,000 cells per well at a cell density of  $1 \times 10^7$  cells/ml. The CRISPR encoded plasmids contain the MAS promoter driving the Cas9 coding sequence with an rbcSE9 terminator and the *Arabidopsis thaliana* U6 promoter driving the sgRNA with a poly-T10 terminator. The CRISPR plasmids along with GRON were introduced into protoplasts by PEG mediated delivery at a final concentration of  $0.05 \mu\text{g}/\mu\text{l}$  for the CRISPR and  $8.0 \mu\text{M}$  for the 41-mer,  $0.32 \mu\text{M}$  for the 101-mer and  $0.16 \mu\text{M}$  201-mer GRONs. GRON treatments alone received a final GRON concentration after PEG delivery of  $8.0 \mu\text{M}$  for the 41-mer,  $5.0 \mu\text{M}$  for the 101-mer and  $2.5 \mu\text{M}$  for the 201-mer. Protoplasts were incubated in the dark at  $23^\circ\text{C}$  for 72 hours, and then they were analyzed by flow cytometer in order to determine the percentage of GFP positive protoplasts within a given treatment.

[0310] The CRISPR consists of two components: the plant codon-optimized *Streptococcus pyogenes* Cas9 (SpCas9) and sgRNA both of which were expressed from the same plasmid. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA region contains the spacer sequence used to guide the Cas9 nuclease to the BFP target gene. In this experiment the BFP6 CRISPR (5'GGTGCCGCACGTCACGAAGTCGG 3') was used which targets the *bfp* gene. The GRONs contain the coding sequence of the *bfp* gene near the site of conversion. Table 7 gives a list of the GRONs used.

**[0311] Table 7. List of GRONs used in these examples**

GRON Name	GRON Chemistry	GRON Sequence	CRISPR
BFP4/C 41-mer	Cy3 or 3PS	5' CCCTCGTGACCACCTTCACCTACGGCGTGCAGTGCTTCAGC 3'	BFP6
BFP4/C 101-mer	Cy3 or 3PS	5'CCACCGGCAAGCTGCCCCTGCCCTGGCCCACCCTCGTGACCACCTTCACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGAC 3'	BFP6
BFP4/C 201-mer	Cy3 or 3PS	5'GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCTGCCCTGGCCCACCCTCGTGACCACCTTCACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTT3'	BFP6

**[0312] Results**

**[0313]** Using the BFP6 CRISPR, the Cy3 GRONs at all lengths tested are able to mediate BFP to GFP conversion as efficiently as the 3PS GRONs (Figure 4). Overall, the samples containing the BFP6 CRISPR and GRON have higher levels of BFP to GFP conversion when compared to the GRON only samples (Figure 4), demonstrating the positive impact CRISPRs have on increasing conversion rates.

**[0314] Example 8: CRISPR with GRONs of varying size**

**[0315]** The purpose of this series of examples is to demonstrate BFP to GFP conversion in our *Arabidopsis thaliana* BFP transgenic model system using CRISPRs to create targeted double-stranded breaks in the *bfp* gene and GRONs of varying lengths to mediate conversion. The BFP CRISPR used in these examples targets the *bfp* gene and causes a double-stranded break in the DNA near the site of conversion. The GRONs used with the BFP CRISPR, contains the coding sequence of the *bfp* gene around the site of conversion and are labeled at both the 5' end and the 3' end with 3 phosphothioate linkages and are herein referred to as 3PS GRONs. These GRONs are tested at three different lengths of 60-mers, 101-mers and 201-mers and they are directly compared to the GRON only treatments. See Table 8 for the list of GRONs used in these examples.

**[0316] Methods**

**[0317]** BFP transgenic *Arabidopsis thaliana* protoplasts derived from induced root tissue were seeded on a flat-bottom 96-well plate, at 250,000 cells per well at a cell density of  $1 \times 10^7$  cells/ml. The CRISPR encoded plasmids contain the MAS promoter driving the Cas9 coding sequence with an rbcSE9 terminator and the *Arabidopsis thaliana* U6 promoter driving the sgRNA with a poly-T10 terminator. The CRISPR plasmids along with GRON were introduced into protoplasts by PEG mediated delivery at a final concentration of 0.05  $\mu\text{g}/\mu\text{l}$  for the CRISPR and 0.547  $\mu\text{M}$  for the 60-mer, 0.32

$\mu\text{M}$  for the 101-mer and 0.16  $\mu\text{M}$  201-mer GRONs. GRON treatments alone received a final GRON concentration after PEG delivery of 7.5  $\mu\text{M}$  for the 60-mer, 5.0  $\mu\text{M}$  for the 101-mer and 2.5  $\mu\text{M}$  for the 201-mer. Protoplasts were incubated in the dark at 23°C for 72 hours, and then they were analyzed by flow cytometry in order to determine the percentage of GFP positive protoplasts within a given treatment.

**[0318]** The CRISPR consists of two components: the plant codon-optimized *Streptococcus pyogenes* Cas9 (SpCas9) and sgRNA both of which were expressed from the same plasmid. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA region contains the spacer sequence used to guide the Cas9 nuclease to the BFP target gene. The BFP CRISPR spacer sequence is 5'GTCGTGCTGCTTCATGTGGT3'. In this example the BFP CRISPR was used which targets the *bfp* gene. The GRONs contain the coding sequence of the *bfp* gene near the site of conversion. Table 8 gives a list of the GRONs used.

**[0319] Table 8. List of GRONs used in these examples.**

GRON Name	GRON Chemistry	GRON Sequence
BFP4/C 60-mer (1 wobble)	3PS	5'GTGACCACCTTCACCTACGGCGTGAGTCTTCAGCCGCTACCCAGACCACATGAAGCAG 3'
BFP4/C 101-mer (1 wobble)	3PS	5'CCACCGGCAAGCTGCCGTGCCCTGGCCACCCTCGTGACACCTTCACCTACGGCGTGAGTCTTCAGCCGCTACCCAGACCACATGAAGCAGCAGCAGAC 3'
BFP4/C 201-mer (1 wobble)	3PS	5'GGGCGAGGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCGTGCCCTGGCCACCCTCGTGACACCTTCACCTACGGCGTGAGTCTTCAGCCGCTACCCAGACCACATGAAGCAGCAGCAGCAGCAGACTTCTTCAAGTCCGCATGCCCGAAGGCTACGTCAGGA GCGCACCATCTT 3'

### **[0320] Results**

**[0321]** Using the BFP CRISPR, GRONs at lengths  $\geq 101$  nt are better at mediating BFP to GFP conversion when directly compared to the 60 nt long GRONs (Figure 5). Overall, the samples containing the BFP CRISPR and GRON have higher levels of BFP to GFP conversion when compared to the GRON only samples (Figure 5), demonstrating the positive impact CRISPRs have on increasing conversion rates. This data further demonstrates that the length of the GRON that is most efficacious in mediating BFP to GFP conversion, when used along with the CRISPR, needs to be  $\geq 101$  nt in length.

### **[0322] Example 9: CRISPR with 2'-O-Me GRONs**

**[0323]** The purpose of this examples is to demonstrate BFP to GFP conversion in our *Arabidopsis thaliana* BFP transgenic model system using CRISPRs to create targeted double-stranded breaks in the *bfp* gene and GRONs to mediate conversion. The BFP CRISPR used in this example targets the *bfp* gene and causes a double-stranded break in

the DNA near the site of conversion. The GRONs used with the BFP CRISPR, contain either the coding or non-coding sequence of the *bfp* gene around the site of conversion with the first ten 5' bases of the GRON being RNA bases instead of DNA bases. These RNA bases are labeled with 2'-O-Me group(s) at either the first 5' RNA base or the first nine 5' RNA bases as depicted in Figure 6. These GRONs are herein referred to as 2'-O-Me GRONs and are directly compared to the 3PS GRONs of similar lengths that contain DNA bases with 3 phosphothioate linkages on both the 5' and 3' ends of the GRON. These GRONs are herein referred to as 3PS GRONs. See Table 9 for the list of GRONs used in these examples.

#### [0324] Methods

[0325] BFP transgenic *Arabidopsis thaliana* protoplasts derived from induced root tissue were seeded on a flat-bottom 96-well plate, at 250,000 cells per well at a cell density of  $1 \times 10^7$  cells/ml. The CRISPR encoded plasmids contain the MAS promoter driving the Cas9 coding sequence with a *rbcSE9* terminator and the *Arabidopsis* U6 promoter driving the sgRNA with a poly-T10 terminator. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The CRISPR plasmids along with GRON were introduced into protoplasts by PEG mediated delivery at a final concentration of  $0.05 \mu\text{g}/\mu\text{l}$  for the CRISPR,  $0.5 \mu\text{M}$  for the 71-mer and  $0.16 \mu\text{M}$  for the 201-mer GRONs. GRON treatments alone received a final GRON concentration after PEG delivery of  $5.0 \mu\text{M}$  for the 71-mer and  $2.5 \mu\text{M}$  for the 201-mer. Protoplasts were incubated in the dark at  $23^\circ\text{C}$  for 72 hours, and then they were analyzed by flow cytometer in order to determine the percentage of GFP positive protoplasts within a given treatment.

[0326] The CRISPR consists of two components: the plant codon-optimized *Streptococcus pyogenes* Cas9 (SpCas9) and sgRNA both of which were expressed from the same plasmid. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA region contains the spacer sequence used to guide the Cas9 nuclease to the BFP target gene. The BFP CRISPR spacer sequence is 5'CTCGTGACCACCTTCACCCA 3'. In this example the BFP CRISPR was used which targets the *bfp* gene. The GRONs contain either the coding or non-coding sequence of the *bfp* gene near the site of conversion. Table 9 shows a list of the GRONs used.

**[0327] Table 9. List of GRONs used in these examples**

GRON Name	GRON Chemistry	GRON Sequence
BFP4/C 71-mer	3PS	5'GCUGCCCGUGCCCTGGCCACCCCTCGTGACCACCTTACCTACGGCGTGCAGTGCTTCAGCC GCTACCCCG3'
BFP4/NC 71-mer	3PS	5'TTCATGTGGTCGGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGGT GGCCAGGG3'
BFP4/C 201-mer	3PS	5'GGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACGGCAAGC TGCCCGTGCCCTGGCCACCCCTCGTGACCACCTTACCTACGGCGTGCAGTGCTTCAGCCGCTA CCCCGACCACATGAAGCAGCAGCACTTCTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAG CGCACCATCTT3'
BFP4/NC 201-mer	3P	5'AAGATGGTGCCTCTGGACGTAGCCTTCGGGCATGGCGGACTGAAGAAGTCGTGCTGCTT CATGTGGTCTGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGGTGGG CCAGGGCACGGGCGAGCTTGCCGGTGGTGCAGATGAACCTCAGGGTCAGCTTGCCGTAGGTGGC ATGCCCTCGCCC3'
BFP4/C 71-mer	2'-O-Me	5'gugcccgugCCCTGGCCACCCCTCGTGACCACCTTACCTACGGCGTGCAGTGCTTCAGCCG TACCCCG 3'
BFP4/NC 71-mer	2'-O-Me	5'uucaugugguCGGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGGT GGCCAGGG3'
BFP4/C 201-mer	2'-O-Me	5'ggcgagggcGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACGGCAAGCTG CCCGTGCCCTGGCCACCCCTCGTGACCACCTTACCTACGGCGTGCAGTGCTTCAGCCGCTACC CCGACCACATGAAGCAGCAGCACTTCTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCG CACCATCTT3'
BFP4/NC 201-mer	2'-O-Me	5'aagauggugGCTCTTGGACGTAGCCTTCGGGCATGGCGGACTGAAGAAGTCGTGCTGCTTCA TGTGGTCGGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGGTGGG CAGGGCACGGGCGAGCTTGCCGGTGGTGCAGATGAACCTCAGGGTCAGCTTGCCGTAGGTGGCA TCGCCCTCGCCC3'

**[0328] Results**

**[0329]** The 71-mer and 201-mer 2'-O-Me GRONs had similar BFP to GFP conversion when compared to the various different types of GRON protections of (0), (1) or (9) using the BFP CRISPRs (Figures 7 and 8). The 2'-O-Me GRONs are more efficacious than their 3PS GRON counterparts at mediating BFP to GFP conversion using the BFP CRISPRs (Figures 7 and 8).

**[0330] Example 10: CRISPR Nickases with GRONs Introduction**

**[0331]** The purpose of this examples is to demonstrate BFP to GFP conversion in our *Arabidopsis thaliana* BFP transgenic model system using CRISPRs to create targeted single-stranded nicks in the *bfp* gene and GRONs to mediate conversion. The BFP1 CRISPR (CR:BFP1) used in this example targets the *bfp* gene and contains mutations in the catalytic residues (D10A in RuvC and H840A in HNH) that causes single-stranded nicks in the DNA of the *bfp* gene near the site of conversion on either the DNA strands complementary or non-complementary to the guide RNA respectively. These CRISPRs are herein referred to as BFP1 CRISPR nickase D10A and BFP1 CRISPR nickase H840A and are used either alone or with BFP5 sgRNA on a separate plasmid. When multiple CRISPR nickases are used together in this example, they can either nick the same DNA strand or opposite DNA strands. When both Cas9 proteins that contain the same

mutations, either D10A or H840A, are used together, they nick the same strand of DNA. Conversely, when two Cas9 proteins are used together and one of them contains the D10A mutation and the other one contains the H840A mutation, they nick opposite strands of the DNA. The GRONs used with the nickase CRISPRs, contains either the coding or the non-coding sequence of the *bfp* gene around the site of conversion with one wobble base located in the PAM sequence of BFP5 CRISPR. These GRONs have 3 phosphothioate linkages on both the 5' and 3' ends and are herein referred to as 3PS GRONs. See Table 10 for the list of GRONs used in these examples. The nickase CRISPRs are directly compared to their CRISPR counterparts that are able to cause targeted double-stranded breaks in the DNA of the *bfp* gene.

### [0332] Methods

[0333] BFP transgenic *Arabidopsis thaliana* protoplasts derived from induced root tissue are seeded on a flat-bottom 96-well plate, at 250,000 cells per well at a cell density of  $1 \times 10^7$  cells/ml. The CRISPR encoded plasmids contain the MAS promoter driving the Cas9 coding sequence with an *rbcSE9* terminator and the *Arabidopsis thaliana* U6 promoter driving the sgRNA with a poly-T10 terminator. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The Cas9 gene contains mutations in the catalytic residues, either D10A in RuvC or H840A in HNH. The CRISPR plasmids along with GRON are introduced into protoplasts by PEG mediated delivery at a final concentration of  $0.05 \mu\text{g}/\mu\text{l}$  for the CRISPR and  $0.16 \mu\text{M}$  for the 201-mer. GRON treatments alone received a final GRON concentration after PEG delivery of  $2.5 \mu\text{M}$  for the 201-mer. Protoplasts were incubated in the dark at  $23^\circ\text{C}$  for 72 hours, and then they were analyzed by flow cytometer in order to determine the percentage of GFP positive protoplasts within a given treatment.

[0334] The CRISPR consists of two components: the plant codon-optimized *Streptococcus pyogenes* Cas9 (SpCas9) and sgRNA both of which were expressed from the same plasmid. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA region contains the spacer sequence used to guide the Cas9 nuclease to the BFP target gene. In this example the BFP1 and BFP5 sgRNA was used that targets different regions the *bfp* gene near the site of conversion. The BFP1 spacer (5'CTCGTGACCACCTTCACCCA 3') targets the coding-strand while the BFP5 spacer (5'GTCGTGCTGCTTCATGTGGT3') targets the non-coding strand of the *bfp*

gene. The GRONs contain either the coding or non-coding sequence of the *bfp* gene near the site of conversion. Table 10 shows a list of the GRONs used.

**[0335] Table 10. List of GRONs used in these examples**

GRON Name	GRON Chemistry	GRON Sequence	CRISPR
BFP4/NC 201-mer (1 wobble;BFP5)	3PS	5' AAGATGTTGCGCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCG TGCTGCTTCATGTGGTCTGGTAGCGGCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTACGAGGGTGG GGCCAGGGCACGGGCAGCTTCCCGTGGTGCAGATGAACCTCAGGGTCAGCTTGCCG TAGGTGGCATCGCCCTCGCCC 3'	BFP1 and BFP5
BFP4/C 201-mer (1 wobble;BFP5)	3PS	5' GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCCTGAAGTTCATCTGCACCACCG GCAAGCTGCCCCGTGCCCTGGCCCA CCTCGTGACCACTTCACTACGGCGTGCAGTGTTCAGCCGCTAC CCAGACCATGAAGCAGCACGACTTCTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGA GCGCACCATCTT 3'	BFP1 and BFP6

**[0336] Results**

**[0337]** Both of the CRISPR nickases (D10A and H840A) are more efficient at mediating BFP to GFP conversion when the BFP1 CRISPR and the BFP5 CRISPR were used together instead of separately (Figure 9). In addition, when the BFP1 and BFP5 D10A CRISPR nickases are used together with the C/201 1W GRON, the BFP to GFP conversion is significantly higher when compared to treatments where these CRISPR nickases are used with the NC/201 1W GRON (Figure 9). When the BFP1 and BFP5 H840A CRISPR nickases are used together roughly the same level of BFP to GFP conversion is observed with either the C/201 or NC/201 1W GRONs (Figure 9). These levels of BFP to GFP conversion are slightly higher than when the BFP5 CRISPR is used alone and slightly lower than when the BFP1 CRISPR is used alone (Figure 9).

**[0338] Example 11: Use of CRISPRs to Target Multiple Genes**

**[0339]** The purpose of this example is to demonstrate conversion of multiple genes simultaneously in a given population of protoplasts derived from the *Arabidopsis thaliana* model system using CRISPRs to create double-stranded breaks in targeted genes and GRONs to mediate conversion. The CRISPRs used in this example target both the BFP and acetohydroxy acid synthase (AHAS) genes in the *Arabidopsis thaliana* genome by introducing into protoplasts plasmid(s) encoding the Cas9 gene and multiple sgRNA targeting these two different genes. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). This will allow Cas9 to cause double-stranded breaks in both the BFP and AHAS genes in the presence of GRONs that will mediate their conversion.

**[0340] Methods**

**[0341]** *Arabidopsis thaliana* protoplasts derived from induced root tissue are seeded on a flat-bottom 96-well plate, at 250,000 cells per well at a cell density of  $1 \times 10^7$  cells/ml. The CRISPR encoded plasmids contain the MAS promoter driving the Cas9 coding sequence with a rbcSE9 terminator and *Arabidopsis thaliana* U6 promoter driving multiple different sgRNAs with a poly-T10 terminator. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The CRISPR plasmids along with GRON are introduced into protoplasts by PEG mediated delivery at a final concentration of  $0.05 \mu\text{g}/\mu\text{l}$  for the CRISPR and  $0.16 \mu\text{M}$  for the 201-mer. GRON treatments alone receive a final GRON concentration after PEG delivery of  $2.5 \mu\text{M}$  for the 201-mer. Protoplasts will be incubated in the dark at  $23^\circ\text{C}$  for 72 hours, and then they are analyzed by flow cytometer and an allele specific PCR assay in order to determine the percentage of both BFP to GFP and AHAS converted protoplasts respectively within a given treatment.

**[0342]** In the an allele specific PCR assay 10-16 replicates of 5,000 genome equivalents of genomic DNA were used in the primary PCR reactions.

**[0343]** The CRISPR consists of two components: the plant codon-optimized *Streptococcus pyogenes* Cas9 (SpCas9) and sgRNA both of which were expressed from the same or multiple plasmids. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA region contains the spacer sequence used to guide the Cas9 nuclease to the targeted genes. In this example different sgRNAs and GRONs are used to target multiple genes near their sites of conversion; BFP spacer (5'CTCGTGACCACCTTCACCCA 3') and AHAS spacer (5'TGGTTATGCAATTGGAAGATCGC 3'). Table 11 describes the GRONs used.

**[0344] Table 11. List of GRONs used in this example**

GRON Name	Target Gene	GRON Chemistry	GRON Sequence
BFP/C 201-mer	BFP	3PS	5'GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCA CCGGCAAGCTGCCCGTGCCCTGGCCACCCTCGTGACCACCTTCACCTACGGCGT GCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCC GCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTT3'
AHAS(W) 574/NC 201-mer	AHAS	3PS	5'AGCTGCTGCAACAGCAACATGTTTCGGGAATATCTCGTCTCTGAGCCGGATC CCCGAGAATGTGTGAGCTCGGTTAGCTTTGTAGAAGCGATCTTCAATTGCATA ACCATGCCAAGATGCTGTTGTTTAATAAAAAGTACCTTCACTGGAAGATTCTCTAC ACGAATAGTGGCTAGCTCTTGACATTCAATATAAA3'

**[0345] Results**

**[0346]** BFP to GFP and AHAS conversion was determined at 144 h post PEG delivery of the BFP and AHAS CRISPR plasmids and the BFP/C 201-mer and AHAS(W)574/NC 201-mer GRONs into the *Arabidopsis thaliana* BFP transgenic line. Flow cytometry data revealed that Treatment 1 resulted in 0.20% BFP to GFP conversion (Table 12). Allele specific PCR assay revealed that Treatment 1 resulted in 0.01% AHAS converted protoplasts (Table 12). GRON only treatments had minimal conversion using both assays (Table 12). This example demonstrates the successful simultaneous conversion of two independent target genes (BFP and AHAS) within a given population of protoplasts derived from the *Arabidopsis thaliana* BFP model system.

**[0347] Table 12.** Measurement of conversion of both the BFP and AHAS genes at 144 h post PEG delivery of CRISPR plasmids and GRONs into a given population of protoplasts derived from the *Arabidopsis thaliana* BFP model system either by: (1) flow cytometry which determines the percentage of GFP positive protoplasts or (2) allele specific PCR which determines the percentage of AHAS converted protoplasts.

Treatment	CRISPR	GRONs	BFP to GFP conversion	AHAS - W574L Conversion
			Flow Cytometry	Allele Specific PCR
1	CR-BFP and CR-AHAS	BFP/C 201-mer and AHAS(W)574/NC 201-mer	0.20%	~0.01%
2	None	BFP/C 201-mer and AHAS(W)574/NC 201-mer	0.01%	~0.001%
3	None	None	0.01%	~0.001%

**[0348] Example 12: Delivery of Cas9 mRNA into Plant Cells**

**[0349]** This example makes use of direct delivery of recombinant Cas9 mRNA into plant cells as an alternative to delivery of CRISPR-Cas expression plasmids. This method includes (1) in vitro synthesis of modified mRNA and (2) Delivery of this modified mRNA into plant cells.

**[0350] Methods**

**[0351]** A Cas9 mRNA will be transcribed in vitro using an RNA polymerase such as T7, T3 or SP6 from a linearized plasmid template including components of a 5'UTR, the coding sequence (CDS) for the protein and a 3'UTR. One RNA polymerase may incorporate a particular modified nucleoside better than another. The 5' UTR may contain elements that improve its stability such as the MiR-122 of hepatitis C virus (Shimakami et al., 2012). In vitro synthesis will incorporate nucleosides that are

protective and ensure good translation in the target plant cells. Recombinant Cas9 mRNA will be capped and contain a polyA tail.

[0352] BFP transgenic *Arabidopsis thaliana* protoplasts derived from induced root tissue are seeded on a flat-bottom 96-well plate at 250,000 cells per well at a cell density of  $1 \times 10^7$  cells/ml. Recombinant Cas9 mRNA will be delivered into plant cells in one of the following means (list not inclusive): cell penetrating peptides (CPP), transfection liposome reagents, poly(ethylene glycol) (PEG) either alone or in combination to allow for delivery of active recombinant Cas9 mRNA into cells.

**[0353] Example 13: CRISPR-Cas for tethering DNA, RNA or proteins.**

[0354] This example makes use of a modified single guide RNA (sgRNA) cassette wherein a linker sequence (may also be referred to as a tethering sequence) is included at the 3' end of the tracrRNA but 5' of the RNA polymerase III termination signal as shown in the example below (Figure 10). The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). Though preferred, the placement of the linker is not limited to the 3' end of the tracrRNA but will be investigated at several positions within the sgRNA cassette. The linker sequence may vary in nucleotide length or contain secondary structure that would improve tethering or increase the number of molecules tethered through triplex interactions.

[0355] The linker will allow for Watson-Crick base pairing with a DNA, RNA or proteins that contains the complementary sequence (Figure 10). Additionally, linker sequence in sgRNA cassettes will be designed to contain advanced secondary and tertiary structure allowing for more complex multifaceted interaction regions that would tether multiple DNA, RNA or protein molecules.

[0356] The overall concept is that a CRISPR-Cas complex will tether biological molecules to the site of nuclease activity thereby increasing the likelihood of gene editing. These biological molecules include the GRON which will mediate conversion of targeted gene(s). Tethering linkers can be added to sgRNA by simply using, for example, Gene Strings or annealed oligos.

[0357] Methods

[0358] BFP transgenic *Arabidopsis thaliana* protoplasts derived from induced root tissue are seeded on a flat-bottom 96-well plate, at 250,000 cells per well at a cell density

of  $1 \times 10^7$  cells/ml. The CRISPR-Cas tethering plasmids contains the MAS promoter driving the Cas9 coding sequence with a *rbcSE9* terminator and

[0359] *Arabidopsis thaliana* U6 promoter driving sgRNAs with a linker sequence that is complementary to a polynucleotide tract of 15-30 bp located on a 201-mer editing GRON targeting *bfp* (as shown in Figure 10). The sgRNA tethering cassette is terminated by a poly-T10 terminator. The CRISPR-Cas plasmids along with GRON are introduced into protoplasts by PEG mediated delivery at a final concentration of  $0.05 \mu\text{g}/\mu\text{l}$  for the CRISPR and  $0.16 \mu\text{M}$  of the 201-mer GRON. GRON treatments alone received a final GRON concentration after PEG delivery of  $2.5 \mu\text{M}$  for the 201-mer. Protoplasts were incubated in the dark at  $23^\circ\text{C}$  for 72 hours, and then they were analyzed by flow cytometer in order to determine the percentage of GFP positive protoplasts within a given treatment.

[0360] The CRISPR consists of two components: the plant codon-optimized *Streptococcus pyogenes* Cas9 (SpCas9) and sgRNA both of which are expressed from the same or different plasmids. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA) and linker. The crRNA region contains the spacer sequence used to guide the Cas9 nuclease to the BFP target gene. In this example CRISPR is used which targets the *bfp* gene. The GRONs contain either the coding or non-coding sequence of the *bfp* gene near the site of conversion.

[0361] **Example 14: CRISPRs with truncated gRNA.**

[0362] The purpose of this example is to demonstrate conversion of BFP to GFP in protoplasts derived from the *Arabidopsis thaliana* BFP model system using CRISPRs to create double-stranded breaks in targeted genes and GRONs to mediate conversion. The CRISPRs used in this example targets the *bfp* gene in the *Arabidopsis thaliana* genome by introducing into protoplasts plasmid(s) encoding the Cas9 gene and one sgRNAs that is two different lengths. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA which guides the Cas9 to the target genes is called the spacer and it is typically 20-nt in length (CR:BFP1 20-nt), however, in these examples we tested the effectiveness of using a smaller length spacer of 17-nt (CR:BFP1 17-nt) in mediating BFP to GFP conversion.

**[0363] Methods**

**[0364]** *Arabidopsis thaliana* protoplasts derived from induced root tissue are seeded on a flat-bottom 96-well plate, at 250,000 cells per well at a cell density of  $1 \times 10^7$  cells/ml. The CRISPR encoded plasmids contain the MAS promoter driving the Cas9 coding sequence with a rbcSE9 terminator and *Arabidopsis thaliana* U6 promoter driving multiple different sgRNAs with a poly-T10 terminator. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The CRISPR plasmids along with GRON are introduced into protoplasts by PEG mediated delivery at a final concentration of  $0.05 \mu\text{g}/\mu\text{l}$  for the CRISPR and  $0.16 \mu\text{M}$  for the 201-mer. GRON treatments alone receive a final GRON concentration after PEG delivery of  $2.5 \mu\text{M}$  for the 201-mer. Protoplasts will be incubated in the dark at  $23^\circ\text{C}$  for 72 hours, and then they are analyzed by flow cytometer in order to determine the percentage of BFP to GFP within a given treatment.

**[0365]** The CRISPR consists of two components: the plant codon-optimized *Streptococcus pyogenes* Cas9 (SpCas9) and sgRNA both of which were expressed from the same or multiple plasmids. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA region contains the spacer sequence used to guide the Cas9 nuclease to the targeted genes. In these examples, two different length BFP1 spacers of 20-nt (5'CTCGTGACCACCTTCACCCA 3') vs. 17-nt (5'GTGACCACCTTCACCCA 3') were tested. Table 13 describes the GRON used

**[0366] Table 13. List of GRON used in this example**

GRON Name	GRON Chemistry	GRON Sequence
BFP4/NC 201-mer 3W	3PS	5'AAGATGGTGCCTCCTGGACGTAGCCTTCGGGCATGGCGGACTTGAAGAAGTCG TGCTGCTTCATGTGGTCGGGGTAGCGGCTGAAGCACTGCACGCCGTACGTAACGTGGTCACG AGGGTGGGCCAGGGCAGGGCAGCTTGCCGGTGGTGCAGATGAAGTTCAGGGTCAGCTTGCCG TAGGTGGCATCGCCCTCGCC 3'

**[0367] Results**

**[0368]** Reducing the length of the BFP1 protospacer from 20bp to 17bp had similar levels of BFP to GFP conversion of 0.163% vs. 0.177% respectively at 72 h post PEG delivery of plasmids and GRONs into the *Arabidopsis thaliana* BFP model system (Figure 11).

**[0369] Example 15: CRISPRs with amplicon gRNA.**

**[0370]** The purpose of this Example is to demonstrate conversion of BFP to GFP in protoplasts derived from the *Arabidopsis thaliana* BFP model system using CRISPRs to create double-stranded breaks in targeted genes and GRONs to mediate conversion. The CRISPRs used in this Example targets the *bfp* gene in the *Arabidopsis thaliana* genome by introducing into protoplasts plasmid(s) encoding the *Cas9* gene and one sgRNAs that is either encoded on a plasmid or introduced into protoplasts as an amplicon. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA guides the Cas9 to the target genes, where Cas9 creates a double-stranded break and the GRON is used as a template to convert BFP to GFP in a site-directed manner.

**[0371] Methods**

**[0372]** *Arabidopsis thaliana* protoplasts derived from induced root tissue are seeded on a flat-bottom 96-well plate, at 250,000 cells per well at a cell density of  $1 \times 10^7$  cells/ml. The CRISPR encoded plasmids contain the MAS promoter driving the Cas9 coding sequence with a rbcSE9 terminator and *Arabidopsis* U6 promoter driving multiple different sgRNAs with a poly-T<sub>10</sub> terminator. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The CRISPR plasmids along with GRON are introduced into protoplasts by PEG mediated delivery at a final concentration of 0.05  $\mu\text{g}/\mu\text{l}$  for the CRISPR and 0.16  $\mu\text{M}$  for the 201-mer. GRON treatments alone receive a final GRON concentration after PEG delivery of 2.5  $\mu\text{M}$  for the 201-mer. Protoplasts will be incubated in the dark at 23°C for 72 hours, and then they are analyzed by flow cytometer in order to determine the percentage of BFP to GFP within a given treatment.

**[0373]** The CRISPR consists of two components: the plant codon-optimized *Streptococcus pyogenes* Cas9 (SpCas9) and sgRNA both of which were expressed from the same or multiple plasmids. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA region contains the spacer sequence used to guide the Cas9 nuclease to the targeted genes. In these examples, the same BFP6 gRNA (5'GGTGCCGCACGTCACGAAGTCGG 3') was delivered into protoplasts either as an amplicon or encoded on a plasmid. Table 14 describes the GRONs used.

**[0374] Table 14. List of GRONs used in this example**

GRON Name	GRON Chemistry	GRON Sequence
BFP4/NC 201-mer	3PS	5'AAGATGGTGGCTCCTGGACGTAGCCTTCGGGCATGGCGGACTGAAGAAGTCGTGCTGCTCATGTGGTCTGGGTAGCGGCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGGTGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATGAACTTCAGGGTCAGCTGCCGTAGGTGGCATCGCCCTCGCCC3'
BFP4/C 201-mer	3PS	5'GGGCGAGGGGCGATGCCACTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCTCGTGACCACCTTCACCTACGGCGTGCAAGTCTTCAGCCGCTACCCGACCATGAAGCAGCAGCACTTCTCAAGTCCGCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTT3'

**[0375] Results**

**[0376]** Delivery of the BFP6 gRNA as an amplicon (CR:BFP6 (gRNA amplicon)) along with a plasmid containing only Cas9 had similar rates of BFP to GFP conversion when compared to treatments with both the gRNA (gRNA plasmid) and Cas9 being encoded on separate plasmids at 72 h post PEG delivery of plasmids and GRONs into the *Arabidopsis thaliana* BFP model system (Figure 12).

**[0377] Example 16: CRISPRs with Unmodified GRONs.**

**[0378]** The purpose of this example is to demonstrate BFP to GFP conversion in our *Arabidopsis thaliana* BFP transgenic model system using CRISPRs to create targeted double-stranded breaks in the *bjf* gene and GRONs to mediate conversion. The BFP CRISPR used in this example targets the *bjf* gene and causes a double-stranded break in the DNA near the site of conversion. The 3PS GRONs contain DNA bases with 3 phosphothioate linkages on both the 5' and 3' ends of the GRON and are herein referred to as 3PS GRONs. The 3PS GRONs were directly compared to their unmodified GRON counterparts in mediated BFP to GFP conversion using the BFP CRISPRs in our BFP transgenic *Arabidopsis thaliana* model system. See Table 15 for the list of GRONs used in these examples

**[0379] Table 15. List of GRONs used in this example**

GRON Name	GRON Chemistry	GRON Sequence
BFP4/C 41-mer	3PS	5'CCCTCGTGACCACCTTCACCTACGGCGTGCAAGTCTTCAGC 3'
BFP4/NC 41-mer	None	5'GCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGG3'

**[0380] Methods**

**[0381]** BFP transgenic *Arabidopsis thaliana* protoplasts derived from induced root tissue were seeded on a flat-bottom 96-well plate, at 250,000 cells per well at a cell density of  $1 \times 10^7$  cells/ml. The CRISPR encoded plasmids contain the MAS promoter driving the Cas9 coding sequence with a rbcSE9 terminator and the *Arabidopsis* U6 promoter driving the sgRNA with a poly-T10 terminator. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The CRISPR plasmids along with GRON were introduced into protoplasts by PEG mediated delivery at a final concentration of 0.05  $\mu\text{g}/\mu\text{l}$  for the CRISPR, 0.16  $\mu\text{M}$  for the 41-mer GRONs. GRON treatments alone received a final GRON concentration after PEG delivery of 0.8  $\mu\text{M}$  for the 41-mer. Protoplasts were incubated in the dark at 23°C for 72 hours, and then they were analyzed by flow cytometer in order to determine the percentage of GFP positive protoplasts within a given treatment.

**[0382]** The CRISPR consists of two components: the plant codon-optimized *Streptococcus pyogenes* Cas9 (SpCas9) and sgRNA both of which were expressed from the same plasmid. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA region contains the spacer sequence used to guide the Cas9 nuclease to the BFP target gene. The BFP CRISPR spacer sequence is 5'CTCGTGACCACCTTCACCCA 3'. In this example the BFP CRISPR was used which targets the bfp gene. The GRONs contain the non-coding sequence of the bfp gene near the site of conversion. Table 16 shows a list of the GRONs used

**[0383] Table 16. List of GRONs used in this example.**

GRON	Spacer	CRISPR
BFP4/C 41-mer	3PS	5'CCCTCGTGACCACCTTCACCTACGGCGTGCAAGTGCCTCAGC 3'
BFP4/NC 41-mer	None	5'GCTGAAGCACTGCACGCCGTAGGTGAAGGTGGTCACGAGGG3'

**[0384] Results**

**[0385]** The 41-mer 3PS GRONs are more efficacious than their unmodified GRON counterparts at mediating BFP to GFP conversion using the BFP CRISPRs (Figure 13).

**[0386] Example 17: TALENs and GRONs in Flax.**

**[0387]** The purpose of this example is to demonstrate EPSPS conversion in flax at both 24 hours in protoplasts and 3-weeks in microcalli after delivery of TALEN plasmids and GRONs. The TALENs used in this example targets the *epsps* gene in the *Linum*

*usitatissimum* genome by introducing into shoot tip derived protoplasts plasmid(s) encoding TALENs creates a double-stranded break and the GRON is used as a template to convert the *epsps* gene in a site-directed manner.

#### [0388] Methods

[0389] Flax protoplasts were isolated from shoot tips obtained from *in vitro* germinated seedlings. The TALEN plasmids along with GRONs were introduced into protoplasts by PEG mediated delivery at a final concentration of 0.05 µg/µl and 0.5 µM respectively. Protoplasts were incubated in the dark at 25°C for up to 48 h in liquid medium, or embedded in alginate beads (5 x 10<sup>5</sup> cells/ml), and cultured in liquid medium to induce cell division and the formation of microcalli. Protoplasts or microcalli samples obtained 24 h or 3 weeks after DNA delivery were analyzed by NGS to determine the percentage of cells (DNA reads) carrying out the target mutation within a given treatment. The percent of indels generated by imperfect NHEJ-mediated DNA repair was also estimated.

[0390] TALEN constructs include two arms (left and right), each consisting of a TAL effector-like DNA binding domain linked to a catalytic DNA cleavage domain of FokI. The TAL effector-like DNA binding domain guides the TALEN arms to specific sites of DNA which allows the FokI endonucleases of each arm to dimerize together and cleave double-stranded DNA. The TALEN encoded plasmids contains MasP::LuEPSPS\_(Left arm)-T2A-LuEPSPS\_(right arm) with a *rbcSE9* terminator. LuEPSPS\_(left arm) sequence is 5'TGGAACAGCTATGCGTCCG 3' and the LuEPSPS\_(right arm) sequence is 5'TGAGTTGCCTCCAGCGGCT 3'. GRONs (144-mers) targeting LuEPSPS with or without wobble bases were used to determine their effect on rate of conversion.

#### [0391] Results

[0392] 24 hour protoplasts and 3-week old microcalli have 0.067% and 0.051% EPSPS conversion respectively as determined by Next Generation Sequencing (Figure 14). Additionally, these data show that the TALEN is active and able to cleave the *epsps* target gene in *Linum usitatissimum* and form indels of 2.60% and 1.84% respectively at 24 hours in protoplasts and up to 3-week in microcalli. Moreover, EPSPS conversion and indels are maintained up to 3 weeks after the TALEN plasmid and GRON are introduced.

#### [0393] Example 18: CRISPRs and GRONs in Flax.

[0394] The purpose of this example is to demonstrate activity of Cas9 in flax microcalli three and six weeks after delivery of a Cas9 plasmid. The CRISPRs used in this example

targets the *epsps* gene in the *Linum usitatissimum* genome by introducing into shoot tip derived protoplasts plasmid(s) encoding the *Cas9* gene and a sgRNAs. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA guides the *Cas9* to the target genes, where *Cas9* creates a double-stranded break in the *epsps* gene in a site-directed manner. The double-stranded breaks in the *epsps* gene when repaired by the ubiquitous NHEJ pathway will cause indels to form around the cleavage site.

#### [0395] Methods

[0396] Flax protoplasts were isolated from shoot tips obtained from *in vitro* germinated seedlings. The CRISPR encoded plasmids contains the MAS promoter driving the *Cas9* coding sequence with an *rbcSE9* terminator and the *Arabidopsis thaliana* U6 promoter driving the sgRNA with a poly-T<sub>10</sub> terminator. The CRISPR plasmids were introduced into protoplasts by PEG mediated delivery at a final concentration of 0.05 µg/µl. Protoplasts were embedded in alginate beads (5 x 10<sup>5</sup> cells/ml), cultured in liquid medium, and incubated in a rotatory shaker (30 rpm) in the dark at 25°C. Microcalli developed from individual cells were analyzed by NGS, 3 and 6 weeks after CRISPR plasmid delivery, to determine the percentage of cells (DNA reads) carrying out indels generated by the error-prone NHEJ-mediated DNA repair pathway.

[0397] The CRISPR consists of two components: the plant codon-optimized *Streptococcus pyogenes* *Cas9* (SpCas9) and sgRNA both of which were expressed from the same plasmid. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA region contains the spacer sequence used to guide the *Cas9* nuclease to the target gene. In this example the CRISPR targets the *epsps* gene.

#### [0398] Results

[0399] 3- and 6-week old microcalli have 46.5% and 54.7% indel formation respectively as determined by Next Generation Sequencing (Figure 15). These data shows that *Cas9* is active and able to cleave the *EPSPS* target gene in *Linum usitatissimum* and form indels. Moreover, these indels are maintained up to 6 weeks after the CRISPR plasmid was introduced.

**[0400] Example 19: Construction of engineered nucleases****[0401] CRISPR-Cas**

**[0402]** For construction of transient CRISPR-Cas9 expression plasmids, a higher plant codon-optimized SpCas9 gene containing a SV40 NLS at both the N- and C-terminal and a 2x FLAG tag on the N-terminal was synthesized as a series of GeneArt® Strings™ (Life Technology, Carlsbad, CA), then cloned downstream of the mannopine synthase (MAS) promoter and upstream of the pea ribulose biphosphate carboxylase (rbcS9) terminator by Gibson's method. Next, an sgRNA cassette consisting of a chimeric gRNA whose expression is driven by the *Arabidopsis* U6 promoter, was synthesized as GeneArt® Strings™, then shuttled into the Cas9 containing construct using Gibson's method forming pBCRISPR. To specify the chimeric gRNA for the respective target sequence, pairs of DNA oligonucleotides encoding the variable 20-nt sgRNA targeting sequences (Extended Data Table 2) were annealed to generate short double strand fragments with 4-bp overhangs. The fragments were ligated into BbsI digested pBCrispr to yield CRISPR-Cas constructs BC-1, BC-2 and BC-3.

**[0403] TALEN**

**[0404]** Design and construction of TALEN expression constructs BT-1 and LuET-1 was based on rules as described in Cermak et al., *Nucleic Acids Res.* **39**, e82 (2011). The target sequence was selected based on the gene editing site and the repeat variable i-residue (RVD) following the rules that NG, HD, NI, and NN recognize T, C, A, and G, respectively. The assembly of TAL effector domain linked to the heterodimeric FokI domains was completed through a commercial service (GeneArt; Life Technologies). TALEN monomers were cloned downstream of the MAS promoter and upstream of the rbcE9 terminator using Gibson's method and expressed as a 2A coupled unit.

**[0405] Cell Culture and protoplast isolation**

**[0406]** Surface-sterilized *Arabidopsis* seeds were germinated on solid ½ MS medium (Minerals and vitamins according to XX; ½ concentrated; 87.7 mM sucrose) at 28°C under a 12 h light/dark cycle. Root material from 2 to 3-week-old seedlings were collected and maintained in ½ MS liquid medium under low light conditions at 28°C. Root cultures were transferred and maintained in MSAR[0.22% 1/2 MS, 87.7 mM sucrose, 11.4 µM IAA, 2.3 µM 2,4-D, 1.5 µM 2iP, pH 5.8] three weeks prior to protoplast isolation. Roots were cut into approximately 6 mm segments and incubated in MSAP

solution [0.22% 1/2 MS, 87.7 mM sucrose, 11.4  $\mu$ M IAA, 2.3  $\mu$ M 2,4-D, 1.5  $\mu$ M 2iP, and 400 mM mannitol, pH 5.8] containing cell wall digesting enzymes [1.25% Cellulase RS, 0.25% Macerozyme R-10, 0.6 M mannitol, 5 mM MES, 0.1% BSA] for 3-4 h in the dark with gentle shaking. The released protoplasts were collected and passed through a sterile 100  $\mu$ m filter and 35  $\mu$ m filter. The protoplast filtrate was mixed with 0.8 times the volume of Optiprep™ Density Gradient Medium (Sigma) and mixed gently. A 60% W5 [154 mM NaCl, 5 mM KCl, 125 mM CaCl<sub>2</sub>•2H<sub>2</sub>O, 5 mM glucose, 10 mM MES, (pH 5.8)] / 40% Optiprep solution followed by 90% W5/10% Optiprep solution were slowly layered onto the filtrate/Optiprep solution to make a gradient, which was centrifuged at 198 RCF for 10 min. The white protoplast layer was collected and mixed with 2 times the volume of W5. Protoplasts were centrifuged at 44 RCF for 10 min and re-suspended in TM solution [14.8 mM MgCl<sub>2</sub>•6H<sub>2</sub>O, 5 mM MES, 572 mM mannitol, (pH 5.8)] at a density of  $1 \times 10^7$  cells/ml. For experiments with Zeocin™ (Life Technologies, Carlsbad, CA) and phleomycin (InvivoGen, San Diego, CA), protoplasts were kept in TM adjusted to pH 7.0 for 90 min on ice before transfection. For antibiotic concentrations see Extended Data Figure 1.

**[0407]** Flax protoplasts were isolated from shoot tips obtained from 3-week-old seedlings germinated *in vitro*. Shoot tips were finely chopped with a scalpel, pre-plasmolyzed for 1 h at room temperature in B-medium [B5 salts and vitamins (Gamborg et al., 1968), 4 mM CaCl<sub>2</sub>, 0.1 M glucose, 0.3 M mannitol, 0.1 M glycine, 250 mg/l casein hydrolysate, 10 mg/l L-cystein-HCL, 0.5% polyvinylpyrrolidone (MW 10,000), 0.1% BSA, 1 mg/l BAP, 0.2 mg/l NAA, and 0.5 mg/l 2,4-D], and incubated in a cell wall digesting enzyme solution containing B-medium supplemented with 0.66% Cellulase YC and 0.16% Macerozyme R-10 over a rotatory shaker (50 rpm) at 25 °C for 5 h. Released protoplasts were sieved and purified by density gradient centrifugation using Optiprep (Sigma) layers, counted with a hemocytometer, and kept stationary overnight in the dark at a density of  $0.5 \times 10^6$  protoplasts/ml in B medium.

**[0408] Protoplast transfection**

**[0409]** In a 96-well flat bottom plate,  $2.5 \times 10^5$  cells per well were transfected with 10 pmol GRON, 10 pmol GRON plus 3.25  $\mu$ g CRISPR-Cas or TALEN expression construct or mock using PEG [270 mM mannitol, 67.5 mM Ca(NO<sub>3</sub>)<sub>2</sub>, 38.4% PEG 1500, (pH 5.8)]. Cells and DNA were incubated with PEG on ice for 30 minutes followed by a wash with 200  $\mu$ l of W5 solution. Finally, 85  $\mu$ l of MSAP++ [MSAP containing 50 nM

phytosulfokine- $\alpha$  and 20  $\mu$ M n-propyl gallate] was added and the cells cultured in low light conditions at 28°C.

**[0410]** After about 18 h of culture, protoplasts were transfected with TALEN plasmid along with GRONs (20  $\mu$ g plasmid and 0.2 nmol GRON/ $10^6$  protoplasts) using PEG mediated delivery. Treated protoplasts were incubated in the dark at 25°C for up to 48 h in B medium, or embedded in alginate beads 24 h after transfection, and cultured in V-KM liquid medium to induce cell division and the formation of microcalli. For the antibiotic experiments,  $1.25 \times 10^5$  cells per well were transfected with 8  $\mu$ M GRON CG13 using the PEG solution described above.

#### **[0411] Cytometry**

Seventy-two h after transfection, cells were analyzed by cytometry using the Attune® Acoustic Focusing cytometer (Applied Biosystems®) with excitation and detection of emission as appropriate for *GFP*. Background level was based on PEG-treated protoplasts without DNA delivery. For antibiotic experiments, protoplasts treated with Zeocin or phleomycin prior to transfection were analyzed by cytometry 48 h after transfection.

#### **[0412] Sequencing analysis**

**[0413]** Genomic DNA was extracted from CRISPR-Cas or TALEN-treated protoplasts using the NucleoSpin® Plant II kit as per the manufacturer's recommendations (Machery-Nagel, Bethlehem, PA). Amplicons flanking the TALEN or CRISPR target region were generated using Phusion® polymerase with 100 ng of genomic DNA and primers BFPP-1 (5'-GGACGACGGCAACTACAAGACC-3')/BFPR-1 (5'-TAAACGGCCACAAGTTCAGC-3') for *Arabidopsis* CRISPR and TALEN; or LuEPF-1 (5'-GCATAGCAGTGAGCAGAAGC-3')/LuEPR-1 5'-AGAAGCTGAAAGGCTGGAAG-3' for *L. usitatissimum* TALEN. The amplicons were purified and concentrated using Qiaquick MinElute columns (Qiagen, Valencia, CA). Deep sequencing of the amplicons was performed by GeneWiz (South Plainfield, NJ) using a 2 x 250 bp MiSeq run (Illumina, San Diego, CA). For data analysis fastq files for read 1 and read 2 were imported into CLC Genomics Workbench 7.0.4 (CLCBio, Boston, MA). Paired reads were merged into a single sequence if their sequences overlapped. A sequence for an amplicon was identified if it or its reverse and complemented sequence contained both forward and reverse primer sequences. Occurrence of a unique sequence in a sample was recorded as its abundance. Percent indel or gene edit was calculated by

dividing the number of reads with the edit or indel by the total number of sequenced reads, and then multiplying by 100.

**[0414]** Sequence of CRISPR-Cas photospacers

Name	Sequence (5' to 3')	Figure Reference
BC-1	CTCGTGACCACCTTCACCCA	1a, 1b, 2a, 2c
BC-2	GTCGTGCTGCTTCATGTGGT	2b
BC-3	GGCTGAAGCACTGCCAGCCCG	2d

**[0415]** TALEN binding domain sequences

Paper Name	Sequence (5' to 3')	Figure Reference
BT-1	Left arm: TGGTCGGGGTAGCGGCTGA Right arm: TCGTGACCACCTTCACCCA	3a, 3b
LuET-1	Left arm: TGGAACAGCTATGCGTCCG Right arm: TGAGTTGCCTCCAGCGGCT	3c, 3d

**[0416]** GRON sequences used

Name	Sequence (5' to 3')	Chemistry	Upper Reference
C61	GCTGAGCACTGCACGCCGTAGGTGAAGTGGTCACGAGGG	Unmodified	2e
C62	GCTTGAAGCACTGCACGCCGTAGGTGAAGTGGTCACGAGG	(*) = 3PS	2a
C63	CCTCCTGGTGCACCCCTTCACTTACGGGGTGCAGTCTTCAGC	(*) = 3PS	2d, 3a
C64	VCCCTCGTGCACCCCTTCACTTACGGGGTGCAGTCTTCAGCH	V=C/Y, H=3DMT dC CPG	2d
C65	GTTGADCACTTCACTTACGGGGTGCAGTCTTCAGGGTACCAGACACATGAAGCTAG	(*) = 3PS	2b
C66	AATGATGCTCTCTGACCTTACGGGGTGCAGTCTTCAGGGTACCAGACACATGAAGCTAG CGGCTGAGCACTGCACGCCGTAGGTGAAGTGGTCACGAGGGTGGCCAGGGCACGCTTGCCTGGTGGTGCAG ATGAATTCAGGGTCAAGTGGCTTACGGGGTGCAGTCTTCAGGGTACCAGACATGAAGCTAG	(*) = 3PS	1b, 2c
C67	GCTGAGCACTGCACGCCGTAGGTGAAGTGGTCACGAGGG CCACCTGTGACCACTTCACTTACGGGGTGCAGTCTTCAGGGTACCAGACATGAAGCTAG AAGTCCGCACTGCACGCCGTAGGTGAAGTGGTCACGAGGG	(*) = 3PS	3b
C68	GCTGAGCACTGCACGCCGTAGGTGAAGTGGTCACGAGGG CCACCTGTGACCACTTCACTTACGGGGTGCAGTCTTCAGGGTACCAGACATGAAGCTAG AAGTCCGCACTGCACGCCGTAGGTGAAGTGGTCACGAGGG	(*) = 3PS	2b
C69	AATGATGCTCTCTGACCTTACGGGGTGCAGTCTTCAGGGTACCAGACATGAAGCTAG CGGCTGAGCACTGCACGCCGTAGGTGAAGTGGTCACGAGGGTGGCCAGGGCACGCTTGCCTGGTGGTGCAG ATGAATTCAGGGTCAAGTGGCTTACGGGGTGCAGTCTTCAGGGTACCAGACATGAAGCTAG	(*) = 3PS	1a
C610	GCTGAGCACTGCACGCCGTAGGTGAAGTGGTCACGAGGG GAACTTCAAGGGTCAAGTGGCTTACGGGGTGCAGTCTTCAGGGTACCAGACATGAAGCTAG (*) = 3PS	Lower Case = RNA bases; (base) = 2'-O-Me; Upper Case=DNA bases	2c
C611	GCTGAGCACTGCACGCCGTAGGTGAAGTGGTCACGAGGG GAACTTCAAGGGTCAAGTGGCTTACGGGGTGCAGTCTTCAGGGTACCAGACATGAAGCTAG (*) = 3PS	Lower Case = RNA bases; (base) = 2'-O-Me; Upper Case=DNA bases	2c
C612	GCTGAGCACTGCACGCCGTAGGTGAAGTGGTCACGAGGG GAACTTCAAGGGTCAAGTGGCTTACGGGGTGCAGTCTTCAGGGTACCAGACATGAAGCTAG (*) = 3PS	V=C/Y, H=3DMT dC CPG	3c, 3d
C613	GCTTGAAGCACTGCACGCCGTAGGTGAAGTGGTCACGAGGG	(*) = 3PS	Extended Data Fig. 1

**[0417] Statistical analysis**

**[0418]** Statistical significance was determined using a Student's t-test with two-tailed distribution. P-values <0.05 were considered as significant. Data are shown as mean and SEM.

**[0419] Results**

**[0420]** CRISPR-Cas nuclease activity and gene editing in *A. thaliana* protoplasts.

**[0421]** Engineered nucleases such as TAL effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease Cas9 (CRISPR-Cas9) can be programmed to target and cleave double-stranded DNA with high specificity. TALENs consist of two arms, both having a TAL effector-like DNA binding domain (TALE) linked to a catalytic DNA nuclease domain of FokI. The TALE domains guide the TALEN arms to specific sites of DNA allowing for dimerization of the FokI endonucleases and subsequent generation of a double strand break (DSB). The CRISPR-Cas9 system consists of a two components; a *Streptococcus pyogenes* Cas9 nuclease (SpCas9) and a chimeric fusion of two RNAs (crRNA and tracrRNA) referred to as an engineered single guide (sgRNA). The sgRNA supports targeted nucleic acid specificity for Cas9 through base pairing of its first twenty 5' bases with the DNA target, subsequently resulting in a site-specific DSB.

**[0422]** In plants, DSBs are typically repaired by the error prone non-homologous end joining (NHEJ) DNA repair pathway resulting in random deletions and/or insertions (indels) at the site of repair. In contrast precision genome editing, relies on nuclease induced DSBs near the targeted change to be repaired by homology directed repair (HDR), a repair pathway that is more precise than NHEJ due to the requirement of a homologous template DNA—in most cases sister chromatid. By harnessing the HDR pathway, it is possible to use an engineered oligonucleotide as the repair template to edit DNA specifically, when cleaved by nucleases or non-specifically when used in combination with double strand break inducing antibiotics.

**[0423]** Fig. 16 depicts CRISPR-Cas9 nuclease activity in *Arabidopsis thaliana* protoplasts derived from the BFP transgenic model system in which a stably integrated *BFP* gene can be converted to *GFP* by editing the codon encoding H66 (CAC→TAC H66Y). When cells were treated with CRISPR-Cas9 (BC-1), NHEJ induced indels were produced at a frequency of 0.79% near the H66 locus of the *BFP* gene by deep

sequencing (Fig. 16a). The majority of indels were single bp and none longer than 9 bp. Conversely, cells treated with GRON only or mock-treated did not exhibit indels (data not shown). These results show that CRISPR-Cas9 nuclease can actively target the *BFP* gene in this transgenic model system.

**[0424]** With regard to the effectiveness of CRISPR-Cas9 in combination with GRON to mediate *BFP* to *GFP* gene editing in protoplasts derived from our transgenic model system, a 7.4-fold improvement in *BFP* to *GFP* editing was observed when both CRISPR-Cas9 and phosphorothioate (PS) modified GRONs (CG6) are introduced concurrently when compared to GRON alone or CRISPR-Cas9 alone treatments (Fig. 16b). These results demonstrate that introducing CRISPR-Cas9 with PS modified GRONs into *Arabidopsis* protoplasts significantly increase the frequency of *BFP* to *GFP* gene editing.

**[0425]** GRONs containing three adjacent PS modifications (herein refer to as 3PS) at both the 5' and 3' ends positively effect *BFP* to *GFP* editing when compared to an unmodified GRON. The 3PS modified GRON (CG2), when combined with CRISPR-Cas9 (BC-1), is more efficacious at *BFP* to *GFP* editing when compared to an unmodified GRON template (CG1; Fig. 17a). In addition, a positive correlation between editing and GRON length (Fig. 17b) was observed. Taken together, these results show that both GRON modification and length can greatly improve the frequency of gene editing in plants such as *Arabidopsis* in the presence of CRISPR-Cas9.

**[0426]** When either the 201 nucleobase (nb) 3PS modified GRON (CG6), or the 201 nb 2'-O-methyl modified GRONs (CG9) or (CG10), consisting of the first ten 5' bases as RNA with either the first RNA base or the first 9 RNA bases modified with 2'-O-methyl are introduced along with CRISPR-Cas9 (BC-1) into *Arabidopsis* protoplasts, no statistical difference in *BFP* to *GFP* editing between them was observed (Fig 17c). Similarly, when either the 201 nb 3PS modified GRON (CG3) or the 201 nb Cy3 modified GRON (CG4), comprising of a 5' Cy3 and an 3' idC reverse base, were introduced along with CRISPR-Cas9 (BC-3) into *Arabidopsis* protoplasts, no statistical difference in editing frequencies was observed (Fig 17d). Overall, these data show that diverse GRON modifications can greatly improve the frequency of gene editing in *Arabidopsis* in the presence of CRISPR-Cas9.

[0427] Based on these results with CRISPR-Cas9 and modified GRONs, it was determined if modified GRONs coupled with TALEN pairs targeting the *BFP* gene result in improved *BFP* to *GFP* gene editing as well. To first show effective nuclease activity at the *BFP* locus *Arabidopsis* protoplasts were treated with TALENs (BT-1) and found 0.51 % indels at or near the expected cleavage site by deep sequencing—indicating that TALENs are as active as CRISPR-Cas9 in this model system (Fig 18a). The majority of deletions were >10 bp but less than 50 bp, while insertions, significantly less abundant than deletions were three bp or less. Next we examined the effectiveness of TALENs coupled with modified GRON to edit *BFP* to *GFP*. A 9.2-fold improvement in the frequency of *BFP* to *GFP* editing was observed when both the BFP TALEN (BT-1) and 3PS GRON (CG7) are introduced when compared to 3PS GRON alone (Fig. 18b). Similar to the CRISPR-Cas experiments described above, these results demonstrate that introducing TALENs with 3PS modified GRONs into *Arabidopsis* protoplasts also significantly increases the frequency of *BFP* to *GFP* gene editing.

[0428] The *EPSPS* (5'-enolpyruvylshikimate-3-phosphate synthase) loci in *Linum usitatissimum* (Common flax) was also used as a target in this system. The *EPSPS* genes encode an enzyme in the shikimate pathway that participates in the biosynthesis of the aromatic amino acids phenylalanine, tyrosine and tryptophan. In plants, *EPSPS* is a target for the herbicide, glyphosate, where it acts as a competitive inhibitor of the binding site for phosphoenolpyruvate.

[0429] TALENs targeting a site near two loci (T97I and P101A) in *L. usitatissimum* that when edited will render *EPSPS* tolerant to glyphosate (Extended Data Fig. 18b) were selected. Delivering TALEN (LuET-1) together with a 144 nb 5' Cy3 modified GRON (CG11) containing the targeted changes at T97I and P101A into protoplasts, gene editing frequencies of 0.19% at both loci and indel frequency at 0.5% seven days after introduction were observed (Fig. 18c, 18d). The majority of indels were 10 bp or less (Fig. 18c). These results demonstrate that introducing TALENs with Cy3 modified GRONs into *L. usitatissimum* protoplasts significantly increase the frequency of *EPSPS* gene editing and further that multiple nucleotide edits can be realized with a single GRON.

[0430] **Example 20: Effect of Two Members of the Bleomycin Family of Antibiotics on Conversion.**

[0431] The purpose of this series of examples was to evaluate the effect of antibiotics on conversion efficiencies.

**[0432] Methods**

[0433] Protoplasts from an *Arabidopsis thaliana* line with multiple copies of a blue fluorescent protein gene were treated with GRON as described in Example 1, with the following modification: before the addition of GRON, the protoplasts were kept for 90 minutes on ice in a solution of TM (14.8 mM MgCl<sub>2</sub> x 6H<sub>2</sub>O, 5 mM 2-(N-morpholino)ethanesulfonic acid, 572 mM mannitol), supplemented with 0, 250, or 1000 µg/ml Zeocin™ or phleomycin. The pH of the solutions was adjusted to 7.0. The percentage of green-fluorescing cells resulting from BFP to GFP conversion was evaluated by flow cytometry as described in Example 1.

**[0434] Results**

[0435] Zeocin and phleomycin at both concentrations used (250 and 1000 µg/ml) resulted in an increase in BFP to GFP gene editing (see Figure 19). Green-fluorescing cells resulting from BFP to GFP gene editing were observed five days after GRON delivery (Figure 20).

**[0436] References**

1. LeCong *et al* 2013 Science : vol. 339 no. 6121 pp. 819-823.
2. Jinek *et al* 2012 Science. 337:816-21
3. Wang *et al* 2008 RNA 14: 903-913
4. Zhang *et al* 2013. Plant Physiol. 161: 20-27

**[0437] Example 21: CRISPRs and GRONs in Rice.**

[0438] The purpose of this experiment is to demonstrate ACCase conversion in *Oryza sativa* at 120 hours after PEG delivery of CRISPR-Cas plasmids and GRONs into protoplasts. The CRISPR-Cas used in this experiment targets the *accase* gene in the rice genome by introducing into protoplasts plasmid(s) encoding the Cas9 gene and a sgRNAs. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA guides the Cas9 to the target genes, where Cas9 creates a double-stranded break in the *accase* gene and the GRON is used as a template to convert the *accase* gene in a site-directed manner.

**[0439] Methods**

**[0440]** Rice protoplasts were isolated from calli. The CRISPR-Cas encoded plasmids contains the corn ubiquitin promoter driving the Cas9 coding sequence with an *rbcSE9* terminator and a rice U6 promoter promoter driving the sgRNA with a poly-T<sub>10</sub> terminator. The CRISPR-Cas plasmids were introduced into protoplasts by PEG mediated delivery at a final concentration of 0.05 µg/µl. GRONs with the following sequence, 5' V C TGA CCT GAA CTT GAT CTC AAT TAA CCC TTG CGG TTC CAG AAC ATT GCC TTT TGC AGT CCT CTC AGC ATA GCA CTC AAT GCG GTC TGG GTT TAT CTT GCT TCC AAC GAC AAC CCA AGC CCC TCC TCG TAG CTC TGC AGC CAT GGG AAT GTA GAC AAA GGC AGG CTG ATT GTA TGT CCT AAG GTT CTC AAC AAT AGT CGA GCC H 3', were used at a final concentration of 0.8 µM. Protoplasts were embedded in agarose (2.5 x 10<sup>6</sup> cells/ml), cultured in liquid medium, and incubated in a rotatory shaker (60 rpm) in the dark at 28°C. Individual samples were analyzed by NGS, 120 hours after CRISPR-Cas plasmid and/or GRON delivery, to determine the percentage of cells (DNA reads) carrying the ACCase conversion and having indels in the *accase* gene.

**[0441]** The CRISPR consists of two components: the plant codon-optimized *Streptococcus pyogenes* Cas9 (SpCas9) and sgRNA both of which were expressed from the same plasmid. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA region contains the spacer sequence (5'-ACGAGGAGGGGCTTGGGTTGTGG-3) used to guide the Cas9 nuclease to the target gene. In this experiment the CRISPR targets the *accase* gene.

**[0442] Results**

**[0443]** At 120 h, rice protoplasts have 0.026% ACCase conversion as determined by Next Generation Sequencing. GRON only controls with no CRISPR-Cas showed minimal Accase conversion of 0.002% at 120 hours and the untreated controls showed no conversion. Additionally, these data show that the CRISPR-Cas is active and able to cleave the *ACCcase* target gene and form indels of 8.0%.

**[0444] References**

5. LeCong *et al* 2013 Science : vol. 339 no. 6121 pp. 819-823.
6. Jinek *et al* 2012 Science. 337:816-21
7. Wang *et al* 2008 RNA 14: 903-913

8. Zhang et al 2013. *Plant Physiol.* 161: 20-27

**[0445] Example 22: CRISPRs and GRONs in Rice**

**[0446] Summary:** Targeted *ACCase* mutations have been identified in nineteen week-old calli by PCR and DNA sequencing analyses.

**[0447] Introduction**

**[0448]** The purpose of this experiment is to demonstrate *ACCase* conversion in *Oryza sativa* calli after PEG delivery of CRISPR-Cas plasmids and GRONs into protoplasts. The CRISPR-Cas used in this experiment targets the *ACCase* gene in the rice genome by introducing into protoplasts plasmid(s) encoding the *Cas9* gene and a sgRNAs. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA guides the Cas9 to the target gene, where Cas9 creates a either a double-stranded break or a nick at a targeted location in the *ACCase* gene and the GRONs are used as a template to convert the *ACCase* gene in a site-directed manner.

**[0449] Results**

**[0450]** Targeted *OsACCase* mutations described in the tables below, have been identified in nineteen week-old calli by PCR and DNA sequencing analyses.

**[0451] Methods**

**[0452]** Rice protoplasts were isolated from suspension cultures initiated from mature seed-derived embryogenic calli. The CRISPR-Cas plasmids with GRONs at a final concentration of 0.05  $\mu\text{g}/\mu\text{l}$  and 0.8  $\mu\text{M}$ , respectively, were introduced into protoplasts by PEG mediated delivery method. An exemplary range for CRISPR-Cas plasmids at a final concentration include, but is not limited to 0.01 to 0.2  $\mu\text{g}/\mu\text{l}$ . An exemplary range for GRON at a final concentration include, but is not limited to 0.01 to 4  $\mu\text{M}$ . The CRISPR-Cas encoded plasmids contains the corn ubiquitin promoter driving the *Cas9* coding sequence with an *rbcSE9* terminator and a rice U6 promoter driving the sgRNA with a poly-T<sub>10</sub> terminator. Sequence information of the GRONs are described in Table 3. Following the PEG treatment, protoplasts were embedded in agarose (1.25 x 10<sup>6</sup> cells/ml), cultured in liquid medium, and incubated in a rotatory shaker (60 rpm) in the dark at 28°C. An exemplary range for embedding protoplasts in agarose include, but is not limited to 0.625 x 10<sup>6</sup> to 2.5 x 10<sup>6</sup> cells/ml. Samples from each treatment were analyzed by Next Generation Sequencing after 4 weeks post CRISPR-Cas plasmid and/or

GRON treatment to determine the proportion of cells (DNA reads) carrying the ACCase conversion. Microcalli from converted treatments were released onto solid selection medium containing clethodim (0.25-0.5  $\mu\text{M}$ ) or sethoxydim (2.5  $\mu\text{M}$ ). Individual callus lines growing on this selection medium after 19 weeks of culture were analyzed by in-house screening methods as well as DNA sequencing in order to identify individual calli containing the targeted ACCase conversions.

**[0453]** The CRISPR consists of two components: the Cas9, such as a plant codon-optimized *Streptococcus pyogenes* Cas9 (SpCas9) and sgRNA both of which were expressed from plasmid(s). Cas9 and sgRNA can also be delivered by mRNA/RNA or protein/RNA respectively. The sgRNA is a fusion of CRISPR RNA (crRNA) and transactivating crRNA (tracrRNA). The crRNA region contains the spacer with sequences described in Table 2, which were used to guide the Cas9 nuclease to the target gene. In this experiment the CRISPR targets the rice *ACCase* gene.

**[0454]** List of conversions in OsACCase at two different locations within the gene, Site 1 and Site 2. For each site, all combinations of conversion events are possible.

OsACCase	
Site1	Site 2
I1781A	D2078G
I1781L	D2078K
I1781M	D2078T
I1781N	S2079F
I1781S	K2080E
I1781T	C2088F
I1781V	C2088G
G1783C	C2088H
A1786P	C2088K
	C2088L
	C2088N
	C2088P
	C2088Q
	C2088R
	C2088S
	C2088T
	C2088V
	C2088W

[0455] List of CRISPR-Cas gRNA spacer sequences used in this experiment. Spacer length may vary up to  $\pm 20$  bp. Mismatched within the spacer sequence can be tolerated up to 10 bp.

Sample ID	Spacer RNA Sequence (5' to 3')	OsACCase	
		Site 1	Site 2
1	CAUAAGAUGCAGCUAGACAG	X	
2	AGCUAGACAGUGGUGAAAUU	X	
3	UAGACAGUGGUGAAAUUAGG	X	
4	AGACAGUGGUGAAAUUAGGU	X	
5	GUGGGUUAUUGAUUCUGUUG	X	
6	UGGGUUAUUGAUUCUGUUGU	X	
7	UAUUGAUUCUGUUGUGGGCA	X	
8	UCUGUUGUGGGCAAGGAAGA	X	
9	GUGGGCAAGGAAGAUGGACU	X	
10	CAAGGAAGAUGGACUUGGUG	X	
11	CUUGGUGUGGAGAAUAUACA	X	
12	CUAUUGCCAGUGCUUAUUCU	X	
13	UGC UUAUUCUAGGGCAUAUA	X	
14	UUUACACUUAACAUUUGUGAC	X	
15	UUUGUGACUGGAAGAACUGU	X	
16	ACUGGAAGAACUGUUGGAAU	X	
17	GGAGCUUAUCUUGCUCGACU	X	
18	AUAUGCCCUAGAAUAAGCAC	X	
19	GAUAAGAUGCAGCUAGACAG	X	
20	GGCUAGACAGUGGUGAAAUU	X	
21	GAGACAGUGGUGAAAUUAGG	X	
22	GGACAGUGGUGAAAUUAGGU	X	
24	GGGGUUAUUGAUUCUGUUGU	X	
25	GAUUGAUUCUGUUGUGGGCA	X	
26	GCUGUUGUGGGCAAGGAAGA	X	
28	GAAGGAAGAUGGACUUGGUG	X	
29	GUUGGUGUGGAGAAUAUACA	X	
30	GUAUUGCCAGUGCUUAUUCU	X	
31	GGCUUAUUCUAGGGCAUAUA	X	
32	GUUACACUUAACAUUUGUGAC	X	
33	GUUGUGACUGGAAGAACUGU	X	
34	GCUGGAAGAACUGUUGGAAU	X	
36	GUAUGCCCUAGAAUAAGCAC	X	
37	CGACUAUUGUUGAGAACCUU		X
38	UGCCUUGUCUACAUCCCA		X
39	CCAUGGCUGCAGAGCUACG		X

Sample ID	Spacer RNA Sequence (5' to 3')	OsACCase	
		Site 1	Site 2
40	AUGGCUGCAGAGCUACGAGG		X
41	UGGCUGCAGAGCUACGAGGA		X
42	GGCUGCAGAGCUACGAGGAG		X
43	CAGAGCUACGAGGAGGGGCU		X
44	AGAGCUACGAGGAGGGGCUU		X
45	ACGAGGAGGGGCUUGGGUUG		X
46	GCAUUGAGUGCUAUGCUGAG		X
47	UAUGCUGAGAGGACUGCAA		X
48	GACUGCAAAGGCAAUGUUC		X
49	GGCAAUGUUCUGGAACCGCA		X
50	GCAAUGUUCUGGAACCGCAA		X
51	GGUAAUUGAGAUCAAGUUC		X
52	UGAGAUCAAGUUCAGGUCAG		X
53	GUUCAGGUCAGAGGAACUCC		X
54	AACUCCAGGAUUGCAUGAGU		X
55	CAAGCCGACUCAUGCAAUCC		X
56	UUGAUCUCAAUUAACCCUUG		X
57	UCUCAGCAUAGCACUCA AUG		X
58	GCAUAGCACUCA AUGCGGUC		X
59	CAUAGCACUCA AUGCGGUCU		X
60	UCCUCGUAGCUCUGCAGCCA		X
61	AGCCAUGGGAAUGUAGACAA		X
62	AUGGGAAUGUAGACAAAGGC		X
63	CAGGCUGAUUGUAUGUCCUA		X
64	GGACUAUUGUUGAGAACCUU		X
65	GGCCUUUGUCUACAUUCCCA		X
66	GCCAUGGCUGCAGAGCUACG		X
67	GUGGCUGCAGAGCUACGAGG		X
68	GGGCUGCAGAGCUACGAGGA		X
69	GAGAGCUACGAGGAGGGGCU		X
70	GGAGCUACGAGGAGGGGCUU		X
71	GCGAGGAGGGGCUUGGGUUG		X
72	GAUGCUGAGAGGACUGCAA		X
73	GGAGAUCAAGUUCAGGUCAG		X
74	GACUCCAGGAUUGCAUGAGU		X
75	GAAGCCGACUCAUGCAAUCC		X
76	GUGAUCUCAAUUAACCCUUG		X
77	GCUCAGCAUAGCACUCA AUG		X
78	GAUAGCACUCA AUGCGGUCU		X
79	GCCUCGUAGCUCUGCAGCCA		X

Sample ID	Spacer RNA Sequence (5' to 3')	OsACCase	
		Site 1	Site 2
80	GGCCAUGGGAAUGUAGACAA		X
81	GUGGGAAUGUAGACAAAGGC		X
82	GAGGCUGAUUGUAUGUCCUA		X

[0456] A list of GRON sequences suitable for use in this experiment are provided in the table below (V=CY3; H=3'DMT dC CPG).

1	VGTCATAGCACATAAGATGCAGCTAGACAGTGGTCAAATTAGGTGGGTTAT TGATTCTGTTGTGGGCAAGGAAGATGGACTTGGTGTGGAGAATGCTCATGG AAGTGCTGCTATTGCCAGTGCTTATTCTAGGGCATATAAGGAGACATTTAC ACTTACATTTGTGACTGGAAGAACTGTTGGAATAGGAGCTTATCTTGCTCH
2	VGAGCAAGATAAGCTCCTATTCCAACAGTTCTTCCAGTCACAAATGTAAGT GTAAATGTCTCCTTATATGCCCTAGAATAAGCACTGGCAATAGCAGCACTT CCATGCAGATTCTCCACACCAAGTCCATCTTCCCTTGCCCACAACAGAATCA ATAACCCACCTAATTTCAACCACTGTCTAGCTGCATCTTATGTGCTATGACH
3	VGTCATAGCACATAAGATGCAGCTAGACAGTGGTCAAATTAGGTGGGTTATGATT CTGTTGTGGGCAAGGAAGATGGACTTGGTGTGGAGAATATACATTGCAGTGCTGCT ATTGCCAGTGCTTATTCTAGGGCATATAAGGAGACATTTACTTACATTTGTGAC TGGAAGAACTGTTGGAATAGGAGCTTATCTTGCTCH
4	VGAGCAAGATAAGCTCCTATTCCAACAGTTCTTCCAGTCACAAATGTAAGTGTAAA TGTCTCCTTATATGCCCTAGAATAAGCACTGGCAATAGCAGCACTGCAATGTATAT TCTCCACACCAAGTCCATCTTCCCTTGCCCACAACAGAATCAATAACCCACCTAATT TCACCACTGTCTAGCTGCATCTTATGTGCTATGACH
5	VGTCATAGCACATAAGATGCAGCTAGACAGTGGTCAAATTAGGTGGGTTAT TGATTCTGTTGTGGGCAAGGAAGATGGACTTGGTGTGGAGAATATACATGG AAGTGCTCCAATTGCCAGTGCTTATTCTAGGGCATATAAGGAGACATTTAC ACTTACATTTGTGACTGGAAGAACTGTTGGAATAGGAGCTTATCTTGCTCH
6	VGAGCAAGATAAGCTCCTATTCCAACAGTTCTTCCAGTCACAAATGTAAGT GTAAATGTCTCCTTATATGCCCTAGAATAAGCACTGGCAATGGTAGCACTT CCATGTATATTCTCCACACCAAGTCCATCTTCCCTTGCCCACAACAGAATCAA TAACCCACCTAATTTCAACCACTGTCTAGCTGCATCTTATGTGCTATGACH
7	VCTGACCTGAACTTGATCTCAATTAACCCCTTGCGGTTCCAGAACATTGCCTT TTGCAGTCTCTCAGCATAGCACTCAATGCGGTCTGGGTTTATCTTGCTTCC AACGACAACCCAAGCCCCTCCTCGTAGCTCTGCAGCCATGGGAATGTAGAC AAAGGCAGGCTGATTGTATGTCCTAAGGTTCTCAACAATAGTCGAGCCH
8	VGGCTCGACTATTGTTGAGAACCCTTAGGACATACAATCAGCCTGCCTTTGTC TACATTCCCATGGCTGCAGAGCTACGAGGAGGGGCTTGGGTTGTGGTTGGT AGCAAGATAAACCCAGACCGCATTGAGTGCTATGCTGAGAGGACTGCAAA AGGCAATGTTCTGGAACCGCAAGGGTTAATTGAGATCAAGTTCAGGTCAGH
9	VCTGACCTGAACTTGATCTCAATTAACCCCTTGCGGTTCCAGAACATTGCCTT TTGCAGTCTCTCAGCATAGCACTCAATGCGGTCTGGGTTTATCTTAAATC AACGACAACCCAAGCCCCTCCTCGTAGCTCTGCAGCCATGGGAATGTAGAC AAAGGCAGGCTGATTGTATGTCCTAAGGTTCTCAACAATAGTCGAGCCH
10	VGGCTCGACTATTGTTGAGAACCCTTAGGACATACAATCAGCCTGCCTTTGTC TACATTCCCATGGCTGCAGAGCTACGAGGAGGGGCTTGGGTTGTGGTTGAT AGCAAGATAAACCCAGACCGCATTGAGAGGTATGCTGAGAGGACTGCAAA

	AGGCAATGTTCTGGAACCGCAAGGGTTAATTGAGATCAAGTTCAGGTCAGH
11	VGGCTCGACTATTGTTGAGAACCTTAGGACATACAATCAGCCTGCCTTTGTC TACATTCCCATGGCTGCAGAGCTACGAGGAGGGGCTTGGGTTGTGGTTGAT AGCGAAATAAACCCAGACCGCATTGAGTGCTATGCTGAGAGGACTGCAA AGGCAATGTTCTGGAACCGCAAGGGTTAATTGAGATCAAGTTCAGGTCAGH
12	VCTGACCTGAACTTGATCTCAATTAACCCTTGCGGTTCCAGAACATTGCCTT TTGCAGTCCTCTCAGCATAGCACTCAATGCGGTCTGGGTTTATTTTCGCTATC AACCACAACCCAAGCGCCTCCTCGTAGCTCTGCAGCCATGGGAATGTAGAC AAAGGCAGGCTGATTGTATGTCCTAAGGTTCTCAACAATAGTCGAGCCH
13	VGGCTCGACTATTGTTGAGAACCTTAGGACATACAATCAGCCTGCCTTTGTC TACATTCCCATGGCTGCAGAGCTACGAGGAGGGGCTTGGGTTGTGGTTGAT AGCAAGATAAACCCAGACCGCATTGAGCGTTATGCTGAGAGGACTGCAA AGGCAATGTTCTGGAACCGCAAGGGTTAATTGAGATCAAGTTCAGGTCAGH
14	VCTGACCTGAACTTGATCTCAATTAACCCTTGCGGTTCCAGAACATTGCCTT TTGCAGTCCTCTCAGCATATTGCTCAATGCGGTCTGGGTTTATCTTGCTATC AACGACAACCCAAGCCCCTCCTCGTAGCTCTGCAGCCATGGGAATGTAGAC AAAGGCAGGCTGATTGTATGTCCTAAGGTTCTCAACAATAGTCGAGCCH

**[0457] Example 23: CRISPRs and GRONs in Flax**

**[0458] Summary:** Targeted *LuEPSPS* mutations have been identified in four week-old calli by PCR and DNA sequencing analyses.

**[0459] Introduction**

**[0460]** The purpose of this experiment is to demonstrate conversion of the *EPSPS* genes in the *Linum usitatissimum* genome in shoot tip derived protoplasts by PEG mediated delivery of CRISPR-Cas plasmids and GRONs. The CRISPR-Cas and GRONs used in this experiment target the *EPSPS* genes in the flax genome. The CRISPR consists of two components: a Cas9, such as a plant codon-optimized *Streptococcus pyogenes* Cas9 (SpCas9) and sgRNA both of which are expressed from plasmid(s). Cas9 and sgRNA can also be delivered by mRNA/RNA or protein/RNA respectively. The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA guides the Cas9 to the targeted genes, where Cas9 creates a either a double-stranded break or nick in the *EPSPS* genes and the GRONs are used as a template to convert the *EPSPS* genes in a site-directed manner.

**[0461] Results**

**[0462]** Targeted *LuEPSPS* (T97I and/or the P101A, P101T or P101S and/or the G96A) mutations have been identified in four week-old calli by PCR and DNA sequencing analyses. Shoots have been regenerated from these converted calli.

**[0463] Methods**

**[0464]** Flax protoplasts were isolated from shoot tips obtained from *in vitro* germinated seedlings. The CRISPR-Cas encoded plasmids contain the MAS promoter driving the Cas9 coding sequence with an *rbcSE9* terminator and the *Arabidopsis thaliana* U6 promoter driving the sgRNA with a poly-T<sub>10</sub> terminator. The CRISPR-Cas plasmids were introduced into protoplasts by PEG mediated delivery at a final concentration of 0.05 µg/µl. GRONs targeting each of the two flax *LuEPSPS* genes (Table 2) were used at a final concentration of 4.0 µM. An exemplary range for CRISPR-Cas plasmids at a final concentration include, but is not limited to 0.01 to 0.2 µg/µl. An exemplary range for GRON at a final concentration include, but is not limited to 0.01 to 4 µM. Protoplasts were embedded in alginate beads (5 x 10<sup>5</sup> cells/ml), cultured in liquid medium, and incubated in a rotatory shaker (30 rpm) in the dark at 25°C. An exemplary range for embedding protoplasts in alginate beads include, but is not limited to 3.0 X 10<sup>5</sup> to 7.5 X 10<sup>5</sup> cells/ml. Microcalli developed from individual cells were analyzed by Next Generation Sequencing, 3 and 7 weeks after CRISPR-Cas plasmid and GRON delivery, to determine the proportion of cells (DNA reads) carrying the targeted mutations in the *LuEPSPS* genes. Larger calli were grown from 8-week-old converted microcalli plated over solid regeneration medium, and shoots started differentiating from regenerated calli after about 4-8 weeks. Converted calli and shoots with the targeted *EPSPS* gene mutations were identified by PCR and DNA sequencing analyses.

**[0465]** The CRISPR consists of two components: the plant codon-optimized *Streptococcus pyogenes* Cas9 (SpCas9) and sgRNA both of which were expressed from plasmid(s). The sgRNA is a fusion of CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). The crRNA region contains the spacer with sequences described in the table below, which were used to guide the Cas9 nuclease to the *EPSPS* targeted genes.

**[0466]** List of CRISPR-Cas gRNA spacer sequences used in this experiment. Spacer length may vary up to ±20 bp. Mismatched within the spacer sequence can be tolerated up to 10 bp.

Sample ID	Spacer RNA Sequence (5' to 3')	LuEPSPS Genes	
		1	2
1	CAGAAGCGCGCCAUUGUUGA	X	X
2	CGCGCCAUUGUUGAAGGUUG	X	

3	CGCGCCAUUGUUGAAGGUCG		X
4	GCCAUUGUUGAAGGUUGUGG	X	
5	GCCAUUGUUGAAGGUCGUGG		X
6	AGGUUGUGGUGGUGUGUUUC	X	
7	AGGUCGUGGUGGUGUGUUUC		X
8	UGUGGUGGUGUGUUUCCGGU	X	
9	CGUGGUGGUGUGUUUCCGGU		X
10	UGUGUUUCCGGUCGGUAAAC	X	X
11	UGUUUCCGGUCGGUAAACUG		X
12	AACGAUAUUGAACUUUCCU	X	
13	AACGAUAUCGAACUUUCCU		X
14	GAACUUUCCUUGGAAAUGC	X	X
15	ACAGCUGCUGAACAGCCGC	X	X
16	GCUGCUGAACAGCCGCUGG	X	X
17	AACUCAAGCUACAUACUCGA	X	
18	AACUCAAGCUACAUACUCGA		X
19	CGAAUGAGAGAGACCAAU	X	
20	CGAAUGAGAGAGACCGAU		X
21	AGAGAGACCAAUUGGAGAUU	X	
22	CCAAUUGGAGAUUUGGUUGU	X	
23	CCGAUUGGAGAUUAGUUGU		X
24	CCAACAACCAAUCUCCAAU	X	
25	CCAACAACUAAAUCUCCAAU		X
26	AUUGGUCUCUCUCAUUCG	X	
27	AUCGGUCUCUCUCAUUCG		X
28	GUAGCUUGAGUUGCCUCCAG	X	X
29	GCUGUACAGCAGCUGUCAG	X	X
30	UAGCUGUCCAGCAUUUCCA	X	X
31	UUCUUCGCCAGUUUACCGAC	X	
32	UUCUUCCCCAGUUUACCGAC		X
33	ACCACCACAACCUUCAACAA	X	
34	ACCACCACGACCUUCAACAA		X
35	GAGAAGCGCGCCAUUGUUGA	X	X
36	GGCGCCAUUGUUGAAGGUUG	X	
37	GGCGCCAUUGUUGAAGGUCG		X
38	GGGUUGUGGUGGUGUGUUUC	X	
39	GGGUCGUGGUGGUGUGUUUC		X
40	GGUGGUGGUGUGUUUCCGGU	X	
41	GGUGGUGGUGUGUUUCCGGU		X
42	GGUGUUUCCGGUCGGUAAAC	X	X
43	GGUUUCCGGUCGGUAAACUG		X
44	GACGAUAUUGAACUUUCCU	X	

45	GACGAUAUCGAACUUUCCU		X
46	GCAGCUGCUGU AACAGCCGC	X	X
47	GACUCAAGCUACAUACUCGA	X	
48	GACUCAAGCUACAUACUCGA		X
49	GGAAUGAGAGAGACCAAU	X	
50	GGAAUGAGAGAGACCGAU		X
51	GGAGAGACCAAUUGGAGAUU	X	
52	GCAAUUGGAGAUUUGGUUGU	X	
53	GCGAUUGGAGAUUAGUUGU		X
54	GCAACAACCAAUCUCCAAU	X	
55	GCAACAACUAAAUCUCCAAU		X
56	GUUGGUCUCUCUCAUUCG	X	
57	GUCGGUCUCUCUCAUUCG		X
58	GAGCUGUCCAGCAUUCCA	X	X
59	GUCUUCGCCAGUUUACCGAC	X	
60	GUCUUCCCCAGUUUACCGAC		X
61	GCCACCACAACCUUCAACAA	X	
62	GCCACCACGACCUUCAACAA		X

[0467] A list of GRON sequences suitable for use in this experiment are provided in the table below (V=CY3; H=3'DMT dC CPG).

1	VCGGTCGGTAAACTGGCGAAGAACGATATTGAACTTTTCCTTGGAAATGCTGCTATAGCTATGCGTGCGCTGACAGCTGCTGTAACAGCCGCTGGAGGCAACTCAAGGTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
2	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGGACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGCGGACGCATAGCTGTGGCAGCATT TCCAAGGAAAAGTTCAATATCGTTCTTCGCCAGTTTACCGACCGH
3	VCGGTCGGTAAACTGGCGAAGAACGATATTGAACTTTTCCTTGGAAATGCTGGAATAGCTATGCGTGCGCTGACAGCTGCTGTAACAGCCGCTGGAGGCAACTCAAGGTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
4	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGGACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGCGCACGCATAGCTATTCCAGCATT TCCAAGGAAAAGTTCAATATCGTTCTTCGCCAGTTTACCGACCGH
5	VCGGTCGGTAAACTGGCGAAGAACGATATTGAACTTTTCCTTGGAAATGCTGGAATTGCTATGCGTTCTCTGACAGCTGCTGTAACAGCCGCTGGAGGCAACTCAAGGTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
6	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGGACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGTGAACGCATAGCAATTCCAGCATT TCCAAGGAAAAGTTCAATATCGTTCTTCGCCAGTTTACCGACCGH
7	VCGGTCGGTAAACTGGCGAAGAACGATATTGAACTTTTCCTTGGAAATGCTGGAATCGCTATGCGTACTCTGACAGCTGCTGTAACAGCCGCTGGAGGCAACTCAAGGTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
8	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGGACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGTGTACGCATAGCAATTCCAGCATT TCCAAGGAAAAGTTCAATATCGTTCTTCGCCAGTTTACCGACCGH

9	VCGGTCGGTAAACTGGCGAAGAACGATATTGAACTTTTCCTTGGAAATGCT GGAATTGCTATGCGTGCGCTGACAGCTGCTGTAACAGCCGCTGGAGGCAAC TCAAGGTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
10	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGAACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGCGCACGCATAGCTATTCCAGCATT TCCAAGGAAAAGTTCGATATCGTTCTTCCCCAGTTTACCGACCGH
11	VCGGTCGGTAAACTGGCGAAGAACGATATTGAACTTTTCCTTGGAAATGCT GGAATAGCTATGCGTGCTCTGACAGCTGCTGTAACAGCCGCTGGAGGCAAC TCAAGGTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
12	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGGACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGCGCACGCATAGCTGTTCCAGCATT TCCAAGGAAAAGTTCGATATCGTTCTTCCCGAGTTTACCGACCGH
13	VCGGTCGGTAAACTGGGGAAGAACGATATCGAACTTTTCCTTGGAAATGCT GCTATAGCTATGCGTGCGCTGACAGCTGCTGTAACAGCCGCTGGAGGCAAC TCAAGGTTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
14	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGAACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGCGGACGCATAGCTGTTGCAGCATT TCCAAGGAAAAGTTCGATATCGTTCTTCCCCAGTTTACCGACCGH
15	VCGGTCGGTAAACTGGGGAAGAACGATATCGAACTTTTCCTTGGAAATGCT GGAATAGCTATGCGTGCGCTGACAGCTGCTGTAACAGCCGCTGGAGGCAAC TCAAGGTTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
16	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGAACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGCGCACGCATAGCTATTCCAGCATT TCCAAGGAAAAGTTCGATATCGTTCTTCCCCAGTTTACCGACCGH
17	VCGGTCGGTAAACTGGGGAAGAACGATATCGAACTTTTCCTTGGAAATGCT GGAATTGCTATGCGTTCTCTGACAGCTGCTGTAACAGCCGCTGGAGGCAAC TCAAGGTTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
18	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGAACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGTGAACGCATAGCGATTCCAGCATT TCCAAGGAAAAGTTCGATATCGTTCTTCCCCAGTTTACCGACCGH
19	VCGGTCGGTAAACTGGGGAAGAACGATATCGAACTTTTCCTTGGAAATGCT GGAATCGCTATGCGTACTCTGACAGCTGCTGTAACAGCCGCTGGAGGCAAC TCAAGGTTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
20	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGAACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGTGTACGCATAGCTATTCCAGCATT CCAAGGAAAAGTTCGATATCGTTCTTCCCCAGTTTACCGACCGH
21	VCGGTCGGTAAACTGGGGAAGAACGATATCGAACTTTTCCTTGGAAATGCT GGAATCGCTATGCGTGCGCTGACAGCTGCTGTAACAGCCGCTGGAGGCAAC TCAAGGTTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
22	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGAACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGCGGACGCATAGCTATTCCAGCATT TCCAAGGAAAAGTTCGATATCGTTCTTCCCCAGTTTACCGACCGH
23	VCGGTCGGTAAACTGGGGAAGAACGATATCGAACTTTTCCTTGGAAATGCT GGAATAGCTATGCGTGCTCTGACAGCTGCTGTAACAGCCGCTGGAGGCAAC TCAAGGTTCCCTTCCCTCAACTCCTTCCAGCCTTTCAGCTTCTTH
24	VAAGAAGCTGAAAGGCTGGAAGGAGTTGAGGGAAGGAACCTTGAGTTGCC TCCAGCGGCTGTTACAGCAGCTGTCAGGGAACGCATAGCTGTTCCAGCATT TCCAAGGAAAAGTTCGATATCGTTCTTCCCCAGTTTACCGACCGH

[0468] One skilled in the art readily appreciates that the present disclosure is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. The examples provided herein are representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the disclosure.

[0469] It will be readily apparent to a person skilled in the art that varying substitutions and modifications may be made to the disclosure disclosed herein without departing from the scope and spirit of the disclosure.

[0470] The disclosure illustratively described herein suitably may be practiced in the absence of any element or elements, limitation or limitations which is not specifically disclosed herein. Thus, for example, in each instance herein any of the terms “comprising”, “consisting essentially of” and “consisting of” may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and there is no intention that in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the disclosure claimed. Thus, it should be understood that although the present disclosure has been specifically disclosed by preferred embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this disclosure as defined by the appended claims.

[0471] Thus, it should be understood that although the present disclosure has been specifically disclosed by preferred embodiments and optional features, modification, improvement, and variation of the disclosures disclosed may be resorted to by those skilled in the art, and that such modifications, improvements and variations are considered to be within the scope of this disclosure. The materials, methods, and examples provided here are representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the disclosure.

[0472] The disclosure has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the disclosure. This includes the generic description of the disclosure with a

proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

**[0473]** In addition, where features or aspects of the disclosure are described in terms of Markush groups, those skilled in the art will recognize that the disclosure is also thereby described in terms of any individual member or subgroup of members of the Markush group.

**[0474]**

**[0475]** Other embodiments are set forth within the following claims.

**THE EMBODIMENTS OF THE INVENTION FOR WHICH AN EXCLUSIVE PROPERTY OR PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:**

1. A method of causing a genetic change in a plant cell, said method comprising exposing said cell to a clustered regularly interspaced short palindromic repeats (CRISPR) nuclease and a gene repair oligonucleobase (GRON), wherein the GRON includes one or more of the following alterations from conventional RNA and DNA nucleotides:
  - a reverse base at the 3' end thereof;
  - one or more 2'O-methyl nucleotides at the 5' and/or 3' end thereof;
  - a backbone modification selected from the group consisting of one or more 3PS modifications at the 5' and/or 3' ends, wherein 3PS denotes 3 phosphorothioate linkages, one or more O-(2-methoxyethyl) (MOE) modifications at the 5' and/or 3' ends; and
  - one or more fluorescent dyes conjugated thereto at the 5' or 3' end of the GRON, wherein said fluorescent dye is a blocked phosphoroamidite that upon incorporation into an oligonucleotide yields a 3,3,3',3'-tetramethyl N,N'-isopropyl substituted indomonocarbocyanine dye.
2. The method of claim 1, wherein said GRON is single stranded.
3. The method of claim 1 or 2, wherein the GRON has a wobble base pair relative to the target sequence for the genetic change.
4. The method of any one of claims 1 to 3, wherein said GRON is between 15 and 210 nucleotides in length.
5. The method any one of claims 1 to 4, wherein said plant cell is a cell of a plant selected from the group consisting of canola, sunflower, corn, tobacco, sugar beet, cotton, maize, wheat, barley, rice, alfalfa, sorghum, tomato, mango, peach, apple, pear, strawberry, banana, melon, cassava, potato, carrot, lettuce, onion, soy bean, soya spp, sugar cane, pea, chickpea, field pea, fava bean, lentils, turnip, rutabaga, brussel sprouts, lupin, cauliflower, kale, field beans, poplar, pine, eucalyptus, grape, citrus, triticale, alfalfa, rye, oats, turf and forage grasses, flax, oilseed rape, mustard, cucumber, morning glory, balsam, pepper, eggplant, marigold, lotus, cabbage, daisy, carnation, tulip, iris, cassava, and lily.

6. The method of any one of claims 1 to 5, wherein said plant cell is a canola cell, or wherein said plant cell is a corn cell, or wherein said plant cell is a maize cell, or wherein said plant cell is a rice cell, or wherein said plant cell is a sorghum cell, or wherein said plant cell is a potato cell, or wherein said plant is a soy bean cell, or wherein said plant cell is a flax cell, or wherein said plant cell is an oilseed rape cell, or wherein said plant cell is a cassava cell, or wherein said plant cell is a sunflower cell.
7. The method of any one of claims 1 to 6, wherein multiple genetic changes are made.
8. The method of any one of claims 1 to 7, wherein two or more guide RNAs are used.
9. The method of claim 8, wherein each of the two or more guide RNAs is complementary to a different target for genetic change.
10. The method of any one of claims 1 to 9, wherein the CRISPR nuclease includes a nickase.
11. The method of any one of claims 1 to 9, wherein the CRISPR nuclease includes two or more nickases, and/or wherein two or more nickases cuts are on opposite strands of the target nucleic acid sequence, or wherein two or more nickases cuts are on the same strand of the target nucleic acid sequence.
12. A plant cell comprising a clustered regularly interspaced short palindromic repeats (CRISPR) nuclease and a gene repair oligonucleobase (GRON), wherein the GRON includes one or more of the following alterations from conventional RNA and DNA nucleotides:
  - a reverse base at the 3' end thereof;
  - one or more 2'O-methyl nucleotides at the 5' and/or 3' end thereof;
  - a backbone modification selected from the group consisting of one or more 3PS modifications at the 5' and/or 3' ends, wherein 3PS denotes 3 phosphorothioate linkages, one or more O-(2-methoxyethyl) (MOE) modifications at the 5' and/or 3' ends; and
  - one or more fluorescent dyes conjugated thereto at the 5' or 3' end of the GRON, wherein

said fluorescent dye is a blocked phosphoroamidite that upon incorporation into an oligonucleotide yields a 3,3,3',3'-tetramethyl N,N'-isopropyl substituted indomonocarbocyanine dye.

13. The cell of claim 12, wherein said GRON is single stranded.
14. The cell of claim 12 or 13, wherein the GRON has a wobble base pair relative to the target sequence for the genetic change.
15. The cell of any one of claims 12 to 14, wherein said GRON is between 15 and 210 nucleotides in length.
16. The cell of any one of claims 12 to 15, wherein said plant cell is of a plant selected from the group consisting of canola, sunflower, corn, tobacco, sugar beet, cotton, maize, wheat, barley, rice, alfalfa, barley, sorghum, tomato, mango, peach, apple, pear, strawberry, banana, melon, cassava, potato, carrot, lettuce, onion, soy bean, soya spp, sugar cane, pea, chickpea, field pea, fava bean, lentils, turnip, rutabaga, brussel sprouts, lupin, cauliflower, kale, field beans, poplar, pine, eucalyptus, grape, citrus, triticale, alfalfa, rye, oats, turf and forage grasses, flax, oilseed rape, mustard, cucumber, morning glory, balsam, pepper, eggplant, marigold, lotus, cabbage, daisy, carnation, tulip, iris, cassava, and lily.
17. The cell of any one of claims 12 to 16, wherein said plant cell is a canola cell, or wherein said plant cell is a corn cell, or wherein said plant cell is a maize cell, or wherein said plant cell is a rice cell, or wherein said plant cell is a sorghum cell, or wherein said plant cell is a potato cell, or wherein said plant is a soy bean cell, or wherein said plant cell is a flax cell, or wherein said plant cell is an oilseed rape cell, or wherein said plant cell is a cassava cell, or wherein said plant cell is a sunflower cell.
18. The cell of any one of claims 12 to 17, wherein multiple genetic changes are made.
19. The cell of any one of claims 12 to 18, wherein two or more guide RNAs are used.
20. The cell of claim 19, wherein each of the two or more guide RNAs is complementary to a

different target for genetic change.

21. The cell of any one of claims 12 to 20, wherein the CRISPR nuclease includes a nickase.
22. The cell of any one of claims 12 to 20, wherein the CRISPR nuclease includes two or more nickases, and/or wherein two or more nickases cuts are on opposite strands of the target nucleic acid sequence, or wherein two or more nickases cuts are on the same strand of the target nucleic acid sequence.

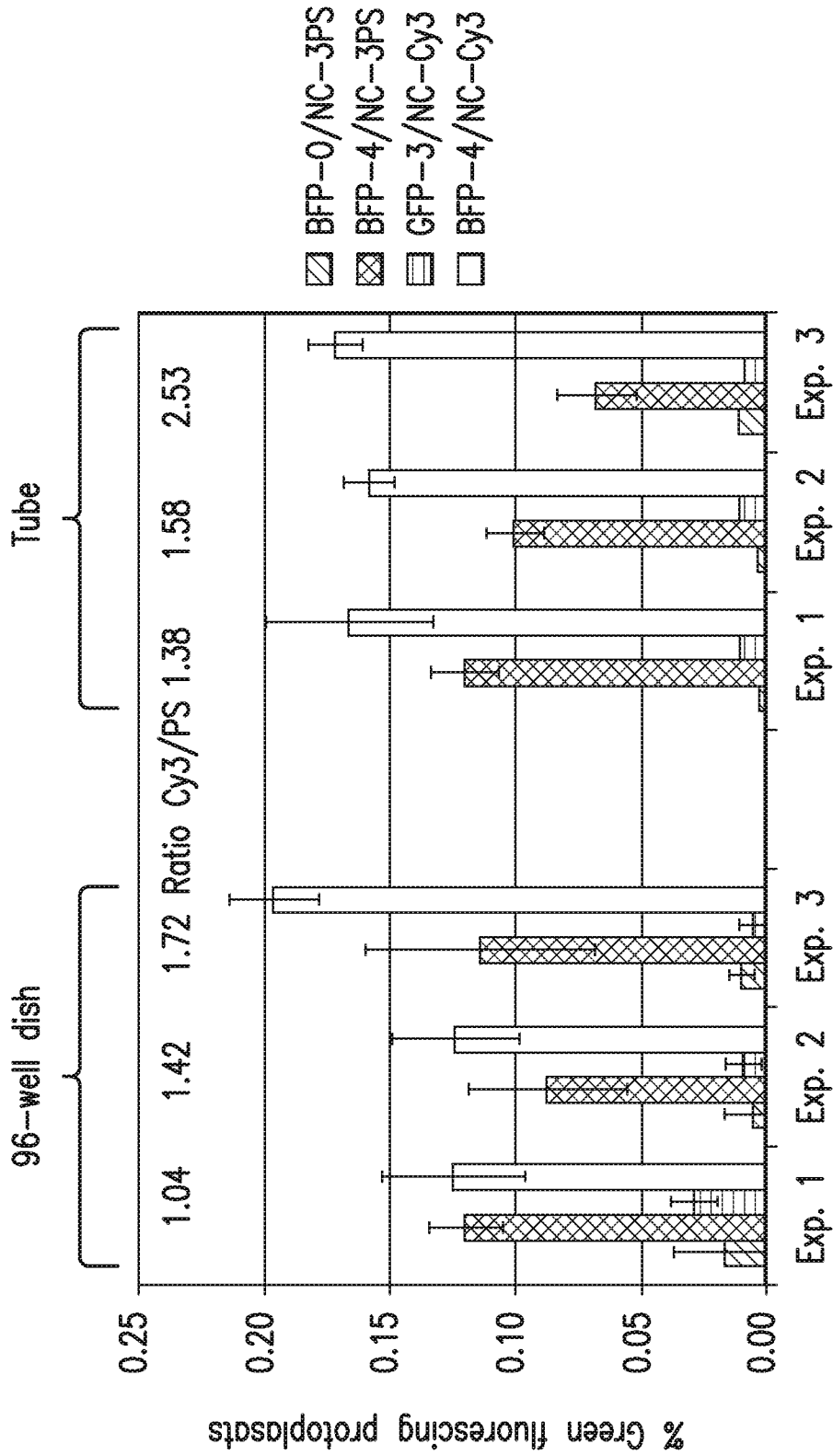


FIG.1

A. Okazaki Fragment GRONs (BFP4 or 0/C or NC/71/5' 2' -0-Me (1))

BFP4/NC: U UCA UGU GG U CGG GGT AGC G GC TGA AGC ACT GCA CGC CGT AGG TGA AGG  
TGG TCA CGA GGG TGG GCC AGG G (71 -mer)

BFP0/NC: U UCA UGU GG U CGG GGT AGC G GC TGA AGC ACT GCA CGC CGT GGG TGA AGG  
TGG TCA CGA GGG TGG GCC AGG G (71 -mer)

BFP4/C: G CUG CCC GUG CCC TGG CCC A CC CTC GTG ACC ACC TTC ACC IAC GGC GTG  
CAG TGC TTC AGC CGC TAC CCC G (71 -mer)

BFP0/C: G CUG CCC GUG CC C TGG CCC A CC CTC GTG ACC ACC TTC ACC CAC GGC GTG  
CAG TGC TTC AGC CGC TAC CCC G (71 -mer)

RNA bases

2' - 0- Me group on first 5' RNA base

(RNA bases without the 2' - 0 -Me groups are in boxes)

B. Okazaki Fragment GRONs ( BFP4 or 0 /C or NC/71/5' 2' -0-Me (9) )

BFP4/NC: U UCA UGU GG U CGG GGT AGC G GC TGA AGC ACT GCA CGC CGT AGG TGA AGG  
TGG TCA CGA GGG TGG GCC AGG G (7 1- mer)

BFP0/NC: U UCA UGU GG U CGG GGT AGC G GC TGA AGC ACT GCA CGC CGT GGG TGA AGG  
TGG TCA CGA GGG TGG GCC AGG G (7 1 -mer)

BFP4/C: G CUG CCC GUG CC C TGG CCC A CC CTC GTG ACC ACC TTC ACC IAC GGC GTG  
CAG TGC TTC AGC CGC TAC CCC G (7 1 -mer)

BFP0/C: G CUG CCC GUG CCC TGG CCC A CC CTC GTG ACC ACC TTC ACC CAC GGC GTG  
CAG TGC TTC AGC CGC TAC CCC G (7 1 -mer)

RNA bases

2' - 0- Me group on all of the RNA bases except for the base closest to the first DNA base of the GRON

(RNA base without the 2' - 0 -Me groups are in boxes)

**FIG.2**

BFP gene

5' CCTTCACCCACGGCGTGCAGTGCTTCAGCCGCTACCccgaccacatgAAGCAGCACGAC 3'

3' GGAAGTGGGTGCCGCACGTCACGAAGTCGGCGATGGGGTGGTGTACTTCGTCGTGCTG 5'

BFP CRISPR spacer sequence

BFP CRISPR PAM Sequence

Conversion site: BFP C>T

*Wobble bases (wt sequence)*

**FIG.3**



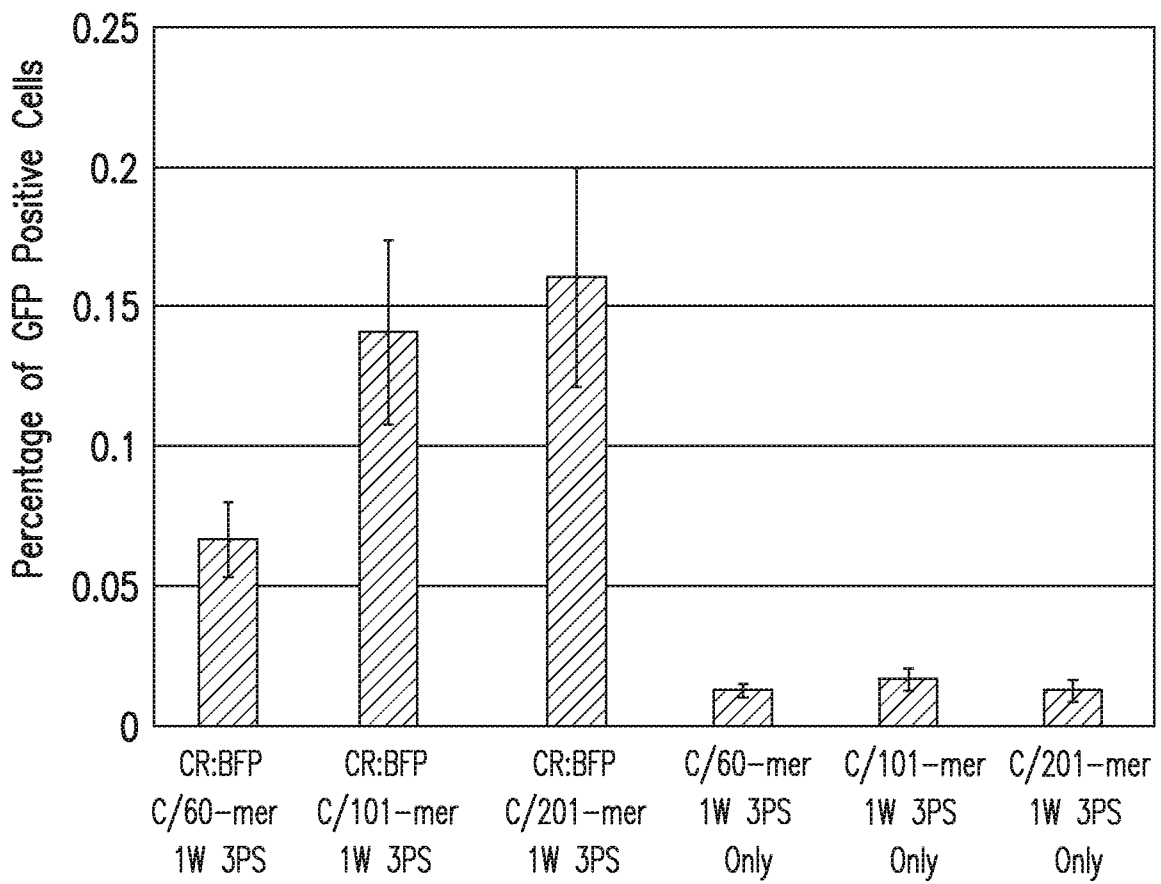


FIG. 5

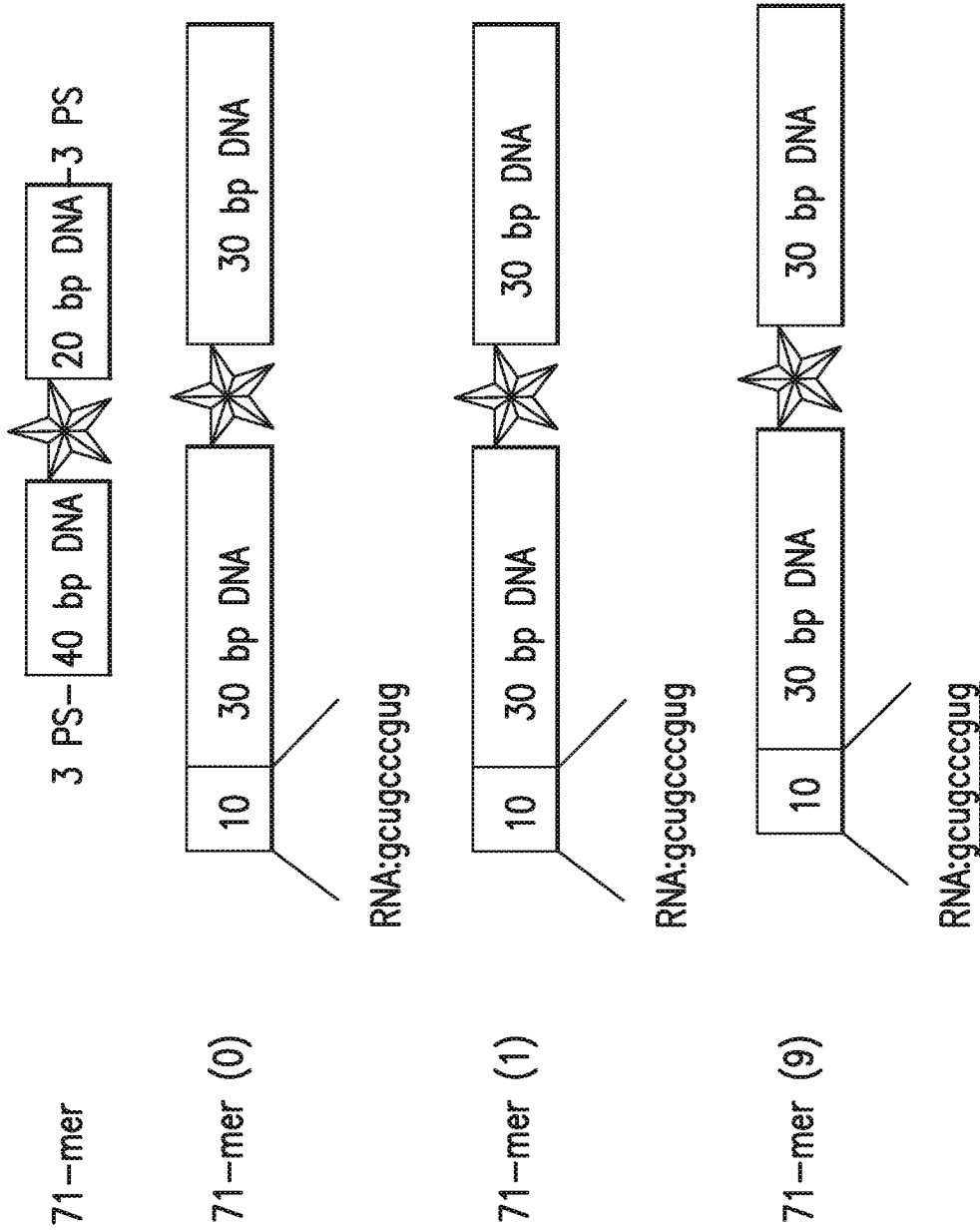
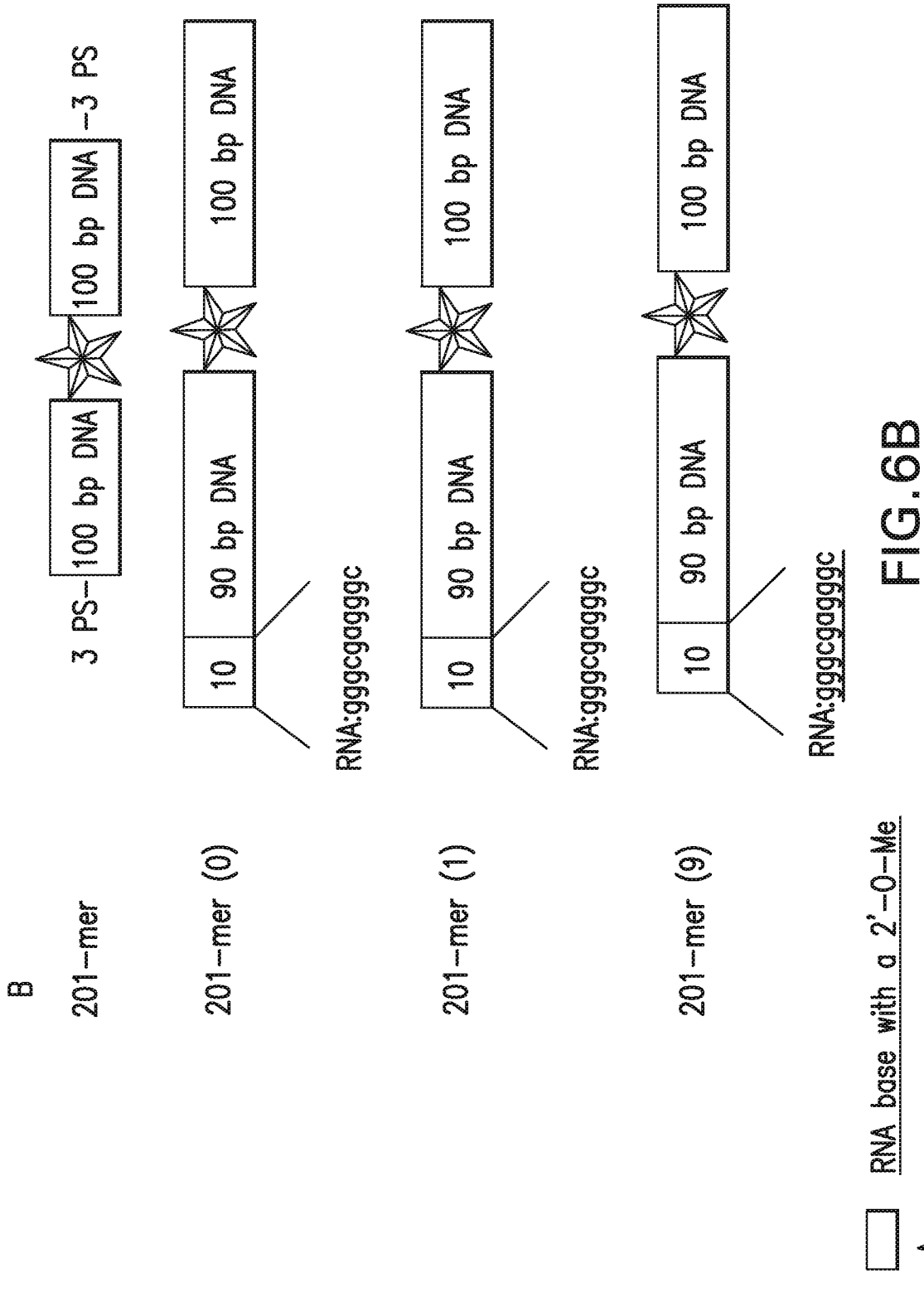


FIG.6A

RNA base with a 2'-O-Me

Site of BFP to GFP conversion C > T



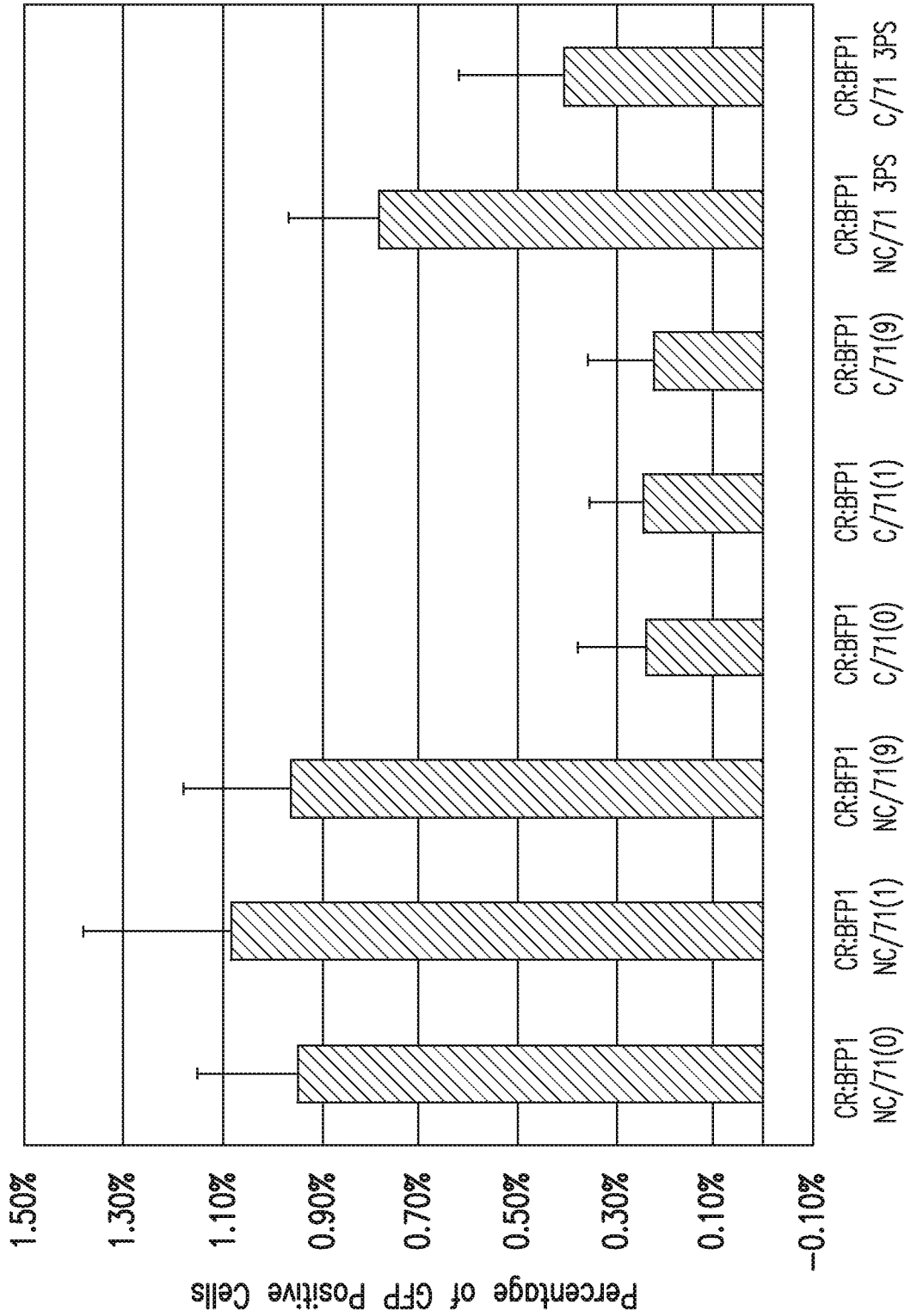


FIG. 7

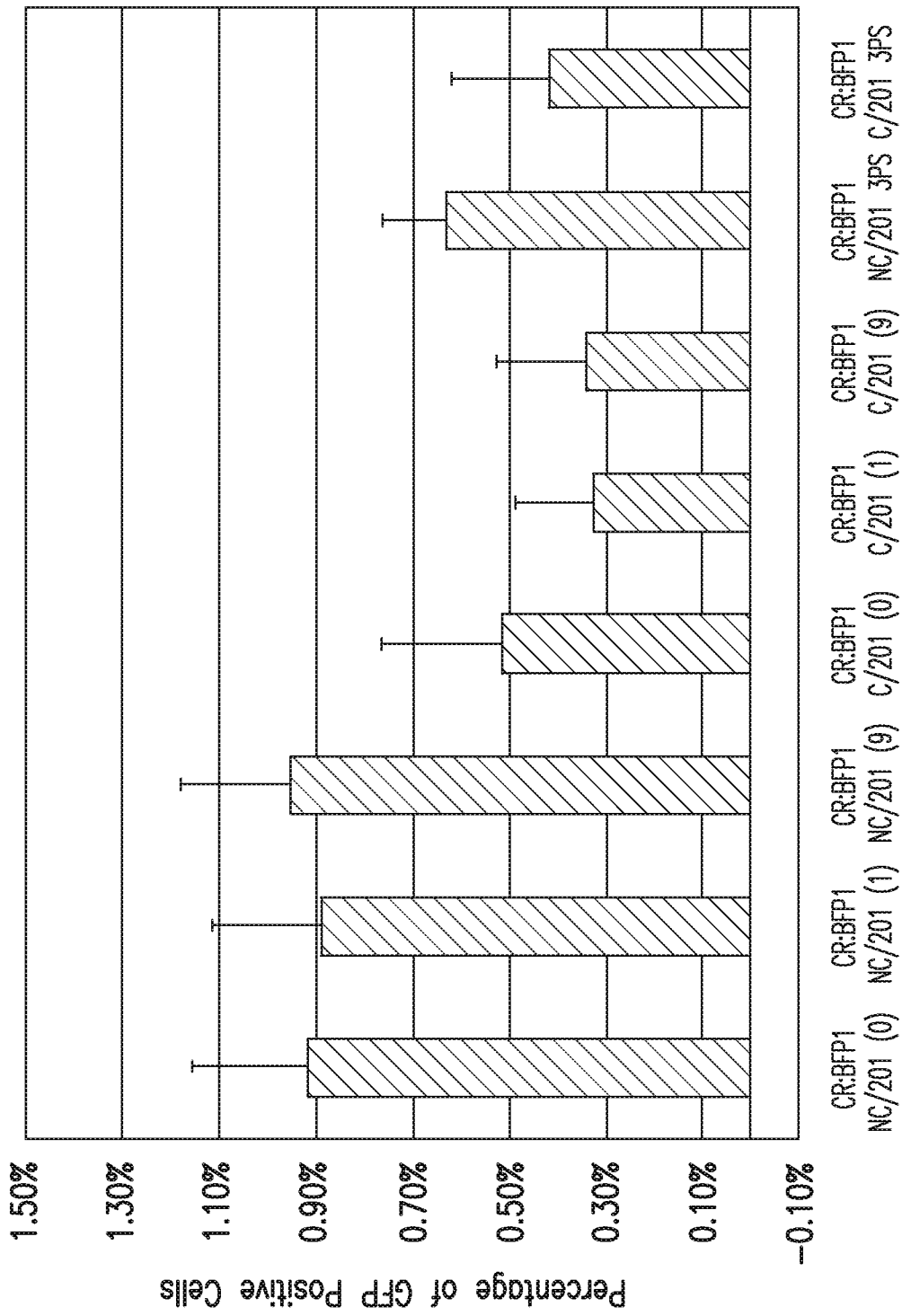


FIG.8

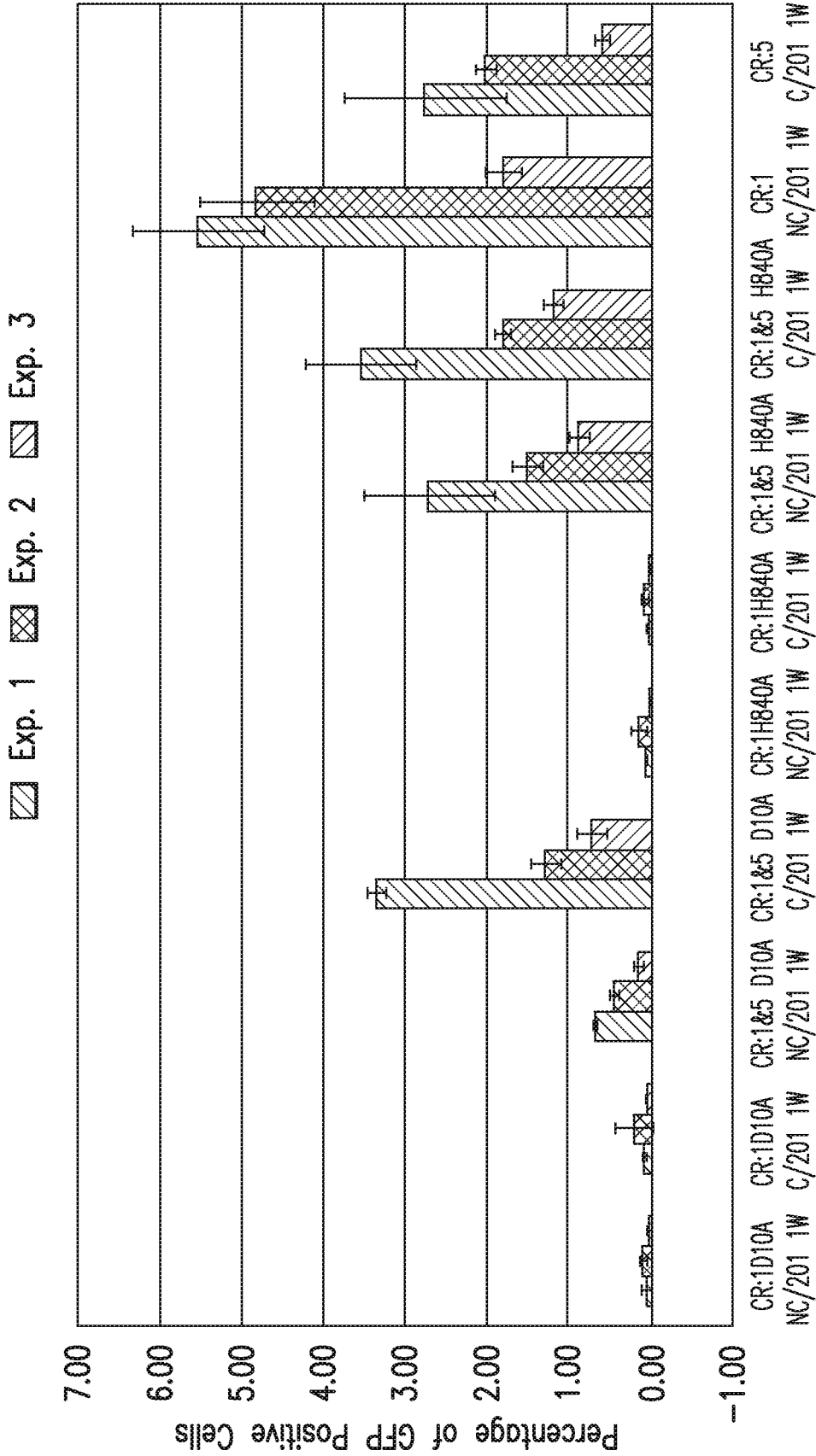


FIG.9



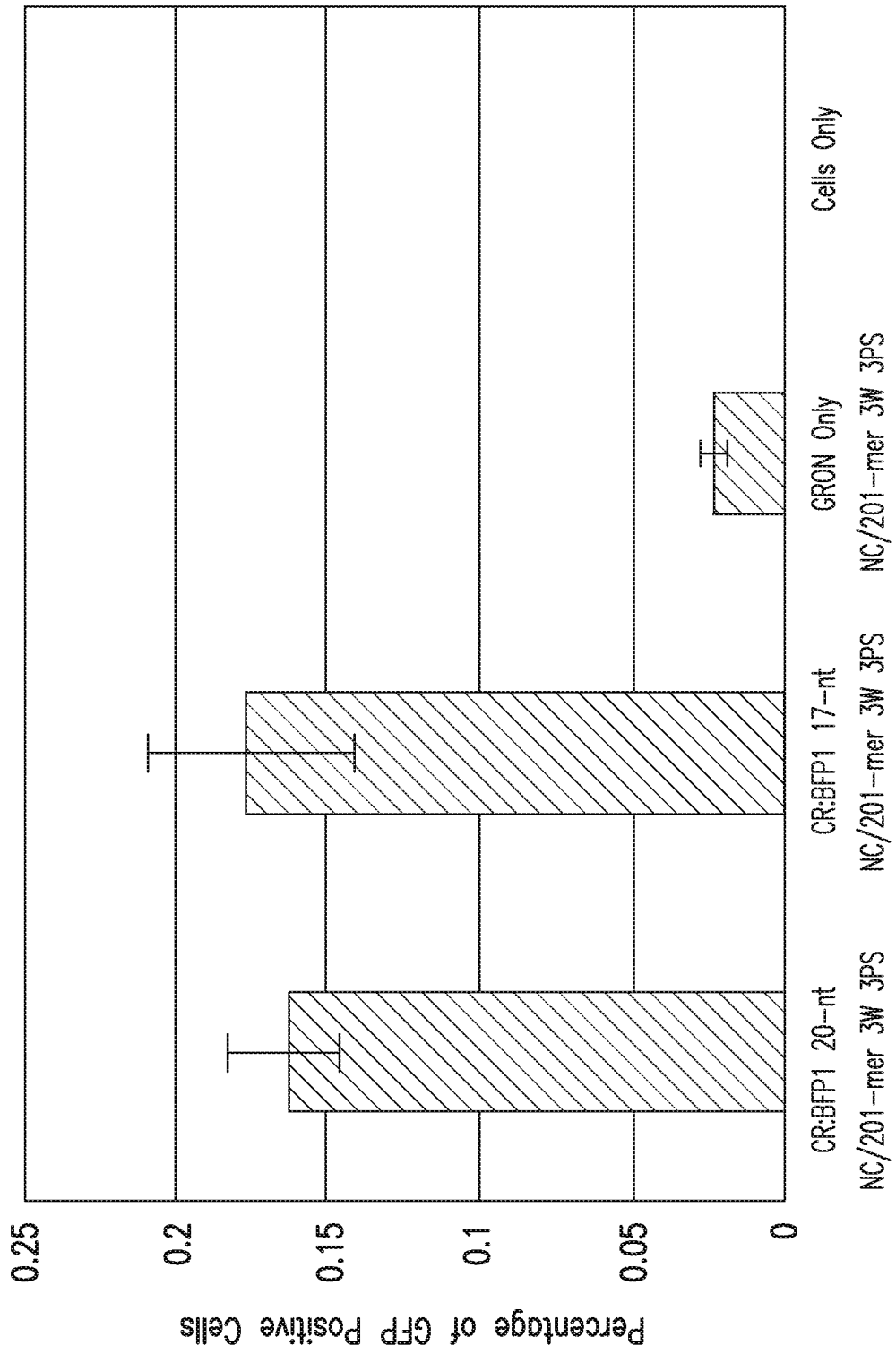


FIG.11

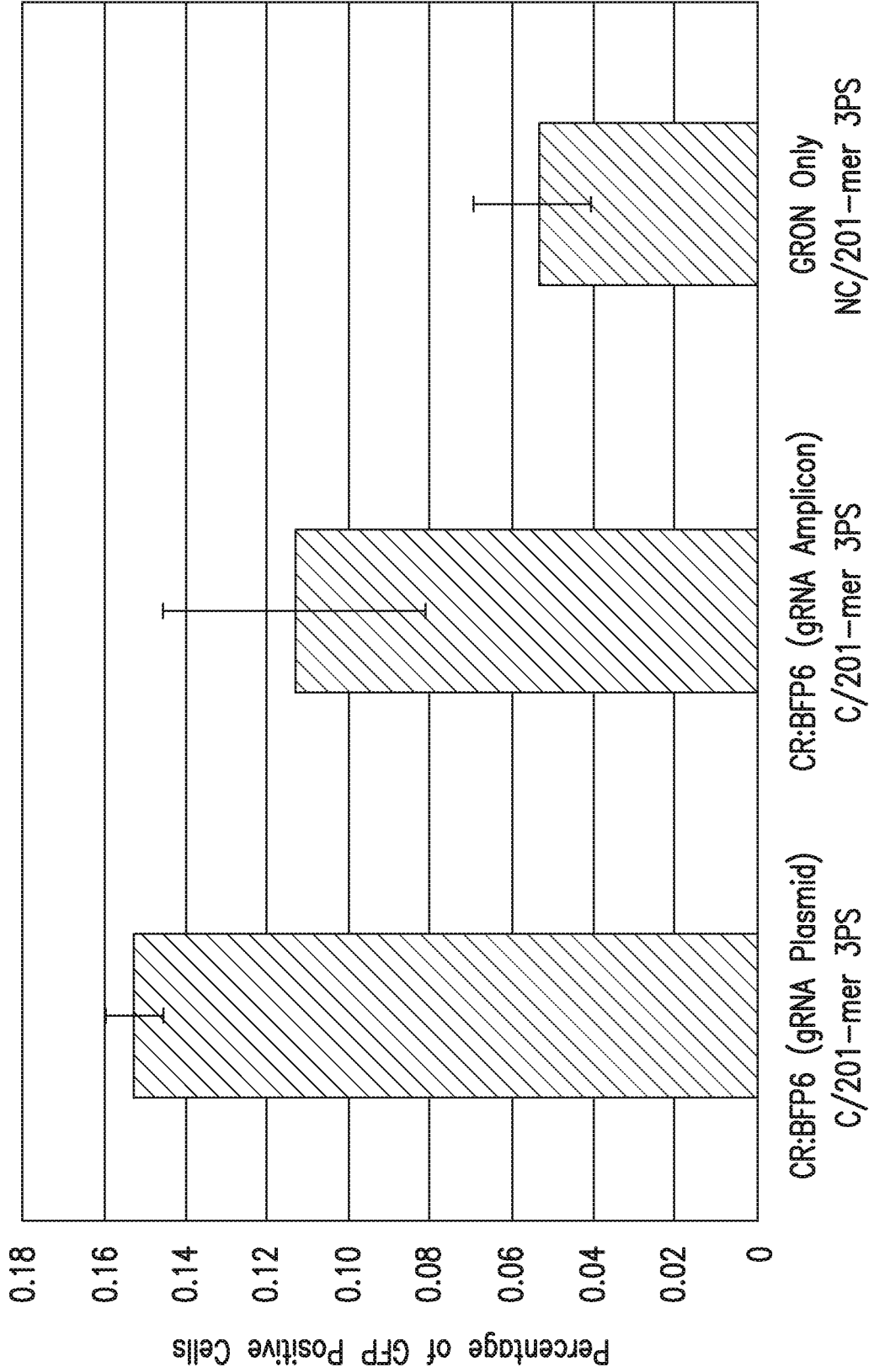


FIG.12

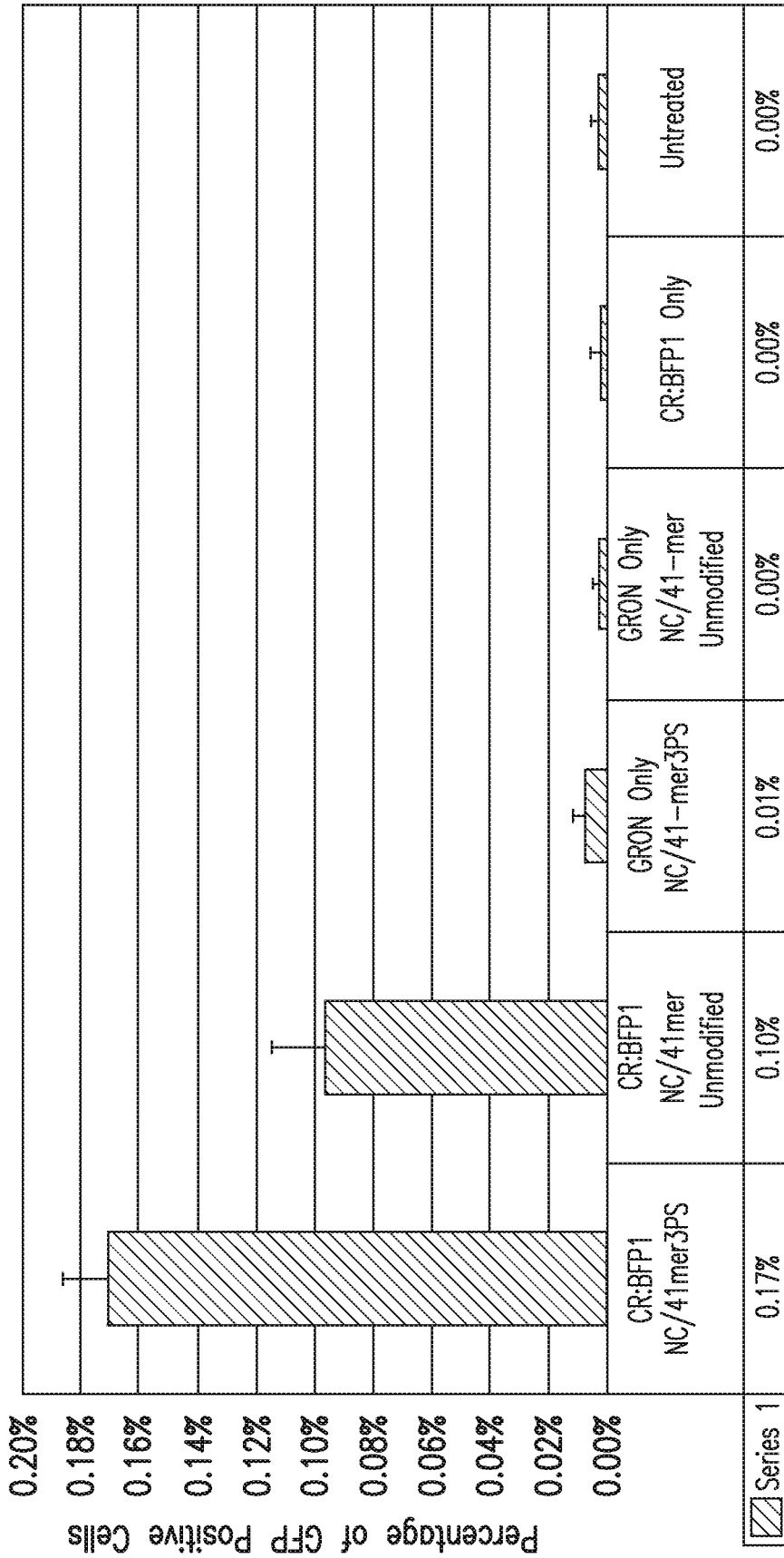


FIG.13

Treatment	Time *	GRON	TALEN	% Conversion	% Indels
Protoplasts**	24 h	Yes	Yes	0.067	2.60
Microcalli**	3 W	Yes	Yes	0.051	1.84

\* denotes time after PEG delivery of TALEN plasmids and GRONS.

\*\* denotes protoplasts and micromalli are not from the same experiment.

### FIG. 14

Treatment**	Time *	CRISPR	% Indels
Microcalli	3W	Yes	46.5
Microcalli	6W	Yes	54.7

\* Time after PEG delivery of CRISPR plasmids

\*\* Treatments are not from the same experiment.

### FIG. 15

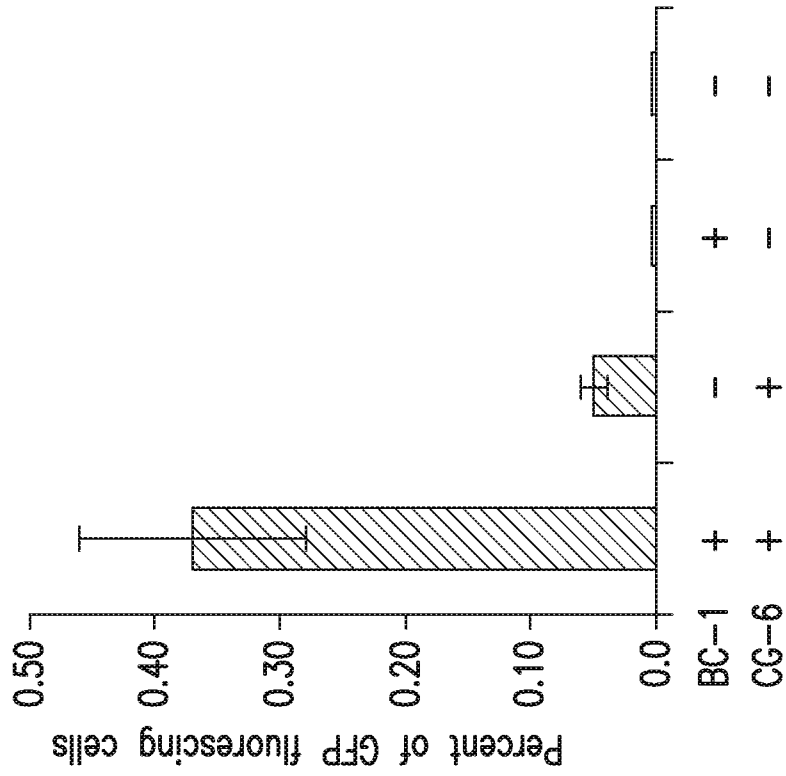


FIG.16B

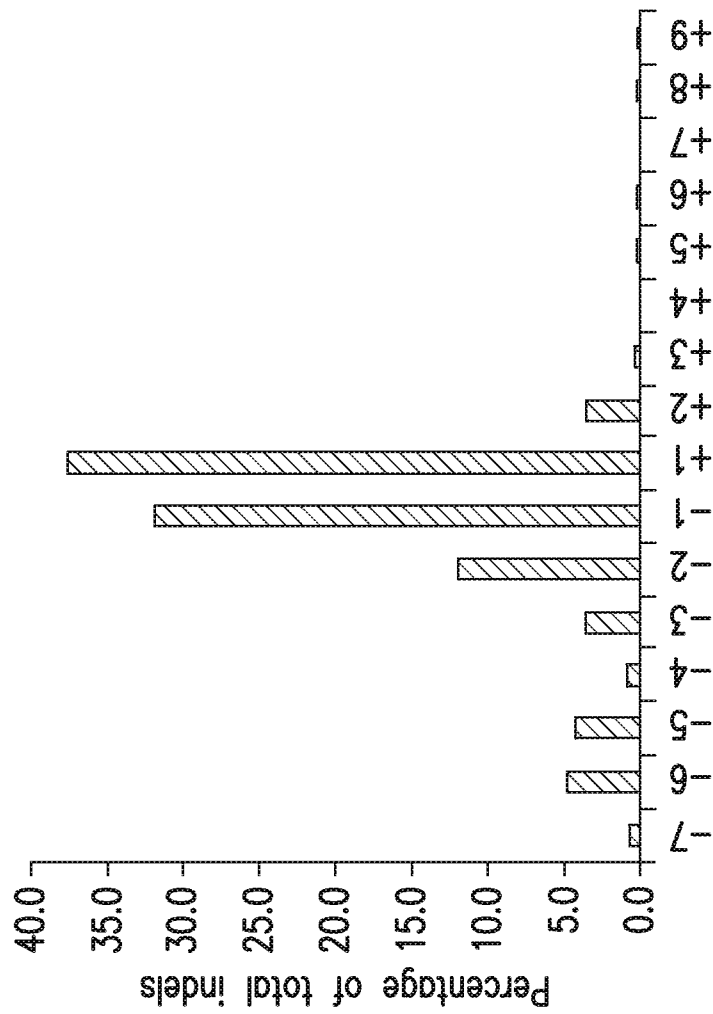


FIG.16A

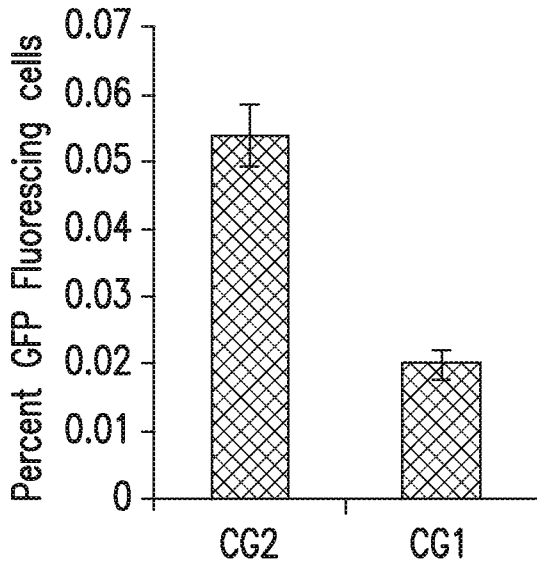


FIG. 17A

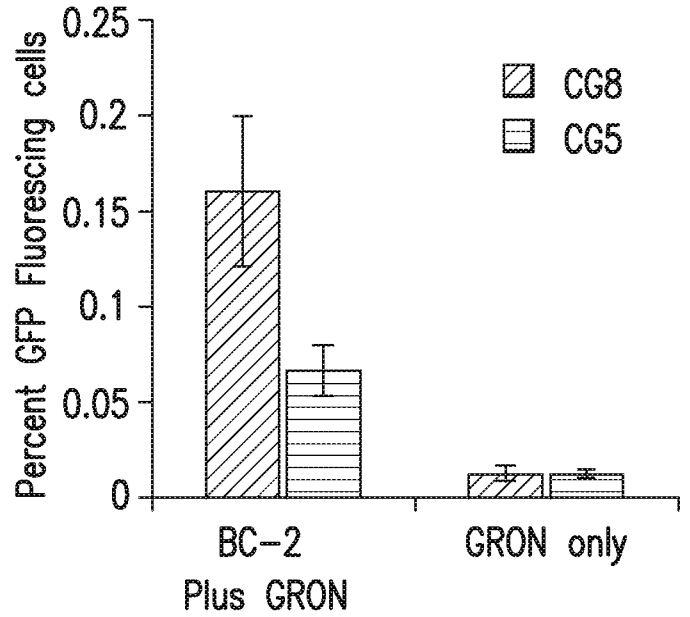


FIG. 17B

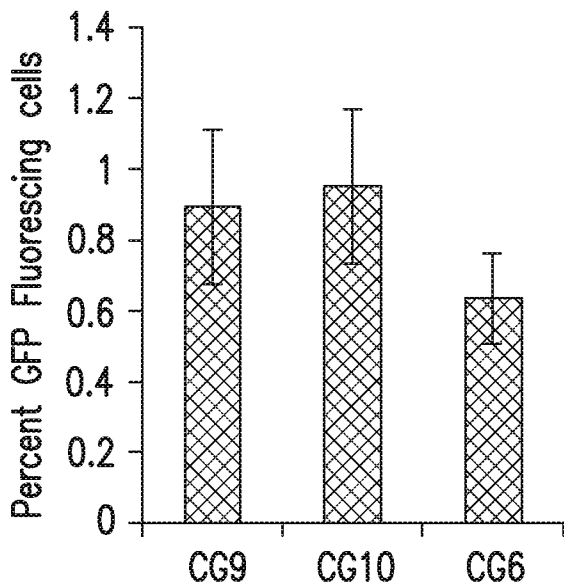


FIG. 17C

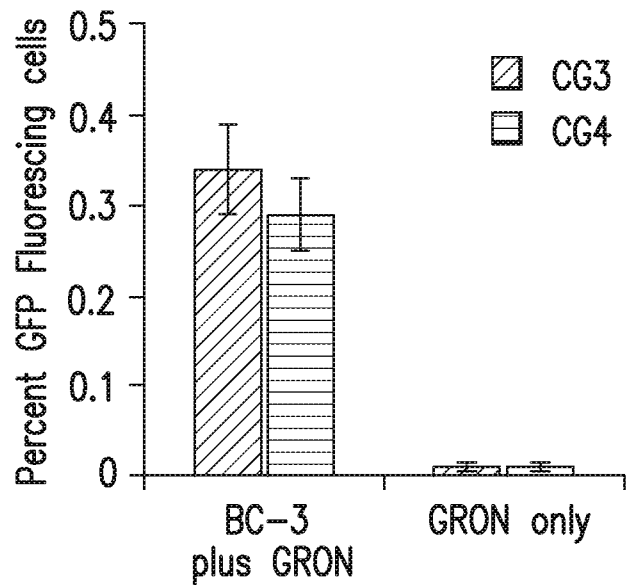


FIG. 17D

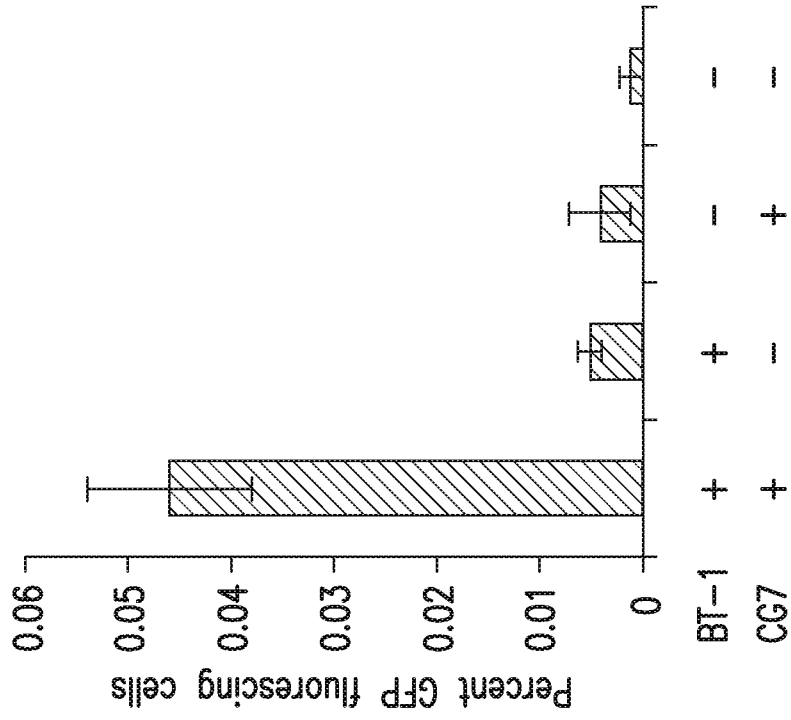


FIG.18B

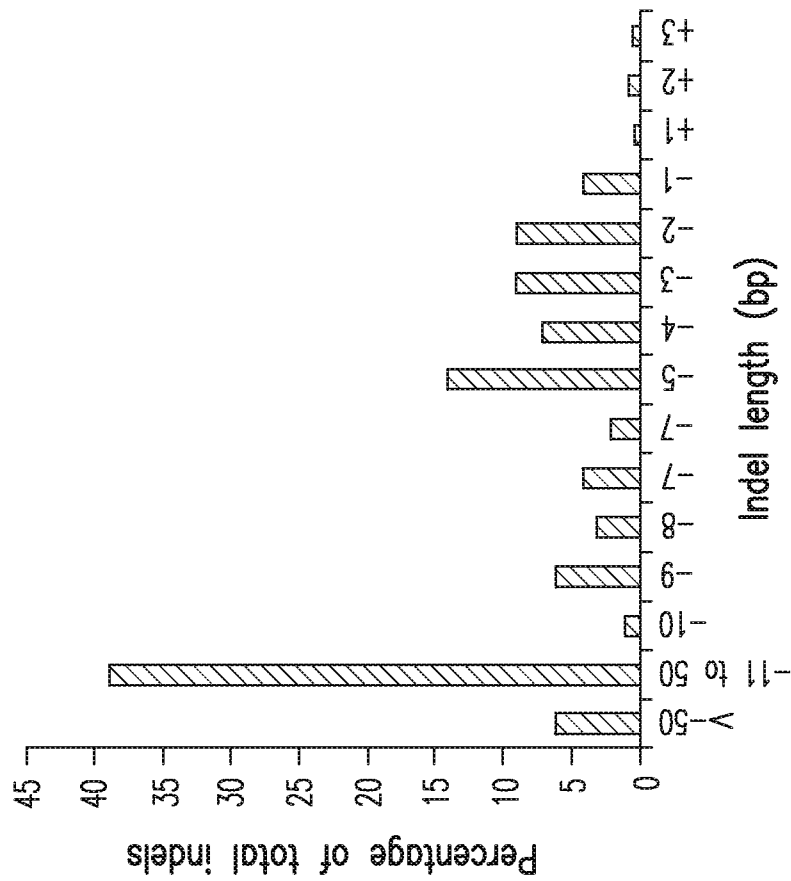


FIG.18A

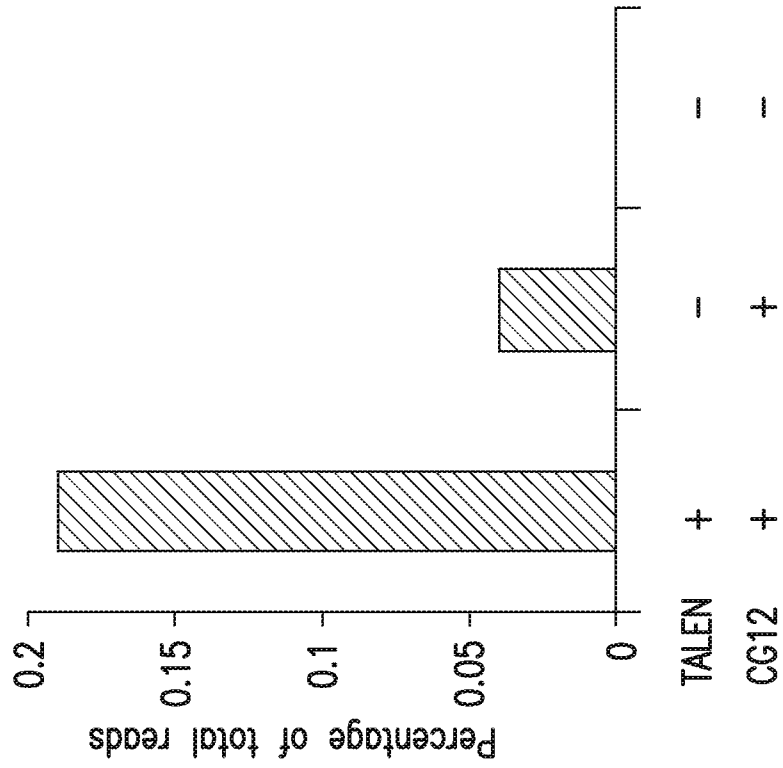


FIG.18D

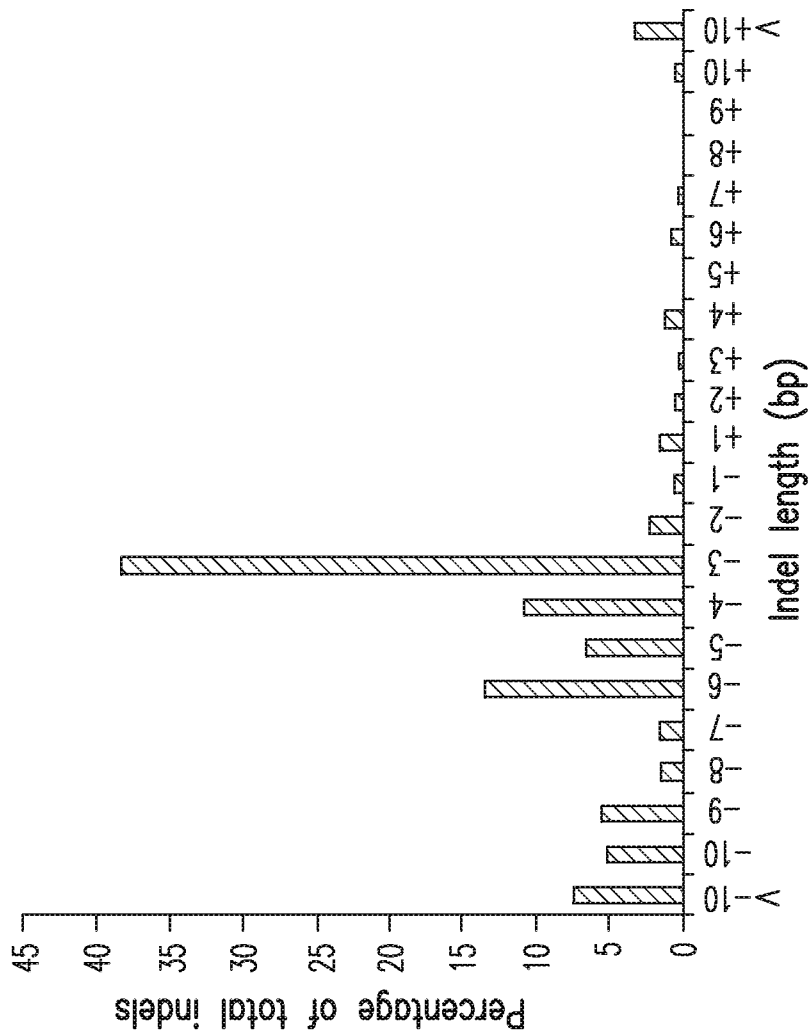


FIG.18C

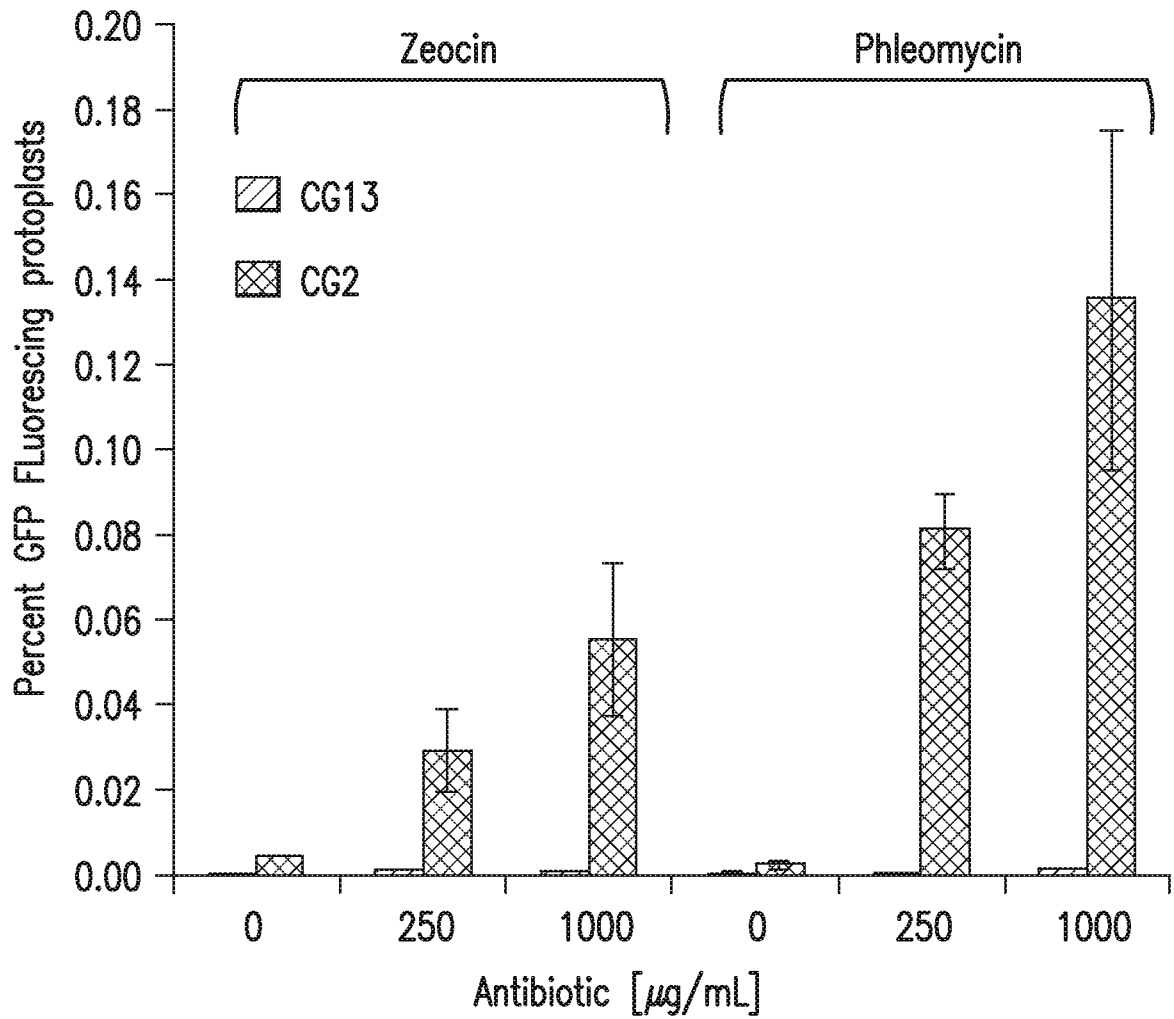


FIG. 19

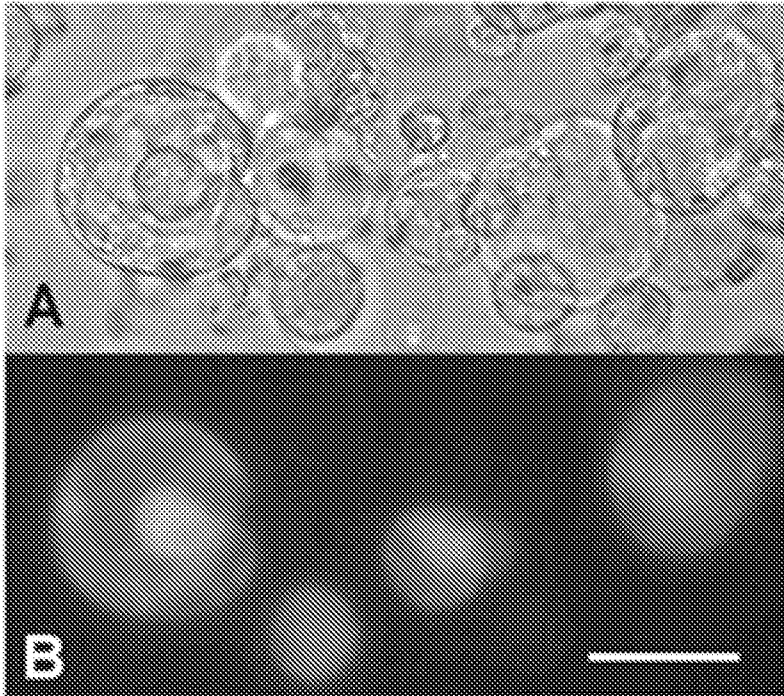


FIG. 20

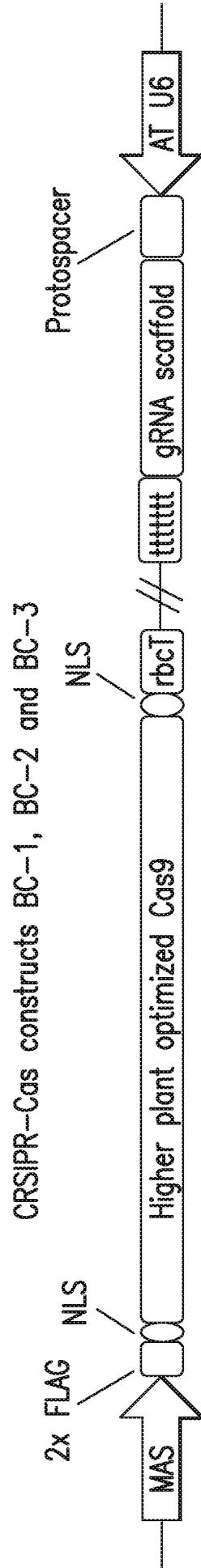


FIG. 21A

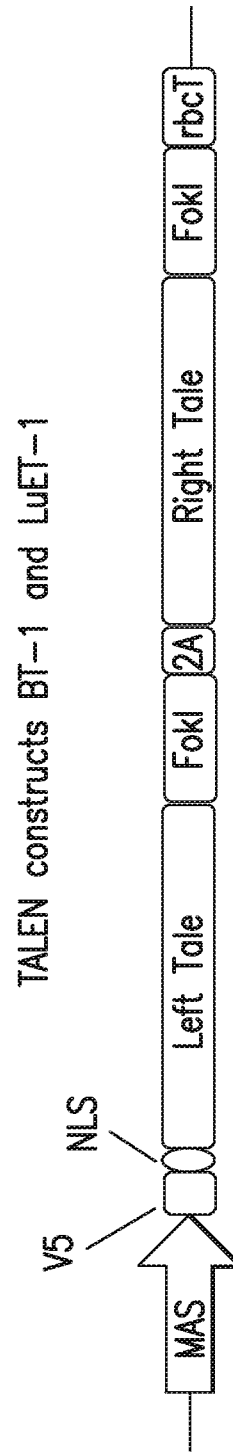


FIG. 21B

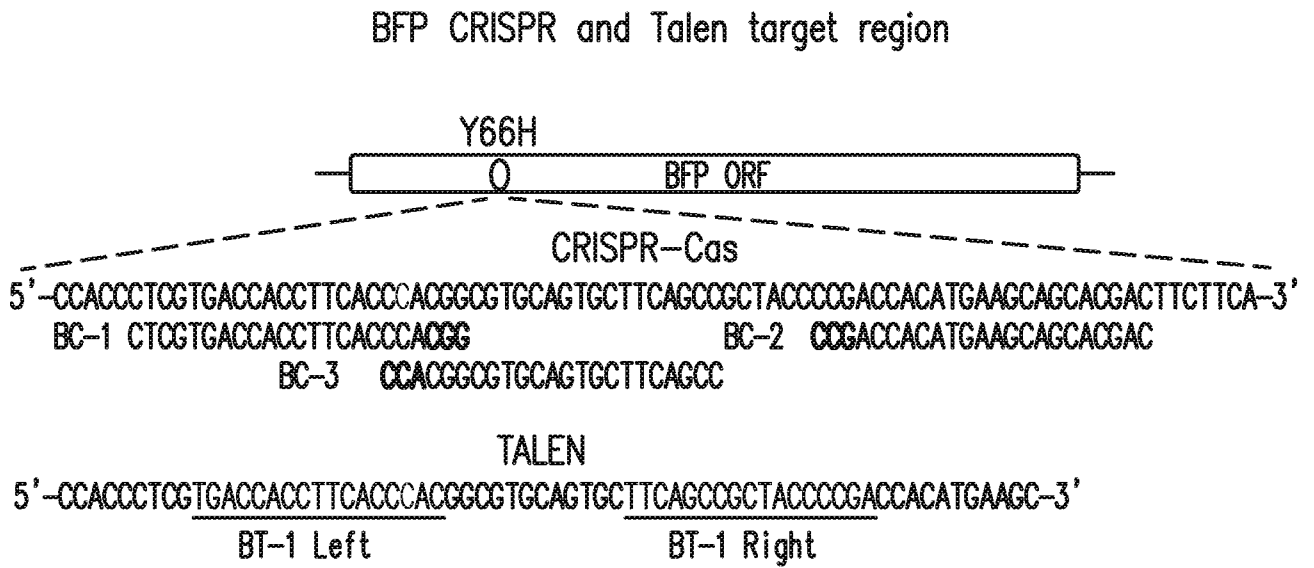


FIG. 22A

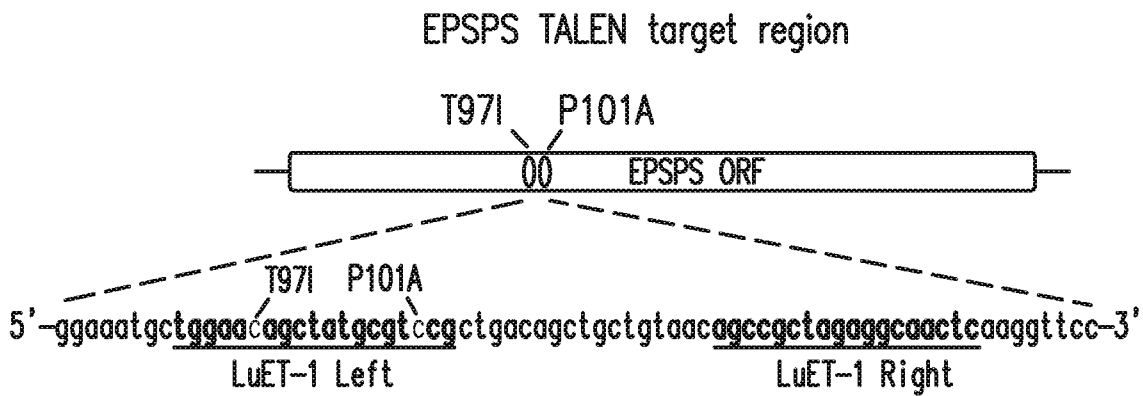


FIG. 22B

AMY310767 - *Alopecurus myosuroides* plastidal ACCase translated protein  
(SEQ ID NO:1)

```

1  MGSTHLPVIG FNASTTPSLS TLRQINSAAA AFQSSSPSRS SKKKSRRVKS
51  IRDDGDGSPV DPAGHGQSIR QLAGIIDLP KEGASAPDVD ISHGSEDHKA
101 SYQMNGILNE SHNGRHASLS KVEYFCTELG GKTPIHSLV ANNGMAAAKF
151 MRSVRTWAND TFGSEKAIQL IAMATPEDMR INAEHIRIAD QFVEVPGGTN
201 NNNYANVQLI VEIAERTGVS AVWPGWGHAS ENPELDPALT AKGIVFLGPP
251 ASSMNALGDK VGSALIAQAA GVPTLAWSGS HVEIPLELCL DSIPEEMYRK
301 ACVTTADEAV ASCQMIGYPA MIKASWGGGG KGIRKVNND D EVKALFKQVQ
351 GEVPGSPIFI MRLASQSRHL EVQLLCDEYG NVAALHSRDC SVQRRHQKII
401 EEGPVTVAPR ETVKELEQAA RRLAKAVGYV GAATVEYLYS METGEYFFLE
451 LNPRLQVEHP VTESIAEVNL PAAQVAVGMG IPLWQIPEIR RFGMDNGGG
501 YDIWRKTAAL ATPFNFDEVD SQWPKGHCVA VRITSENDD GFKPTGGKVK
551 EISFKSKPNV WGYFSVKSGG GIHEFADSQF GHVFAYGETR SAAITMSLA
601 LKEIQIRGEI HTNVDYTVDL LNA PDFRENT IHTGWL DTRI AMRVQAERPP
651 WYISVGGAL YKTITTNAET VSEYVSYLIK GQIPPKHISL VHSTISLNIE
701 ESKYTIEIVR SGQGSYRLRL NGSLIEANVQ TLC DGGLLMQ LDGNSHVIYA
751 EEEAGGTRLL IDGKTCLLQN DHDPSRLLAE TPCKLLRFLI ADGAHV DADV
801 PYAEVEVMKM CPLLSPAAG VINVLLSEQ AMQAGDIAR LDLD DPSAVK
851 RAEPFEGSFP EMSLP IAASG QVHKRCAASL NAARMVLAGY DHAANKV VQD
901 LVWCLDTPAL PFLQWEELMS VLATRLPRRL KSELEGKYNE YKLNVDHVKI
951 KDFPTEMLRE TIEENLACVS EKEMVTIERL VDPLMSLLKS YEGGRESHAH
1001 FIVKSLFEEY LSVEELFSDG IQSDVIERLR LQYSKDLQKV VDIVLSHQGV
1051 RNKTKLILAL MEKLVYPNPA AYRDLIRFS SLNHKRYYKL ALKASELLEQ
1101 TKLSELRTSI ARNLSALDMF TEEKADFSLQ DRKLAINESM GDLVTAPLPV
1151 EDALVSLFDC TDQTLQQRVI QTYISRLYQP QLVKDSIQLK YQDSGVIALW
1201 EFTEGNHEKR LGAMVILKSL ESVSTAIGAA LKDASHYASS AGNTVHIALL
1251 DADTQLNTE DSGDNDQAQD KMDKLSFVLK QDVVMADLRA ADVKVVSCIV
1301 QRDGAIMPMR RTFLLSEEKL CYEEEPILRH VEPPLSALLE LDKLKVKGYN
1351 EMKYTPSRDR QWHIYTLRNT ENPKMLHRVF FRTLVRQPSA GNRFTSDHIT
1401 DVEVGHAEEP LSFTSSSILK SLKIAKEELE LHAIRTGSH MYLCILKEQK
1451 LLDLVPVSGN TVVDVGQDEA TACSLLKEMA LKIHVGVGAR MHLVSVQWE
1501 VKLKLVS DGP ASGSWRVVT NVTGHTCTVD IYREVEDTES QKLVYHSTAL
1551 SSGPLHGVAL NTSYQPLSVI DLKRCSARNN KTTYCYDFPL TFEAAVQKSW
1601 SNISSENNQC YVKATELVFA EKNGSWGTP IIPMQRAGLN DIGMVAWILD
1651 MSTPEFPSGR QIIVIANDIT FRAGSFGPRE DAFPEAVTNL ACEKKLPLIY
1701 LAANSGARIG IADEVKSCFR VGWTDSSPE RGFRIYMTD EDHDRIGSSV
1751 IAHKMQLDSG EIRWVIDSVV GKEDGLGVEN IHGSAAIASA YSRAYEETFT
1801 LTFVTGRTVG IGAYLARLGI RCIQRIDQPI ILTGFSALNK LLGREVYSSH
1851 MQLGGPKIMA TNGVVHLTVP DDLEGVSNIL RWLSYVPANI GGPLPITKSL
1901 DPIDRPVAYI PENTCDPRAA ISGIDDSQ GK WLGGMFDKDS FVETFEGWAK
1951 TVVTGRAKLG GIPVGVIAVE TQTMMLVPA DPGQPSHER SVPRAGQVWF
2001 PDSATKTAQA MLDFNREGLP LFI LANWRGF SGGQRDLFEG ILQAGSTIVE
2051 NLRTYNQPAF VYIPKAAELR GGAWVIDSK INPDRIECYA ERTAKGNVLE
2101 POGILIEIKFR SEELKECMGR LDPELIDLKA RLQANGSLS DGESLQKSIE
2151 ARKKQLLPLY TQIAVRFAEL HDTSLRMAAK GVIRKVV DWE DSRFFYKRL
2201 RRRLEDVLA KEIRGVIGEK FPHKSAIELI KKWYLASEAA AAGSTDWDDD
2251 DAFVAWRENP ENYKEYIKEL RAQRVSRLLS DVAGSSSDLQ ALPQGLSMLL
2301 DKMDPSKRAQ FIEEVMKVLK

```

FIG.23

The *E. coli* EPSPS gene product has the following sequence (UniProt Accession # P0A6D3, SEQ ID NO:2):

MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLDSDVHRHMLNALTAL  
GVSYTLSADRTRCEIIGNGGPLHAEGALELFLGNAGTAMRPLAAALCLGSNDIVLTGEPR  
MKERPIGHLVDALRLGGAKITYLEQENYPPLRLQGGFTGGNVDVDGSVSSQFLTALLMTA  
PLAPEDTVIRIKGDLVSKPYIDITLNLTKTFGVEIENQHYQQFVVKGGQSYQSPGTYLVE  
GDASSASYFLAAAAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICWGDDYISCTRGE  
LNAIDMDMNHIPDAAMTIATAALFAKGTTTTLRNIYNWRVKETDRLFAMATELRKVGAEVE  
EGHDYIRITPPEKLNFAEIATYNDHRMAMCFSLVALSDTPVTILDPKCTAKTFPDYFEQL  
ARISQAA

## FIG. 24

Protein	Genbank Accession #	L82	T97	P101	N111
<i>E. coli</i>	X00557	82	97	101	111*
<i>Arabidopsis thaliana</i>	AF360224	160	179	183	194
<i>Petunia hybrida</i>	M21084.1	155	174	178	189
<i>Brassica napus</i>	X51475.1	155	174	178	189
<i>Zea mays</i>	X63374	84	102	106	117
<i>Oryza sativa</i>	AF413082	150	169	173	184
<i>Arabidopsis thaliana</i>	NM 130093	159	178	182	193

\* No true *E. coli* homologous acid

FIG.25

AMY310767 - *Alopecurus myosuroides* plastidial ACCase cDNA

1 TCGATAAACT TCCTGTTGCA TGCTCTATC TCTATGGACT AACGGTTCCT ATGTGCATGC  
 61 ATCTGTCAGG TTTCCAGACC TGGGGTTTTAC AATCAGTTTA TGGCAGTCTG TGTTTGAAGA  
 121 AACTGCAAC TCTGCTGTCT GTCCAAAGGG AGGACGATGG GATCCACACA TCTGCCATT  
 181 GTCGGGTTTA ATGCATCCAC AACACCATCG CTATCCACTC TTCGCCAGAT AAAGTCAGCT  
 241 GCTGCTGCAT TCCAATCTTC GTCCCTTCA AGGTCATCCA AGAAGAAAAG CCGACGTGTT  
 301 AAGTCAATAA GGGATGATGG CGATGGAAGC GTGCCAGACC CTGCAGGCCA TGGCCAGTCT  
 361 ATTCGCCAAG GTCTCGCTGG CATCATCGAC CTCCCAAAGG AGGGCGCATC AGCTCCAGAT  
 421 GTGGACATTT CACATGGGTC TGAAGACCAC AAGGCCTCCT ACCAAATGAA TGGGATACTG  
 481 AATGAATCAC ATAACGGGAG GCACGCCTCT CTGTCTAAAG TTTATGAATT TTGCACGGAA  
 541 TTGGGTGGAA AAACACCAAT TCACAGTGTA TTAGTCGCCA ACAATGGAAT GGCAGCAGCT  
 601 AAGTTCATGC GGAGTGTCGG GACATGGGCT AATGATACAT TTGGGTCAGA GAAGGCGATT  
 661 CAGTTGATAG CTATGGCAAC TCCGGAAGAC ATGAGAATAA ATGCAGAGCA CATTAGAATT  
 721 GCTGATCAGT TTGTTGAAGT ACCTGGTGGG ACAAACAATA ACAACTATGC AAATGTCCAA  
 781 CTCATAGTGG AGATAGCAGA GAGAACTGGT GTCTCCGCCG TTTGGCCTGG TTGGGGCCAT  
 841 GCATCTGAGA ATCCTGAACT TCCAGATGCA CTAAGTCAA AAGGAATTGT TTTTCTTGGG  
 901 CCACCAGCAT CATCAATGAA CGCACTAGGC GACAAGGTTG GTTCAGCTCT CATTGCTCAA  
 961 GCAGCAGGGG TTCCCACTCT TGCTTGGAGT GGATCACATG TGGAAATTCC ATTAGAATT  
 1021 TGTTTGGACT CGATACCTGA GGAGATGTAT AGGAAAGCCT GTGTTACAAC CGCTGATGAA  
 1081 GCAGTTGCAA GTTGTGAGAT GATTGGTTAC CCTGCCATGA TCAAGGCATC CTGGGGTGGT  
 1141 GGTGGTAAAG GGATTAGAAA GGTTAATAAT GATGACGAGG TGAAAGCACT GTTTAAGCAA  
 1201 GTACAGGGTG AAGTTCCTGG CTCCCGATA TTTATCATGA GACTTGCATC TCAGAGTCGT  
 1261 CATCTTGAAG TCCAGCTGCT TTGTGATGAA TATGGCAATG TAGCAGCACT TCACAGTCGT  
 1321 GATTGCAGTG TGCAACGACG ACACCAAAG ATTATCGAGG AAGGACCAGT TACTGTTGCT  
 1381 CCTCGTAAA CAGTGAAAGA GCTAGAGCAA GCAGCAAGGA GGCTTGCTAA GGCCGTGGGT  
 1441 TACGTCGGTG CTGCTACTGT TGAATATCTC TACAGCATGG AGACTGGTGA AACTATTTT  
 1501 CTGGAGCTTA ATCCACGGTT GCAGGTTGAG CACCCAGTCA CCGAGTCGAT AGCTGAAGTA  
 1561 AATTTGCTG CAGCCAAGT TGCAGTTGGG ATGGGTATAC CCCTTTGGCA GATTCCAGAG  
 1621 ATCAGACGTT TCTACGGAAT GGACAATGGA GGAGGCTATG ATATTTGGAG GAAAACAGCA  
 1681 GCTCTCGCTA CTCCATTCAA CTTTGATGAA GTAGATTCTC AATGGCCGAA GGGTCATTGT  
 1741 GTGGCAGTTA GGATAACCAG TGAGAATCCA GATGATGGAT TCAAGCCTAC TGGTGGAAAA  
 1801 GTAAAGGAGA TAAGTTTTAA AAGTAAGCCA AATGTCTGGG GATATTTCTC AGTTAAGTCT  
 1861 GGTGGAGGCA TTCATGAATT TGCGGATTCT CAGTTTGGAC ACGTTTTTGC CTATGGAGAG  
 1921 ACTAGATCAG CAGCAATAAC CAGCATGTCT CTTGCACTAA AAGAGATTCA AATTCGTGGA  
 1981 GAAATTCATA CAAACGTTGA TTACACGGTT GATCTCTTGA ATGCCCCAGA CTTCAGAGAA  
 2041 AACACGATCC ATACCGGTTG GCTGGATACC AGAATAGCTA TGCCTGTTCA AGCTGAGAGG  
 2101 CCTCCCTGGT ATATTTCACT GGTGGAGGA GCTCTATATA AAACAATAAC CACCAATGCG  
 2161 GAGACCGTTT CTGAATATGT TAGCTATCTC ATCAAGGGTC AGATTCCACC AAAGCACATA  
 2221 TCCCTTGTCC ATTCAACTAT TTCTTTGAAT ATAGAGGAAA GCAAATATAC AATTGAGATT  
 2281 GTGAGGAGTG GACAGGGTAG CTACAGATTG AGACTGAATG GATCACTTAT TGAAGCCAAT  
 2341 GTACAAACAT TATGTGATGG AGGCCTTTTA ATGCAGCTGG ATGGAAATAG CCATGTTATT  
 2401 TATGCTGAAG AAGAAGCGGG TGGTACACGG CTTCTTATTG ATGGAAAAAC ATGCTTGCTA

FIG. 26A-1

2461 CAGAATGACC ATGATCCGTC AAGGTTATTA GCTGAGACAC CCTGCAAAC TCTTCGTTTC  
 2521 TTGATTGCCG ATGGTGCTCA TGTTGATGCT GATGTACCAT ACGCGGAAGT TGAGGTTATG  
 2581 AAGATGTGCA TGCCCCTCTT GTCGCTGCT GCTGGTGTCA TTAATGTTTT GTTGTCTGAG  
 2641 GGCCAGGCGA TGCAGGCTGG TGATCTTATA GCGAGACTTG ATCTCGATGA CCCTTCTGCT  
 2701 GTGAAGAGAG CCGAGCCATT TGAAGGATCT TTTCCAGAAA TGAGCCTTCC TATTGCTGCT  
 2761 TCTGGCCAAG TTCACAAAAG ATGTGCTGCA AGTTTGAACG CTGCTCGAAT GGTCCCTGCA  
 2821 GGATATGACC ATGCGGCCAA CAAAGTTGTG CAAGATTTGG TATGGTGCCT TGATACACCT  
 2881 GCTCTTCTTT TCCTACAATG GGAAGAGCTT ATGTCTGTTT TAGCAACTAG ACTTCCAAGA  
 2941 CGTCTTAAGA GCGAGTTGGA GGGCAAATAC AATGAATACA AGTTAAATGT TGACCATGTG  
 3001 AAGATCAAGG ATTTCCCTAC CGAGATGCTT AGAGAGACAA TCGAGGAAAA TCTTGCATGT  
 3061 GTTTCCGAGA AGGAAATGGT GACAATTGAG AGGCTTGTTG ACCCTCTGAT GAGCCTGCTG  
 3121 AAGTCATACG AGGGTGGGAG AGAAAGCCAT GCCCACTTTA TTGTCAAGTC CCTTTTTGAG  
 3181 GAGTATCTCT CGGTTGAGGA ACTATTCAGT GATGGCATTG AGTCTGACGT GATTGAACGC  
 3241 CTGCGCCTAC AATATAGTAA AGACCTCCAG AAGGTTGTAG ACATTGTTTT GTCTCACCAG  
 3301 GGTGTGAGAA ACAAAACAAA GCTGATACTC GCGCTCATGG AGAACTGGT CTATCCAAAC  
 3361 CCTGCTGCCT ACAGAGATCA GTTGATTGCG TTTTCTTCCC TCAACCATAA AAGATATTAT  
 3421 AAGTTGGCTC TFAAAGCTAG TGAACCTCTT GAACAAACCA AGCTCAGCGA ACTCCGCACA  
 3481 AGCATTGCAA GGAACCTTTC AGCGCTGGAT ATGTTACCCG AGGAAAAGGC AGATTTCTCC  
 3541 TTGCAAGACA GAAAATTGGC CATTAATGAG AGCATGGGAG ATTTAGTCAC TGCCCCACTG  
 3601 CCAGTTGAAG ATGCACTTGT TTCTTTGTTT GATTGTACTG ATCAAACCTCT TCAGCAGAGA  
 3661 GTGATTCAGA CACATATC TCGATTATAC CAGCCTCAAC TTGTGAAGGA TAGCATCCAG  
 3721 CTGAAATATC AGGATTCTGG TGTTATTGCT TTATGGGAAT TCACTGAAGG AAATCATGAG  
 3781 AAGAGATTGG GTGCTATGGT TATCCTGAAG TCACTAGAAT CTGTGTCAAC AGCCATTGGA  
 3841 GCTGCTCTAA AGGATGCATC ACATTATGCA AGCTCTGCGG GCAACACGGT GCATATTGCT  
 3901 TTGTTGGATG CTGATACCCA ACTGAATACA ACTGAAGATA GTGGTGATAA TGACCAAGCT  
 3961 CAAGACAAGA TGGATAAACT TTCTTTTGTA CTGAAACAAG ATGTTGTCAT GGCTGATCTA  
 4021 CGTGCTGCTG ATGTCAAGGT TGTTAGTTGC ATTGTTCAA GAGATGGAGC AATCATGCCT  
 4081 ATGCGCCGTA CCTTCTCTT GTCAGAGGAA AAACCTTGT ACAGGAAGA GCCGATTCTT  
 4141 CGGCATGTGG AGCCTCCACT TTCTGCACTT CTTGAGTTGG ATAAATTGAA AGTGAAAGGA  
 4201 TACAATGAGA TGAAGTATAC ACCGTCACGT GATCGTCAGT GGCATATATA CACACTTAGA  
 4261 AATACTGAAA ATCCAAAAAT GCTGCACAGG GTATTTTTCC GAACACTTGT CAGACAACCC  
 4321 AGTGCAGGCA ACAGGTTTAC ATCAGACCAT ATCACTGATG TTGAAGTAGG ACACGCAGAG  
 4381 GAACCTCTTT CATTACTTC AAGCAGCATA TAAAAATCGT TGAAGATTGC TAAAGAAGAA  
 4441 TTGGAGCTTC ACGGATCAG GACTGGCCAT TCTCATATGT ACTTGTGCAT ATTGAAAGAG  
 4501 CAAAAGCTTC TTGACCTTGT TCCTGTTTCA GGGAAACTG TTGTGGATGT TGGTCAAGAT  
 4561 GAAGCTACTG CATGCTCTCT TTTGAAAGAA ATGGCTTTAA AGATACATGA ACTTGTGGT  
 4621 GCAAGAATGC ATCATCTTTC TGTATGCCAG TGGGAAGTGA AACTTAAGTT GGTGAGCGAT  
 4681 GGGCCTGCCA GTGGTAGCTG GAGAGTTGTA ACAACCAATG TTAAGTGGTCA CACCTGCACT  
 4741 GTGGATATCT ACCGGGAGGT CGAAGATACA GAATCACAGA AACTAGTATA CCACTCCACC  
 4801 GCATTGTCAT CTGGTCCTTT GCATGGTGTT GCACTGAATA CTTCGTATCA GCCTTTGAGT  
 4861 GTTATTGATT TAAAACGTTG CTCTGCCAGG AACAAACAAA CTACATACTG CTATGATTTT  
 4921 CCATTGACAT TTGAAGCTGC AGTGCAGAAG TCGTGGTCTA ACATTTCCAG TGAACAAC

## FIG. 26A-2

4981 CAATGTTATG TTAAAGCGAC AGAGCTTGTG TTTGCTGAAA AGAATGGGTC GTGGGGCACT  
 5041 CCTATAATTC CTATGCAGCG TGCTGCTGGG CTGAATGACA TTGGTATGGT AGCCTGGATC  
 5101 TTGGACATGT CCACTCCTGA ATTTCCCAGC GGCAGACAGA TCATTGTTAT CGCAAATGAT  
 5161 ATTACATTTA GAGCTGGATC ATTTGGCCCA AGGGAAGATG CATTITTCGA AGCTGTAACC  
 5221 AACCTGGCTT GTGAGAAGAA GCTTCCACTT ATCTACTTGG CTGCAAACCT TGGTGCTCGG  
 5281 ATTGGCATTG CTGATGAAGT AAAATCCTGC TTCCGTGTTG GATGGACTGA TGATAGCAGC  
 5341 CCTGAACGTG GATTTAGGTA CATTTATATG ACTGACGAAG ACCATGATCG TATTGGCTCT  
 5401 TCAGTTATAG CACACAAGAT GCAGCTAGAT AGTGGCGAGA TCAGGTGGGT TATTGATTCT  
 5461 GTTGTGGGAA AAGAGGATGG ACTAGGTGTG GAGAACATAC ATGGAAGTGC TGCTATTGCC  
 5521 AGTGCCATT CTAGGGCGTA CGAGGAGACA TTTACACTTA CATTTCGTTAC TGGACGAACT  
 5581 GTTGGAAATCG GAGCCTATCT TGCTCGACTT GGCATACGGT GCATACAGCG TATTGACCAG  
 5641 CCCATTATTT TGACCGGGTT TTCTGCCCTG AACAAAGCTT TGGGGCGGGA GGTGTACAGC  
 5701 TCCCACATGC AGTTGGGTGG TCCCAAAATC ATGGCGACGA ATGGTGTGT CCATCTGACT  
 5761 GTTCCAGATG ACCTTGAAGG TGTTTCTAAT ATATTGAGGT GGCTCAGCTA TGTTCTGCA  
 5821 AACATTGGTG GACCTCTTCC TATTACAAAA TCTTTGGACC CAATAGACAG ACCCGTTGCA  
 5881 TACATCCCTG AGAATACATG TGATCCTCGT GCAGCCATCA GTGGCATTGA TGACAGCCAA  
 5941 GGGAAATGGT TGGGTGGCAT GTTTGACAAA GACAGTTTTG TGGAGACATT TGAAGGATGG  
 6001 GCGAAGACAG TAGTTACTGG CAGAGCAAAA CTTGGAGGGA TTCCTGTTGG TGTTATAGCT  
 6061 GTGGAGACAC AGACCATGAT GCAGCTCGTC CCCGCTGATC CAGGCCAGCC TGATTCCCAC  
 6121 GAGCGGTCTG TTCCTCGTGC TGGGCAAGTT TGGTTTCCAG ATTCTGCTAC CAAGACAGCG  
 6181 CAGGCGATGT TGGACTTCAA CCGTGAAGGA TTACCTCTGT TCATACTTGC TAACTGGAGA  
 6241 GGCTTCTCTG GAGGGCAAAG AGATCTTTTT GAAGGAATTC TGCAGGCTGG GTCAACAATT  
 6301 GTTGAAGACC TTAGGACATA CAATCAGCCT GCCTTTGTAT ATATCCCCAA GGCTGCAGAG  
 6361 CTACGTGGAG GAGCCTGGGT CGTGATTGAT AGCAAGATAA ACCCAGATCG CATCGAGTGC  
 6421 TATGCTGAGA GGAATGAAA GGGTAATGTT CTCGAACCTC AAGGGTTGAT TGAGATCAAG  
 6481 TTCAGGTCAG AGGAACTCAA AGAATGCATG GGTAGGCTTG ATCCAGAATT GATAGATCTG  
 6541 AAAGCAAGAC TCCAGGGAGC AAATGGAAGC CTATCTGATG GAGAATCCCT TCAGAAGAGC  
 6601 ATAGAAGCTC GGAAGAAACA GTTGCTGCCT CTGTACACCC AAATCGCGGT ACGTTTTGCG  
 6661 GAATTGCACG ACACTTCCCT TAGAATGGCT GCTAAAGGTG TGATCAGGAA AGTTGTAGAC  
 6721 TGGGAAGACT CTCGGTCTTT CTTCTACAAG AGATTACGGA GGAGGCTATC CGAGGACGTT  
 6781 CTGGCAAAGG AGATTAGAGG TGTAATTGGT GAGAAGTTTC CTCACAAATC AGCGATCGAG  
 6841 CTGATCAAGA AATGGTACTT GGCTTCTGAG GCAGCTGCAG CAGGAAGCAC CGACTGGGAT  
 6901 GACGACGATG CTTTTGTGCG CTGGAGGGAG AACCTGAAA ACTATAAGGA GTATATCAAA  
 6961 GAGCTTAGGG CTCAAAGGGT ATCTCGGTTG CTCTCAGATG TTGCAGGCTC CAGTTCGGAT  
 7021 TTACAAGCCT TGCCGCAGGG TCTTCCATG CTACTAGATA AGATGGATCC CTCTAAGAGA  
 7081 GCACAGTTTA TCGAGGAGGT CATGAAGGTC CTGAAATGAT CAAATGATAC CAACACATCC  
 7141 AATACAGTAT GTGCATGATA TCTGTTTCTC TTGAAGTACA TATATAGATG GATACAAGGC  
 7201 GGCTGTAACT GATGGTAGCT AATCTGGGCC AACCTTACT TTTGTGAACT TGCTGGTGGC  
 7261 CTTTATTATT CAAGGCACAG CTCGCCCTCG GACCCCTCC GGCTGGTTGA TGATGAGTGT  
 7321 AACTGGATGT GTTAGTTCTG CTGCCACAGA ATTCGAGAAG GATAGGGGCA TGCGGGTTTT  
 7381 GCCTCCTGTT GGCAAGAACA CTGGTGATTT TGAGTTCCTG TTATGTGGAC TGTGGTAGTC  
 7441 TTGTTTCGCT GTAGTTCGTG GATGTTCTAT CTCCTGTAAT TCTAGTCTTG GGAGAGTGAT  
 7501 TCAGATGTCC ATTCAATTTT GAACTTGAAT AATAATATGC TTTGTAGGCC TATGCGTACC  
 7561 AGTATGTGGA ATAAATGTTT GTTGAGTTA

## FIG. 26A-3

AMY310767 - *Alopecurus myosuroides* plastidal ACCase translated protein

```

1   MGSTHLPIVG FNASTTPSLS TLRQINSAAA AFQSSSPSRS SKKKSRRVKS
51  IRDDGDGSPV DPAGHGQSIR OGLAGIIDLP KEGASAPDVD ISHGSEDHKA
101 SYQMNGILNE SHNGRHASLS KVYEFCTELG GKTPIHSLV ANNGMAAAKF
151 MRSVRTWAND TFGSEKAIQL IAMATPEDMR INAEHIRIAD QFVEVPGGTN
201 NNNYANVQLI VEIAERTGVS AVWPGWGHAS ENPELPDALT AKGIVFLGPP
251 ASSMNALGDK VGSALIAQAA GVPTLAWSGS HVEIPLELCL DSIPEEMYRK
301 ACVTTADEAV ASCQMIGYPA MIKASWGGGG KGIRKVNNDG EVKALFKQVQ
351 GEVPGSPIFI MRLASQSRHL EVQLLCDEYG NVAALHSRDC SVQRRHQKII
401 EEGPVTVAPR ETVKELEQAA RRLAKAVGYV GAATVEYLYS METGEYFFLE
451 LNPRLQVEHP VTESIAEVNL PAAQVAVGMG IPLWQIPEIR RFGYMDNGGG
501 YDIWRKTAAL ATPFNFDEVD SQWPKGHCVA VRITSENPDG GFKPTGGKVK
551 EISFKSKPNV WGYFSVKSGG GIHEFADSQF GHVFAYGETR SAAITSMCLA
601 LKEIQIRGEI HTNVDYTVDL LNAPDFRENT IHTGWLDTRI AMRVQAERPP
651 WYISVVGAL YKTITTNAET VSEYVSYLIK GQIPPKHISL VHSTISLNIE
701 ESKYTIEIVR SGQGSYRLRL NGSLIEANVQ TLCDCGGLLMQ LDGNSHVIYA
751 EEEAGGTRLL IDGKTCLLQN DHDPSRLLAE TPCKLLRFLI ADGAHVADAV
801 PYAEVEVMKM CPLLSPAAG VINVLLSEGQ AMQAGDLIAR LDLDPPSAVK
851 RAEPFEGSFP EMSLPIAASG QVHKRCAASL NAARMVLGYD DHAANKVVQD
901 LVWCLDTPAL PFLQWEELMS VLATRLPRRL KSELEGKYNE YKLNVDHVKI
951 KDFPTEMLRE TIEENLACVS EKEMVTIERL VDPLMSLLKS YEGGRESHAH
1001 FIVKSLFEEY LSVEELFSDG IOSDVIERLR LQYSKDLQKV VDIVLSHQGV
1051 RNKTKLILAL MEKLVYPNPA AYRDQLIRFS SLNHKRYYKL ALKASELLEQ
1101 TKLSELRTSI ARNLSALDMF TEEKADFSLQ DRKLAINESM GDLVTAPLPV
1151 EDALVSLFDC TDQTLQQRVI QTYISRLYQP QLVKDSIQLK YQDSGVIALW
1201 EFTEGNHEKR LGAMVILKSL ESVSTAIGAA LKDASHYASS AGNTVHIALL
1251 DADTQLNTE DSGDNDQAQD KMDKLSFVLK QDVVMADLRA ADVKVVSCIV
1301 QRDGAIMPMR RTFLLSEEKL CYEEEPILRH VEPPLSALLE LDKLKVKGYN
1351 EMKYTPSRDR QWHIYTLRNT ENPKMLHRVF FRTLVRQPSA GNRFTSDHIT
1401 DVEVGHAEEP LSFTSSSILK SLKIAKEELE LHAIRTGSH MYLCILKEQK
1451 LLDLVPVSGN TVVDVGQDEA TACSLKEMA LKIHVGVGAR MHHLSVCQWE
1501 VKLKLVSDBG ASGSWRVVT NVTGHTCTVD IYREVEDTES QKLVYHSTAL
1551 SSGPLHGVAL NTSYQPLSVI DLKRC SARNN KTTYCYDFPL TFEAAVQKSW
1601 SNISSENNQC YVKATELVFA EKNGSWGTP IIPMORAAGLN DIGMVAWILD
1651 MSTPEFPSGR QIIVIANIT FRAGSFGPRE DAFPEAVTNL ACEKKLPLIY
1701 LAANSGARIG IAEVKSCFR VGWTDSSPE RGFRIYMTD EDHDRIGSSV
1751 IAHKMLDSG EIRWVIDSVV GKEDGLGVEN IHGSAIASA YSRAYEETFT
1801 LTFVTGRTVG IGAYLARLGI RCIQRIDQPI ILTGFSALNK LLGREVYSSH
1851 MQLGGPKIMA TNGVVHLTVP DDLEGVSNIL RWLSYVPANI GGPLPITKSL
1901 DPIDRPVAYI PENTCDPRAA ISGIDDSQGK WLGGMFDKDS FVETFEGWAK
1951 TVVTGRAKLG GIPVGVIAVE TQTMQLVPA DPGQPDHER SVPRAGQVWF
2001 PDSATKTAQA MLDFNREGLP LFILANWRGF SGGQDRLFEG ILQAGSTIVE
2051 NLRTYNQPAF VYIPKAAELR GGAWVIDSK INPDRIEYCA ERTAKGNVLE
2101 PQGLIEIKFR SEELKECMGR LDPOLIDLKA RLQGANGLS DGESLQKISIE
2151 ARKKQLPLY TQIAVRFAEL HDTSLRMAAK GVIRKVDWE DSRFFYKRL
2201 RRRLSQVLA KEIRGVIGEK FPHKSAIELI KKWYLASEAA AAGSTDWDDD
2251 DAFVAWRENP ENYKEYIKEL RAQRVSRLLS DVAGSSSDLQ ALPQGLSMLL
2301 DKMDPSKRAQ FIEEVMKVLK

```

FIG.26B

## Oryza sativa plastidal ACCase Os05g22940 rice cDNA(Cypress)

1 ATGACATCCA CACATGTGGC GACATTGGGA GTTGGTGCCC AGGCACCTCC TCGTCACCAG  
 61 AAAAAAGTCAG CTGGCACTGC ATTTGTATCA TCTGGGTCAT CAAGACCCTC ATACCGAAAG  
 121 AATGGTCAGC GTECTCGGTC ACTTAGGGAA GAAAGCAATG GAGGAGTGTC TGATTCCAAA  
 181 AAGCTTAACC ACTCTATTCG CCAAGGTCTT GCTGGCATCA TTGACCTCCC AAATGACGCA  
 241 GCTTCAGAAG TTGATATTTT ACATGGTTCC GAAGATCCCA GGGGGCCTAC GGTCCCAGGT  
 301 TCCTACCAAA TGAATGGGAT TATCAATGAA ACACATAATG GGAGGCATGC TTCAGTCTCC  
 361 AAGGTTGTTG AGTTTTGTAC GGCACCTGGT GGCAAAACAC CAATTCACAG TGTATTAGTG  
 421 GCCAACAATG GAATGGCAGC AGCTAAGTTC ATGCGGAGTG TCCGAACATG GGCTAATGAT  
 481 ACTTTTGGAT CAGAGAAGGC AATTCAGCTG ATAGCTATGG CAACTCCGGA GGATCTGAGG  
 541 ATAAATGCAG AGCACATCAG AATTGCCGAT CAATTTGTAG AGGTACCTGG TGGAACAAAC  
 601 AACAACTACT ATGCAAATGT CCAACTCATA GTGGAGATAG CAGAGAGAAC AGGTGTTTCT  
 661 GCTGTTTGGC CTGGTTGGGG TCATGCATCT GAGAATCCTG AACTTCCAGA TGCCTGACT  
 721 GCAAAAGGAA TTGTTTTTCT TGGGCCACCA GCATCATCAA TGCATGCATT AGGAGACAAG  
 781 GTTGGCTCAG CTCTCATTGC TCAAGCAGCT GGAGTTCCAA CACTTGCTTG GAGTGGATCA  
 841 CATGTGGAAG TTCCTCTGGA GTGTTGCTTG GACTCAATAC CTGATGAGAT GTATAGAAAA  
 901 GCTTGTGTTA CTACCACAGA GGAAGCAGTT GCAAGTTGTC AGGTGGTTGG TTATCCTGCC  
 961 ATGATTAAGG CATCTTGGGG TGGTGGTGGT AAAGGAATAA GGAAGGTTCA TAATGATGAT  
 1021 GAGGTTAGGA CATTATTTAA GCAAGTTCAA GGCGAAGTAC CTGGTTCCCC AATATTTATC  
 1081 ATGAGGCTAG CTGCTCAGAG TCGACATCTT GAAGTTCAGT TGCTTTGTGA TCAATATGGC  
 1141 AACGTAGCAG CACTTCACAG TCGAGATTGC AGTGTACAAC GGCACACCA AAAGATAATC  
 1201 GAGGAAGGAC CAGTTACTGT TGCTCCTCGT GAGACTGTGA AAGAGCTTGA GCAGGCAGCA  
 1261 CGGAGGCTTG CTAAGCTGT GGGTTATGTT GGTGCTGCTA CTGTTGAATA CCTTTACAGC  
 1321 ATGGAACTG GTGAATATTA TTTTCTGGAA CTTAATCCAC GGCTACAGGT TGAGCATCCT  
 1381 GTCAGTGAAT GGATAGCTGA AGTAAATTTG CCTGCGGCTC AAGTTGCTGT TGGAAATGGGT  
 1441 ATACCCCTTT GGCAGATTCC AGAGATCAGG CGCTTCTACG GAATGAACCA TGGAGGAGGC  
 1501 TATGACCTTT GGAGGAAAAC AGCAGCTCTA GCGACTCCAT TTAACCTTGA TGAAGTAGAT  
 1561 TCTAAATGGC CAAAAGGCCA CTGCGTAGCT GTTAGAATAA CTAGCGAGGA TCCAGATGAT  
 1621 GGGTTTAAGC CTAAGCTGT AAAAGTAAAG GAGATAAGTT TCAAGAGTAA ACCAAATGTT  
 1681 TGGCCTATT TCTCAGTAAA GTCTGGTGGG GGCATCCATG AATTCGCTGA TTCTCAGTTC  
 1741 GGACATGTTT TTGCGTATGG AACTACTAGA TCGGCAGCAA TAACTACCAT GGCTCTTGCA  
 1801 CTAAAAGAGG TTCAAATTCG TGGAGAAATT CATTCAAACG TAGACTACAC AGTTGACCTA  
 1861 TTAAATGCCT CAGATTTTGT AGAAAATAAG ATTCATACTG GTTGGCTGGA TACCAGGATA  
 1921 GCCATGCGTG TTCAAGCTGA GAGGCCTCCA TGGTATATTT CAGTCGTTGG AGGGGCTTTA  
 1981 TATAAAACAG TAACTGCCAA CACGGCCACT GTTTCTGATT ATGTTGGTTA TCTTACCAAG  
 2041 GGCCAGATTC CACCAAAGCA TATATCCCTT GTCTATACGA CTGTTGCTTT GAATATAGAT  
 2101 GGGAAAAAAT ATACAATCGA TACTGTGAGG AGTGGACATG GTAGCTACAG ATTGCGAATG  
 2161 AATGGATCAA CGGTTGACGC AAATGTACAA ATATTATGTG ATGGTGGGCT TTTAATGCAG  
 2221 CTGGATGGAA ACAGCCATGT AATTTATGCT GAAGAAGAGG CCAGTGGTAC ACGACTTCTT  
 2281 ATTGATGGAA AGACATGCAT GTTACAGAAAT GACCATGACC CATCAAAGTT ATTAGCTGAG  
 2341 ACACCATGCA AACTTCTTCG TTTCTTGGTT GCTGATGGTG CTCATGTTGA TGCTGATGTA

FIG.26C-1

2401 CCATATGCGG AAGTTGAGGT TATGAAGATG TGCATGCCCC TCTTATCACC CGCTTCTGGT  
 2461 GTCATACATG TTGTAATGTC TGAGGGCCAA GCAATGCAGG CTGGTGATCT TATAGCTAGG  
 2521 CTGGATCTTG ATGACCCTTC TGCTGTAAAG AGAGCTGAGC CGTTCGAAGA TACTTTTCCA  
 2581 CAAATGGGTC TCCCTATTGC TGCTTCTGGC CAAGTTCACA AATTATGTGC TGCAAGTCTG  
 2641 AATGCTTGTC GAATGATCCT TCGGGGTAT GAGCATGATA TTGACAAGGT TGTGCCAGAG  
 2701 TTGGTATACT GCCTAGACAC TCCGGAGCTT CCTTTCCTGC AGTGGGAGGA GCTTATGTCT  
 2761 GTTTTAGCAA CTAGACTTCC AAGAAATCTT AAAAGTGAGT TGGAGGGCAA ATATGAGGAA  
 2821 TACAAAGTAA AATTTGACTC TGGGATAATC AATGATTTCC CTGCCAATAT GCTACGAGTG  
 2881 ATAATTGAGG AAAATCTTGC ATGTGGTTCT GAGAAGGAGA AGGCTACAAA TGAGAGGCTT  
 2941 GTTGAGCCTC TTATGAGCCT ACTGAAAGTCA TATGAGGGTG GGAGAGAAAG TCATGCTCAC  
 3001 TTTGTTGTCA AGTCCCTTTT TGAGGAGTAT CTCTATGTTG AAGAATTGTT CAGTGATGGA  
 3061 ATTCAGTCTG ATGTGATTGA GCGTCTGCGC CTTCAACATA GTAAAGACCT ACAGAAGGTC  
 3121 GTAGACATTG TGTTGTCCCA CCAGAGTGTT AGAAATAAAA CTAAGCTGAT ACTAAAACCT  
 3181 ATGGAGAGTC TGGTCTATCC AAATCCTGCT GCCTACAGGG ATCAATTGAT TCGCTTTTCT  
 3241 TCCCTTAATC ACAAAGCGTA TTACAAGTTG GCACTTAAAG CTAGTGAECT TCTTGAACAA  
 3301 ACAAAACTTA GTGAGCTCCG TGCAAGAATA GCAAGGAGCC TTTAGAGCT GGAGATGTTT  
 3361 ACTGAGGAAA GCAAGGGTCT CTCCATGCAT AAGCGAGAAA TTGCCATTAA GGAGAGCATG  
 3421 GAAGATTTAG TCACTGCTCC ACTGCCAGTT GAAGATGCGC TCATTTCTTT ATTTGATTGT  
 3481 AGTGATACAA CTGTTCAACA GAGAGTGATT GAGACTTATA TAGCTCGATT ATACCAGCCT  
 3541 CATCTTGTAAGGACAGTAT CAAAATGAAA TGGATAGAAT CGGGTGTTAT TGCTTTATGG  
 3601 GAATTTCTCG AAGGGCATT TTTGATGCAAGA AATGGAGGAG CGGTTCTTGG TGACAAAAGA  
 3661 TGGGGTGCCA TGGTCATTGT CAAGTCTCTT GAATCACTTT CAATGGCCAT TAGATTTGCA  
 3721 CTAAGGAGAG CATCACACTA CACTAGCTCT GAGGGCAATA TGATGCATAT TGCTTTGTTG  
 3781 GGTGCTGATA ATAAGATGCA TATAATTCAA GAAAGTGGTG ATGATGCTGA CAGAATAGCC  
 3841 AACTTCCCT TGATACTAAA GGATAATGTA ACCGATCTGC ATGCCTCTGG TGTGAAAACA  
 3901 ATAAGTTTCA TTGTTCAAAG AGATGAAGCA CGGATGACAA TCGCTCGTAC CTTCTTTTGG  
 3961 TCTGATGAAA AGCTTTCTTA TGAGGAAGAG CCAATTTCTCC GGCATGTGGA ACCTCCTCTT  
 4021 TCTGCACTTC TTGAGTTGGA CAAGTTGAAA GTGAAAGGAT ACAATGAAAT GAAGTATACC  
 4081 CCATCACGGG ATCGTCAATG GCATATCTAC AACTTAGAA ATACTGAAAA CCCCAAAATG  
 4141 TTGCACCGGG TATTTTTCCG AACCTTGTG AGGCAACCCA GTGTATCCAA CAAGTTTTCT  
 4201 TCGGGCCAGA TTGGTGACAT GGAAGTTGGG AGTGCTGAAG AACCTCTGTC ATTTACATCA  
 4261 ACCAGCATAT TAAGATCTTT GATGACTGCT ATAGAGGAAT TGGAGCTTCA CGCAATTAGA  
 4321 ACTGGCCATT CACACATGTA TTTGCATGTA TTGAAAGAAC AAAAGCTTCT TGATCTTGTT  
 4381 CCAGTTTCAG GGAATACAGT TTTGGATGTT GGTCAAGATG AAGCTACTGC ATATCACTT  
 4441 TTAAGGAAA TGGCTATGAA GATACATGAA CTTGTTGGTG CAAGATGCA CCATCTTTCT  
 4501 GTATGCCAAT GGAAGTGAA ACTTAAGTTG GACTGCGATG GTCCTGCCAG TGGTACCTGG  
 4561 AGGATTGTAA CAACCAATGT TACTAGTCAC ACTTGCACTG TGGATATCTA CCGTGAGATG  
 4621 GAAGATAAAG AATCACGGAA GTTAGTATAC CATCCCGCCA CTCCGGCGGC TGGTCTCTG  
 4681 CATGGTGTGG CACTGAATAA TCCATATCAG CCTTTGAGTG TCATTGATCT CAAACGCTGT  
 4741 TCTGCTAGGA ATAATAGAAC TACATACTGC TATGATTTTC CACTGGCATT TGAAACTGCA  
 4801 GTGAGGAAGT CATGGTCCTC TAGTACCTCT GGTGCTTCTA AAGGTGTTGA AAATGCCCAA  
 4861 TGTTATGTTA AAGCTACAGA GTTGGTATTT GCGGACAAAC ATGGGTGATG GGGCACTCCT

FIG.26C-2

4921 TTAGTTCAAA TGGACCGGCC TGCTGGGCTC AATGACATTG GTATGGTAGC TTGGACCTTG  
 4981 AAGATGTCCA CTCCTGAATT TCCTAGTGGT AGGGAGATTA TTGTTGTTGC AAATGATATT  
 5041 ACGTTCAGAG CTGGATCATT TGGCCCAAGG GAAGATGCAT TTTTTGAAGC TGTTACCAAC  
 5101 CTAGCCTGTG AGAAGAAACT TCCTCTTATT TATTTGGCAG CAAATTCTGG TGCTCGAATT  
 5161 GGCATAGCAG ATGAAGTGAA ATCTTGCTTC CGTGTGGGGT GGTCTGATGA TGGCAGCCCT  
 5221 GAACGTGGGT TTCAGTACAT TTATCTAAGC GAAGAAGACT ATGCTCGTAT TGGCACTTCT  
 5281 GTCATAGCAC ATAAGATGCA GCTAGACAGT GGTGAAATTA GGTGGGTTAT TGATTCTGTT  
 5341 GTGGGCAAGG AAGATGGACT TGGTGTGGAG AATATACATG GAAGTGCTGC TATTGCCAGT  
 5401 GCTTATTCTA GGGCATATAA GGAGACATTT ACACTTACAT TTGTGACTGG AAGAACTGTT  
 5461 GGAATAGGAG CTTATCTTGC TCGACTTGGC ATCCGGTGCA TACAGCGTCT TGACCAGCCT  
 5521 ATTATTCTTA CAGGCTATTC TGCACTGAAC AAGCTTCTTG GGCGGGAAGT GTACAGCTCC  
 5581 CACATGCAGT TGGGTGGTCC CAAAATCATG GCAACTAATG GTGTTGTCCA TCTTACTGTT  
 5641 TCAGATGACC TTGAAGGCGT TTCTAATATA TTGAGGTGGC TCAGTTATGT TCCTGCCTAC  
 5701 ATTGGTGGAC CACTTCCAGT AACAACACCG TTGGACCCAC CGGACAGACC TGTTGCATAC  
 5761 ATTCCTGAGA ACTCGTGTGA TCCTCGAGCG GCTATCCGTG GTGTTGATGA CAGCCAAGGG  
 5821 AAATGGTTAG GTGGTATGTT TGATAAAGAC AGCTTTGTGG AAACATTTGA AGGTTGGGCT  
 5881 AAGACAGTGG TFACTGGCAG AGCAAAGCTT GGTGGAATTC CAGTGGGTGT GATAGCTGTG  
 5941 GAGACTCAGA CCATGATGCA AACTATCCCT GCTGACCCTG GTCAGCTTGA TTCCCGTGAG  
 6001 CAATCTGTTC CTCGTGCTGG ACAAGTGTGG TTTCCAGATT CTGCAACCAA GACTGCGCAG  
 6061 GCATTGCTGG ACTTCAACCG TGAAGGATTA CCTCTGTTCA TCCTCGCTAA CTGGAGAGGC  
 6121 TTCTCTGGTG GACAAAGAGA TCTTTTTGAA GGAATTCTTC AGGCTGGCTC GACTATTGTT  
 6181 GAGAACCTTA GGACATACAA TCAGCCTGCC TTTGTCTACA TTCCCATGGC TGCAGAGCTA  
 6241 CGAGGAGGGG CTTGGGTTGT GGTTGATAGC AAGATAAACC CAGACCGCAT TGAGTGCTAT  
 6301 GCTGAGAGGA CTGCAAAAGG CAATGTTCTG GAACCGCAAG GGTTAATTGA GATCAAGTTC  
 6361 AGGTCAGAGG AACTCCAGGA TTGCATGAGT CGGCTTGACC CAACATTAAT TGATCTGAAA  
 6421 GCAAAACTCG AAGTAGCAAA TAAAAATGGA AGTGCTGACA CAAAATCGCT TCAAGAAAAT  
 6481 ATAGAAGCTC GAACAAAACA GTTGATGCCT CTATATACTC AGATTGCGAT ACGGTTTGCT  
 6541 GAATTGCATG ATACATCCCT CAGAATGGCT GCGAAAGGTG TGATTAAGAA AGTTGTGGAC  
 6601 TGGGAAGAAT CACGATCTTT CTTCTATAAG AGATTACGGA GGAGGATCTC TGAGGATGTT  
 6661 CTTGCAAAAAG AAATTAGAGC TGTAGCAGGT GAGCAGTTTT CCCACCAACC AGCAATCGAG  
 6721 CTGATCAAGA AATGGTATTC AGCTTCACAT GCAGCTGAAT GGGATGATGA CGATGCTTTT  
 6781 GTTGCTTGGG TGGATAACCC TGAAAACTAC AAGGATTATA TTCAATATCT TAAGGCTCAA  
 6841 AGAGTATCCC AATCCCTCTC AAGTCTTTCA GATTCCAGCT CAGATTTGCA AGCCCTGCCA  
 6901 CAGGGTCTTT CCATGTTACT AGATAAGATG GATCCCTCTA GAAGAGCTCA ACTTGTTGAA  
 6961 GAAATCAGGA AGGTCCTTGG TTGAATCATA TGATGCCAAA ACTATTATTG GAGGCACAAA  
 7021 TAGCTTGTGG ACCCTGTCCG ATTGTTGGTG AGTGTATATT GGATTTGTTA GTTCTGCCAG  
 7081 ATGAAAGTGC AAGTCTGATG ATTCATGATA CCGTCAGTTG GCAAGAACAC CGGTTAACCT  
 7141 GAGTGCTTGT TTACAAATGG TCCTTTATGA CAATCGTTGT TTCGCGCTAG TTCCGTGATC  
 7201 TACTATCATC TGTTAGACGC TGTAATTAGT GAGTCTCCGC GGATCCACAG TATACGGTTG  
 7261 AGCTGTTGAT TCAATTTTGG ACACGAATAA TATGATTTTG TAGGCATAAA TGCCTCTGTA  
 7321 TGTGAAATAA ATTGTCTGTT GAGTTAACAC ACAAGATGAC AATATGTTTG TGCTCTACTG  
 7381 CTATTGTCCA TGAATACTGA TTGCGGAATC AACCACATGC ATTAT

## FIG.26C-3

*Oryza sativa* plastidal ACCase Os05g22940 rice protein (Cypress)

1	MTSTHVATLG	VGAQAPPRHQ	KKSAGTAFVS	SGSSRPSYRK	NGQRTRSLRE
51	ESNGGVSDSK	KLNHSIROQL	AGIIDLPNDA	ASEVDISHGS	EDPRGPTVPG
101	SYQMNGIINE	THNGRHASVS	KVVEFCTALG	GKTPIHSLV	ANNGMAAAKF
151	MRSVRTWAND	TFGSEKAIQL	IAMATPEDLR	INAEHIRIAD	QFVEVPGGTN
201	NNNYANVQLI	VEIAERTGVS	AVWPGWGHAS	ENPELDPALT	AKGIVFLGPP
251	ASSMHALGDK	VGSALIAQAA	GVPTLAWSGS	HVEVPLECCL	DSIPDEMYRK
301	ACVTTTTEEAV	ASCQVVGYP	MIKASWGGGG	KGIRKVHND	EVRTLFLKQVQ
351	GEVPGSPIFI	MRLAAQSRHL	EVQLLCDQYG	NVAALHSRDC	SVQRRHQKII
401	EEGPVTVAPR	ETVKELEQAA	RRLAKAVGYV	GAATVEYLYS	METGEYFFLE
451	LNPRLOVEHP	VTEWIAEVNL	PAAQVAVGMG	IPLWQIPEIR	RFYGMNHGGG
501	YDLWRKTAAL	ATPFNFDEVD	SKWPKGHCVA	VRITSEDPDD	GFKPTGGKVK
551	EISFKSKPNV	WAYFSVKS	GIHEFADSQF	GHVFAYGTR	SAAITTMALA
601	LKEVQIRGEI	HSNVDYTVDL	LNASDFRENK	IHTGWLDTRI	AMRVQAERPP
651	WYISVVGAL	YKTVTANTAT	VSDYVGYLTK	GQIPPKHISL	VYTTVALNID
701	GKKYTIDTVR	SGHGSYRLRM	NGSTVDANVQ	ILCDGGLLMQ	LDGNSHVIYA
751	EEEASGTRLL	IDGKTCMLQN	DHDPSKLLAE	TPCKLLRFLV	ADGAHVADAV
801	PYAEVEVMKM	CMPLLSPASG	VIHVVMSEGO	AMQAGDLIAR	LDLDDPSAVK
851	RAEPFEDTFP	QMGLPIAASG	QVHKLCAASL	NACRMILAGY	EHDIDKVVPE
901	LVYCLDTPEL	PFLQWEELMS	VLATRLPRNL	KSELEGKYEE	YKVKFDSGII
951	NDFPANMLRV	IIEENLACGS	EKEKATNERL	VEPLMSLLKS	YEGGRESHAH
1001	FVVKSLFEEY	LYVEELFSDG	IQSDVIERLR	LQHSKDLQKV	VDIVLSHQSV
1051	RNKTKLILKL	MESLVYPNPA	AYRDQLIRFS	SLNHKAYYKL	ALKASELLEQ
1101	TKLSELRARI	ARSLSELEMF	TEESKGLSMH	KREIAIKESM	EDLVTAPLPV
1151	EDALISLFD	SDTTVQQRVI	ETYIARLYQP	HLVKDSIKMK	WIESGVIALW
1201	EFPEGHFDAR	NGGAVLGDKR	WGAMVIVKSL	ESLSMAIRFA	LKETSHTYSS
1251	EGNMMHIAL	GADNKMHI IQ	ESGDDADRIA	KLPLILKDNV	TDLHASGVKT
1301	ISFIVORDEA	RMTMRRTFLW	SDEKLSYEEE	PILRHVEPPL	SALLELDKLK
1351	VKGYNEMKYT	PSRDRWHIY	TLRNTENPKM	LHRVFFRTL	RQPSVSNKFS
1401	SGQIGDMEVG	SAEPLSFTS	TSILRSLMTA	IEEELHAIR	TGHSHMYLHV
1451	LKEQKLLDLV	PVSGNTVLDV	GQDEATAYSL	LKEMAMKTHE	LVGARMHHL
1501	VCQWEVKLKL	DCDGPASGTW	RIVTTNVTSH	TCTVDIYREM	EDKESRKL
1551	HPATPAAGPL	HGVALNPNYQ	PLSVIDLKRC	SARNNRTTYC	YDFPLAFETA
1601	VRKSWSSSTS	GASKGVENAQ	CYVKATELVF	ADKHGSWGTP	LVQMDRPA
1651	NDIGMVAWTL	KMSTPEFPSG	REIIVVANDI	TFRAGSFGPR	EDAFFEAVTN
1701	LACEKKLPLI	YLAANS	GIADEVKSCF	RVGWSDDGSP	ERGFQYIYLS
1751	EEDYARIGTS	VIAHKMLDS	GEIRWVIDSV	VGKEDGLGVE	NIHGSAAIAS
1801	AYSRAYKET	TLTFVTGRTV	GIGAYLARLG	IRCIQRDQP	IILTGY
1851	KLLGREVYSS	HMLG	ATNGVVHLTV	SDDLEGV	LRWLSYVPAY
1901	IGGPLPV	LDPPDRPVAY	IPENSCD	AIRGVDD	QWLGGMFDKD
1951	SFVETFE	KTVV	GGIPVGVIAV	ETQ	ADPGQLDSRE
2001	QSVPRAGQVW	FPDSATKTAQ	ALLDFNREGL	PLFILANWRG	FSGGQ
2051	GILQAGSTIV	ENLR	FVYIPMAAEL	RGGAWV	KINPDRIECY
2101	AERTAKGNVL	EPOGLIEIKF	RSEELQDCMS	RLDPTLIDLK	AKLEVANKNG
2151	SADTKSLQEN	IEARTKQLMP	LYTQIAIRFA	ELHDTSLRMA	AKGVIK
2201	WEESRSFFYK	RLRRRI	LAKEIRAVAG	EQFSHQPAIE	LIKKWYSASH
2251	AAEWDDDDAF	VAWMDNPENY	KDYIQYLK	RVSQSLSSLS	DSSSDLQALP
2301	OGLSMLLDKM	DPSRRAQLVE	EIRKVLG		

FIG. 26D

*Oryza sativa* plastidal ACCase Os05g22940 rice genomic DNA - AP008211  
reverse complemented fragment

```

1 TCATTCTTAT ATATTTTCAT CTGTCAGATT TCACACATCT GGGGATTTAT CTTCTCTTTG
61 TATGGCACTA CACATTTGAG AAACCGTGCA ATTCTACTGT TTGGTCAGCA GGACAACAAT
121 GACATCCACA CATGTGGCGA CATTGGGAGT TGGTGCCAG GCACCTCCTC GTCACCAGAA
181 AAAGTCAGCT GGCCTGCAT TTGTATCATC TGGGTCATCA AGACCCTCAT ACCGAAAGAA
241 TGGTCAGCGT ACTCGGTCAC TTAGGGAAGA AAGCAATGGA GGAGTGTCTG ATTCCAAAAA
301 GCTTAACCAC TCTATTCGCC AAGGTGACCA CTAGCTACTT TACATATGCT ATAATTTGTG
361 CCAAACATAA ACATGCAATG GCTGCTATTA TTTAAACGTT AATGTTGAAA TAGCTGCTAT
421 AGGATACAGC AAAAAATATAT AATTGACTGG GCAAGATGCA ACAATTGTTT TTRACTAAAG
481 TTAGTTATCT TTTGCTGTAA AAGACAACCTG TTTTTTACAT AAAATGGTAT TAATAACCTT
541 GTAATATTCA ATGCAACATG TTCTCAAGTA AAAAAAACA TTGCCTGGTT GTATAAGCAA
601 ATGTGTCGTT GTAGACATCT TATTAACCT TTTTGTGATA TCTATTACCG TAGGGAACAG
661 GGGAGCTGTT TAAATCTGTT ATCATAGAGT AATATGAGAA AAGTGGATTG TGCGACTTTG
721 GCATGTATAC CTGCTCAATT TCAAATATAT GTCTATGTGC AGGTCTTGCT GGCATCATTG
781 ACCTCCAAA TGACGCAGCT TCAGAAGTTG ATATTTTACA GTAAGGACTT TATATTTTAT
841 AATAATTATT ATATAATTTT CTGACATGTT TTGAGAACCT CAAAACATGT GATTGCACCT
901 TCCTTTTTTA TGTCTGGTTC AGAAACTGAT AAGTTTTGAC AGTGTTTAGG ATGGATCTTT
961 GATGCGCACA GTGCTTTCTA ATGTTTTCAT TTTTGAAAGT AATGTTTTAG GAAGAAATAT
1021 CTGATTAAT TTATACTTTA TCTTTACAAA AGTCAAATGC GTTCTGTATC AATTGCGGTT
1081 TGTAATATGG CAAGAACATG CTTTCAGAAAT TTGTTCATAA AATGCTTTCT TTCTATTATT
1141 ATGTAGAACA AATACCTAAT ACTTTGTTCA CCTTTTATAG TGGACACCTC TCACAGCTTT
1201 TTCAGTAAGT GATGCAATTT TGTACATTTG TAAGATGTGT TCCAGAAACC TTTTCTCCTG
1261 CAATTCTAAT GTACCCACTC AAAGTGGTAT CACCAAAGAT CTCCATCTGA TTGAAAAAAA
1321 GCTGCGTGAA GTATGCTTAT TTATGCTAAC CATAATGAT TTATACTGTT TTATAGTACA
1381 ATGCTTATTT ATGCTAACCA TACATAATTT TATTCTGTTT TCTAGTACAT TATTTGTGCC
1441 CCTGACCATA AATGATCCTT TCTTTTACAG TGGTCCGAA GATCCCAGGG GGCCTACGGT
1501 CCCAGGTTCC TACCAAATGA ATGGGATTAT CAATGAAACA CATAATGGGA GGCATGCTTC
1561 AGTCTCCAAG GTTGTGAGT TTTGTACGGC ACTTGGTGGC AAAACACCAA TTCACAGTGT
1621 ATTAGTGGCC AACAATGGAA TGGCAGCAGC TAAGTTCATG CGGAGTGTCC GAACATGGGC
1681 TAATGATACT TTTGGATCAG AGAAGGCAAT TCAGCTGATA GCTATGGCAA CTCCGGAGGA
1741 TCTGAGGATA AATGCAGAGC ACATCAGAAT TGCCGATCAA TTTGTAGAGG TACCTGGTGG
1801 AACAAACAAC AACAATATG CAAATGTCCA ACTCATAGTG GAGGTTAGTT CAGCTCATCC
1861 CTCAACACAA CATTTTCGTT TCTATTTAAG TTAGGGAAAA ATCTCTACGA CCCTCCAATT
1921 TCTGAACATC CAATTTTCAC CATCAACTGC AATCACAGAT AGCAGAGAGA ACAGGTGTTT
1981 CTGCTGTTTG GCCTGGTTGG GGTGATGCAT CTGAGAATCC TGAAC TTCCA GATGCGCTGA
2041 CTGCAAAAAGG AATTGTTTTT CTTGGGCCAC CAGCATCATC AATGCATGCA TTAGGAGACA
2101 AGGTTGGCTC AGCTCTCATT GCTCAAGCAG CTGGAGTTCC AACACTTGCT TGGAGTGGAT
2161 CACATGTGAG CCTTGTCTTC TCTTTTTTAG CTTATCATCT TATCTTTTCG GTGATGCATT
2221 ATCCCAATGA CACTAAACCA TAGGTGGAAG TTCCTCTGGA GTGTTGCTTG GACTCAATAC
2281 CTGATGAGAT GTATAGAAAA GCTTGTGTTA CTACCACAGA GGAAGCAGTT GCAAGTTGTC
2341 AGGTGGTTGG TTATCCTGCC ATGATTAAGG CATCTTGGGG TGGTGGTGGT AAAGGAATAA

```

**FIG.26E-1**

2401 GGAAGGTTTG TTCTTCTTGT AGTTATCAAG AGATTGTTTG GATTGCAAGT GTTTAGTGCC  
 2461 CATAGTTAAC TCTGGTCTTT CTAACATGAG TAACTCAACT TTCTTGCAGG TTCATAATGA  
 2521 TGATGAGGTT AGGACATTAT TTAAGCAAGT TCAAGGCGAA GTACCTGGTT CCCCAATATT  
 2581 TATCATGAGG CTAGCTGCTC AGGTGGGGCC TTTTATGGAA GTTACACCTT TTCCCTTAAT  
 2641 GTTGAGTTAT TCCGGAGTTA TTATGGTTAT GTTCTGTATG TTTGATCTGT AAATTATTGA  
 2701 AATTCACCTC CATTGGTTCT CCAGATTAGC AGACCTACAA TTCTACATAT GGTTTATACT  
 2761 TTATAAATAC TAGGATTTAG GGATCTTCAT ATAGTTTATA CATGGTATTT AGATTTCAAT  
 2821 TGTAACCTA TTGAAGACAT CCTGATTGTT GTCTTATGTA GAGTCGACAT CTTGAAGTTC  
 2881 AGTTGCTTTG TGATCAATAT GGCAACGTAG CAGCACTTCA CAGTCGAGAT TGCAGTGTAC  
 2941 AACGGCGACA CAAAAGGTC TGCTGTCTCA GTTAAATCAC CCCTCTGAAT GATCTACTTC  
 3001 TTGCCTGCTG CGTTGGTCAG AGGAATAATG GTTGTATTCT ACTGAACAGA TAATCGAGGA  
 3061 AGGACCAGTT ACTGTTGCTC CTCGTGAGAC TGTGAAAGAG CTTGAGCAGG CAGCACGGAG  
 3121 GCTTGCTAAA GCTGTGGGT ATGTTGGTGC TGCTACTGTT GAATACCTTT ACAGCATGGA  
 3181 AACTGGTGAA TATTATTTTC TGGAACTTAA TCCACGGCTA CAGGTCGGCT CCTTTGACAT  
 3241 TCTTCAGGAA TTAATTTCTG TTGACCACAT GATTTACATT GTCAAATGGT CTCACAGGTT  
 3301 GAGCATCCTG TCACTGAGTG GATAGCTGAA GTAAATTTGC CTGCGGCTCA AGTTGCTGTT  
 3361 GGAATGGGTA TACCCCTTTG GCAGATTCCA GGTAAATGCTT CTTCAATTTAG TTCCTGCTCT  
 3421 TTGTTAATTG AATGAGCTCT TATACAGACC ATGAGACACA TTCTACTGTT AATTCATAGT  
 3481 ATCCCTGAC TTGTTAGTGT TAGAGATACA GAGATGTATC ACAAATTCAT TGTATCTCCT  
 3541 CAAGGACTGT AAAAATCCTA TAATTAATTT TCTGAAAATT TGTTCTTTTA AGCAGAAAAA  
 3601 AAATCTCTAA ATTATCTCCC TGTATACAGA GATCAGGCGC TTCTACGGAA TGAACCATGG  
 3661 AGGAGGCTAT GACCTTTGGA GGAAAACAGC AGCTCTAGCG ACTCCATTTA ACTTTGATGA  
 3721 AGTAGATTCT AAATGGCCAA AAGGCCACTG CGTAGCTGTT AGAATAACTA GCGAGGATCC  
 3781 AGATGATGGG TTTAAGCCTA CTGGTGGAAA AGTAAAGGTG CGGTTTCTG ATGTTAGGTG  
 3841 TATGAATTGA ACACATTGCT ATATTGCAGC TAGTGAAATG ACTGGATCAT GGTTCTCTTA  
 3901 TTTTCAGGAG ATAAGTTTCA AGAGTAAACC AAATGTTTGG GCCTATTTCT CAGTAAAGGT  
 3961 AGTCCTCAAT ATTGTTGCAC TGCCACATTA TTTGAGTTGT CCTAACAATT GTGCTGCAAT  
 4021 TGTTAGTTTT CAACTATTTG TTGTTCTGTT TGGTTGACTG GTACCCTCTC TTTGAGTCT  
 4081 GGTGGAGGCA TCCATGAATT CGCTGATTCT CAGTTCGGTA TGTAAGTTA AAAGAGTAAT  
 4141 ATTGTCTTTG CTATTTATGT TTGTCCTCAC TTTTAAAAGA TATTGCCTTC CATTACAGGA  
 4201 CATGTTTTTG CGTATGGAAC TACTAGATCG GCAGCAATAA CTACCATGGC TCTTGCCTA  
 4261 AAAGAGGTTT AAATTCGTGG AGAAATTCAT TCAAACGTAG ACTACACAGT TGACCTATTA  
 4321 AATGTAAGGA CTAATATCT GCTTATTGAA CCTTGCTTTT TGGTCCCTA ATGCCATTTT  
 4381 AGTCTGGCTA CTGAAGAACT TATCCATCAT GCCATTTCTG TTATCTTAAA TTCAGGCCTC  
 4441 AGATTTTAGA GAAAATAAGA TTCATACTGG TTGGCTGGAT ACCAGGATAG CCATGCGTGT  
 4501 TCAAGCTGAG AGGCCTCCAT GGTATATTTT AGTCGTTGGA GGGGCTTTAT ATGTAAGACA  
 4561 AACTATGCCA CTCATTAGCA TTTATGTGAA GCAAATGCGG AAAACATGAT CAATATGTCG  
 4621 TCTTATTTAA ATTTATTTAT TTTTGTGCTG CAGAAAACAG TAACTGCCAA CACGGCCACT  
 4681 GTTTCTGATT ATGTTGGTTA TCTTACCAAG GGCCAGATTC CACCAAAGGT ACTATTCTGT  
 4741 TTTTTCAGGA TATGAATGCT GTTTGAATGT GAAAACCATT GACCATAAAT CCTTGTTTGC  
 4801 AGCATATATC CCTTGTCTAT ACGACTGTTG CTTTGAATAT AGATGGGAAA AAATATACAG  
 4861 TAAGTGTGAC ATTCTTAATG GGGAAACTTA ATTTGTTGTA AATAATCAAT ATCATATTGA

FIG.26E-2

SUBSTITUTE SHEET (RULE 26)

4921 CTCGTGTATG CTGCATCATA GATCGATACT GTGAGGAGTG GACATGGTAG CTACAGATTG  
 4981 CGAATGAATG GATCAACGGT TGACGCAAAT GTACAAATAT TATGTGATGG TGGGCTTTTA  
 5041 ATGCAGGTAA TATCTTCTTC CTAGTTAAAG AAGATATATC TTGTTCAAAG AATTCTGATT  
 5101 ATTGATCTTT TAATGTTTTT AGCTGGATGG AAACAGCCAT GTAATTTATG CTGAAGAAGA  
 5161 GGCCAGTGGT ACACGACTTC TTATTGATGG AAAGACATGC ATGTTACAGG TAATGATAGC  
 5221 CTTGTTCTTT TTAGTTCTAG TCACGGTGTT TGCTTGCTAT TTGTTGTATC TATTTAATGC  
 5281 ATTCACTAAT TACTATATTA GTTTGCATCA TCAAGTAAA ATGGAECTTC TTTCTTGCAG  
 5341 AATGACCATG ACCCATCAAA GTTATTAGCT GAGACACCAT GCAAECTTCT TCGTTTCTTG  
 5401 GTTGCTGATG GTGCTCATGT TGATGCTGAT GTACCATATG CGGAAGTTGA GGTTATGAAG  
 5461 ATGTGCATGC CCCTCTTATC ACCCGCTTCT GGTGTCATAC ATGTTGTAAT GTCTGAGGGC  
 5521 CAAGCAATGC AGGTACATTC CTACATTCCA TTCATTGTGC TGTGCTGACA TGAACATTTT  
 5581 AAGTAAATAC CTGTAECTTG TTTATTATTC TAGGCTGGTG ATCTTATAGC TAGGCTGGAT  
 5641 CTTGATGACC CTTCTGCTGT TAAGAGAGCT GAGCCGTTTC AAGATACTTT TCCACAAATG  
 5701 GGTCTCCCTA TTGCTGCTTC TGGCCAAGTT CACAAATTAT GTGCTGCAAG TCTGAATGCT  
 5761 TGTCGAATGA TCCTTGCGGG GTATGAGCAT GATATTGACA AGGTAACAT CATGTCCTCT  
 5821 TGTTTTTTCT TTTGTTTATC ATGCATTCTT ATGTTTATCA TGTCCTCTGG CAAATCTAGA  
 5881 TTCCGCTGTC GTTTCACACA GATTTTTCTC ATTCTCATAA TGGTGCCAAA CATAAATATG  
 5941 CTGCTATATT CATCAATGTT TTCACTCGAT TTCTAATTTT GCTTTTGAGT TTTAACTTTT  
 6001 AGTACAATCC ATATCTAATC TCCTTTGGCA ACAGTGAATC CATTATATAT ATTTTTATTA  
 6061 AACTGCTTTC TTTTTCAGGT TGTGCCAGAG TTGGTATACT GCCTAGACAC TCCGGAGCTT  
 6121 CCTTTCCTGC AGTGGGAGGA GCTTATGTCT GTTTTAGCAA CTAGACTTCC AAGAAATCTT  
 6181 AAAAGTGAGG TATATTATGG TTGACAAGAT AGCTAGTCTC ATGCTCTAAG GACTTGTACA  
 6241 TTTCCGACCA TAGGTTAATT TTCCATATCA AGTTCTAATG TACGATATAA AAGTAGTACT  
 6301 GGCCATAAAC AGTATTGGTG GTTGACTATC TTTGTTGTGT AAGATCAAGT ATTTCTTTTT  
 6361 CATGCTTAGT TTGTCAATAC TTCACATTTA TCACTGACTT GTCGAGCTAA ATGAGATTTT  
 6421 ATTTGATTTT TGTGCTCCAT TATTTTTGTA TATATATATA TATATTTAAC TATGACTATA  
 6481 TGTTATGCCT CAAACGTTTC AAACCTTTTC AGTTGGAGGG CAAATATGAG GAATACAAAG  
 6541 TAAAATTTGA CTCTGGGATA ATCAATGATT TCCCTGCCAA TATGCTACGA GTGATAATTG  
 6601 AGGTCAGTTA TTCAATTTGT TGTGATAATC ACTGCCTTAA CTGTTCTGTT TTTTAAACAAG  
 6661 CGGTTTTATA GGAAAATCTT GCATGTGGTT CTGAGAAGGA GAAGGCTACA AATGAGAGGC  
 6721 TTGTTGAGCC TCTTATGAGC CTACTGAAGT CATATGAGGG TGGGAGAGAA AGTCATGCTC  
 6781 ACTTTGTTGT CAAGTCCCTT TTTGAGGAGT ATCTCTATGT TGAAGAATTG TTCAGTGATG  
 6841 GAATTCAGGT TAACTTACCT ATTCGCATTA AACAAATCAT CAGTTGTTTT ATGATAAAGT  
 6901 CAAAATGTTT ATATTTCCCA TTCTTCTGTG GATCAAATAT ATCACGGACA TGATATAGTT  
 6961 TCCTTAGGCT ATATAATGGT TCTTCATCAA ATAATATTGC AGGAAACAGT ATAGCAAATC  
 7021 ATTTGTATAT ACTCGAGATG GAAATTTGTTA GAAACATCAT TGAATAAATC TGTCCTTTGT  
 7081 TACGCTGTTT TTGTAGTCTG ATGTGATTGA GCGTCTGCGC CTTCAACATA GTAAAGACCT  
 7141 ACAGAAGGTC GTAGACATTG TGTTGTCCCA CCAGGTAAAT TTCTTCATGG TCTGATGACT  
 7201 TCACTGCGAA TGGTTACTGA ACTGTCTTCT TGTTCTGACA ATGTGACTTT TCTTTGTAGA  
 7261 GTGTTAGAAA TAAAATAAG CTGATACTAA AACTCATGGA GAGTCTGGTC TATCCAAATC  
 7321 CTGCTGCCTA CAGGGATCAA TTGATTCGCT TTTCTTCCCT TAATCACAAA GCGTATTACA  
 7381 AGGTGACCAG GATAAACATA AATAAACGTG AATTTTTCAA TGACCTTTTC TTCTGACATC

FIG.26E-3

7441 TGAATCTGAT GAATTTCTTG CATATTAATA CAGTTGGCAC TTAAGCTAG TGAACTTCTT  
 7501 GAACAAACAA AACTTAGTGA GCTCCGTGCA AGAATAGCAA GGAGCCTTTC AGAGCTGGAG  
 7561 ATGTTTACTG AGGAAAGCAA GGGTCTCTCC ATGCATAAGC GAGAAATTGC CATTAAAGGAG  
 7621 AGCATGGAAG ATTTAGTCAC TGCTCCACTG CCAGTTGAAG ATGCGCTCAT TTCTTTATTT  
 7681 GATTGTAGTG ATACAACGTG TCAACAGAGA GTGATTGAGA CTTATATAGC TCGATTATAC  
 7741 CAGGTATGAG AAGAAAGACC TTTTGAAATT ATTTATATTA ACATATCCTA GTAAAACAGC  
 7801 ATGCTCATCA TTTCTTAAAA AAAGTTTACA GCACCTGATG TTTGGTACT GACCGCATCA  
 7861 TTAATAAAA GTTACTTGTT GTGGAGAGAT GTATTTTGGG ACTTGTGGCA CATGCAGTAA  
 7921 CATGCTACTG CTCGATATGT TTGCTAACTT GACAACAATA TTTTTCAGCC TCATCTTGTA  
 7981 AAGGACAGTA TCAAAATGAA ATGGATAGAA TCGGGTGTTA TTGCTTTATG GGAATTTCTT  
 8041 GAAGGGCATT TTGATGCAAG AAATGGAGGA GCGGTTCTTG GTGACAAAAG ATGGGGTGCC  
 8101 ATGGTCATTG TCAAGTCTCT TGAATCACTT TCAATGGCCA TTAGATTTGC ACTAAAGGAG  
 8161 ACATCACACT ACACTAGCTC TGAGGGCAAT ATGATGCATA TTGCTTTGTT GGGTGTGAT  
 8221 AATAAGATGC ATATAATTCA AGAAAGGTAT GTTCATATGC TATGTTGGTG CTGAAATAGT  
 8281 TATATATGTA GTTAGCTGGT GGAGTTCTGG TAATTAACCT ATCCCATTGT TCAGTGGTGA  
 8341 TGATGCTGAC AGAATAGCCA AACTTCCCTT GATACTAAAG GATAATGTAA CCGATCTGCA  
 8401 TGCCCTGGT GTGAAAACAA TAAGTTTCAT TGTTCAAAGA GATGAAGCAC GGATGACAAT  
 8461 GCGTCGTACC TTCCTTTGGT CTGATGAAAA GCTTTCTTAT GAGGAAGAGC CAATTCTCCG  
 8521 GCATGTGGAA CCTCCTCTTT CTGCACTTCT TGAGTTGGTA CGTGATATCA TCAAAATGAT  
 8581 AATGTTTTGG TATGGCATTG ATTATCTTCT ATGCTCTTTG TATTTATTCA GCCTATTGTG  
 8641 GATACAGGAC AAGTTGAAAAG TGAAAGGATA CAATGAAATG AAGTATACCC CATCACGGGA  
 8701 TCGTCAATGG CATATCTACA CACTTAGAAA TACTGAAAAC CCCAAAATGT TGCACCGGGT  
 8761 ATTTTTCCGA ACCCTTGTC AAGCAACCCAG TGTATCCAAC AAGTTTTCTT CGGGCCAGAT  
 8821 TGGTGACATG GAAGTTGGGA GTGCTGAAGA ACCTCTGTCA TTTACATCAA CCAGCATATT  
 8881 AAGATCTTTG ATGACTGCTA TAGAGGAATT GGAGCTTAC GCAATTAGAA CTGGCCATTC  
 8941 ACACATGTAT TTGCATGTAT TGAAAGAACA AAAGCTTCTT GATCTTGTT CAGTTTCAGG  
 9001 GTAAGTGCGC ATATTTCTTT TTGGGAACAT ATGCTTGCTT ATGAGGTTGG TCTTCTCAAT  
 9061 GATCTTCTTA TCTTACTCAG GAATACAGTT TTGGATGTTG GTCAAGATGA AGCTACTGCA  
 9121 TATTCACCTT TAAAAGAAAT GGCTATGAAG ATACATGAAC TTGTTGGTGC AAGAATGCAC  
 9181 CATCTTCTG TATGCCAATG GGAAGTGAAC CTTAAGTTGG ACTGCGATGG TCCTGCCAGT  
 9241 GGTACCTGGA GGATTGTAAC AACCAATGTT ACTAGTCACA CTTGCACTGT GGATGTAAGT  
 9301 TTAATCCTCT AGCATTITGT TTTCTTTGGA AAAGCATGTG ATTTAAGCC GGCTGGTCCT  
 9361 CATACCGAGA CCTAGTGATC TTTATATAGT GTAGACATTT TTCTAACTGC TTTTAATTGT  
 9421 TTTAGATCTA CCGTGAGATG GAAGATAAAG AATCACGGAA GTTAGTATAC CATCCCGCCA  
 9481 CTCCGGCGGC TGGTCCTCTG CATGGTGTGG CACTGAATAA TCCATATCAG CCTTTGAGTG  
 9541 TCATTGATCT CAAACGCTGT TCTGCTAGGA ATAATAGAAC TACATACTGC TATGATTTTC  
 9601 CACTGGTGAG TTGACTGCTC CCTTATATTC AATGCATTAC CATAGCAAAT TCATATTCGT  
 9661 TCATGTTGTC AAAATAAGCC GATGAAAATT CAAAACGTGA GGCATTTGAA ACTGCAGTGA  
 9721 GGAAGTCATG GTCCTCTAGT ACCTCTGGTG CTTCTAAAGG TGTTGAAAAT GCCCAATGTT  
 9781 ATGTTAAAGC TACAGAGTTG GTATTTGCGG ACAAACATGG GTCATGGGGC ACTCCTTTAG  
 9841 TTCAAATGGA CCGGCCTGCT GGGCTCAATG ACATTGGTAT GGTAGCTTGG ACCTTGAAGA  
 9901 TGTCCACTCC TGAATTTCTT AGTGGTAGGG AGATTATTGT TGTTGCAAAT GATATTACGT

FIG. 26E-4

9961 TCAGAGCTGG ATCATTTGGC CCAAGGGAAG ATGCATTTTT TGAAGCTGTT ACCAACCTAG  
 10021 CCTGTGAGAA GAAACTTCCT CTTATTTATT TGGCAGCAA TTCTGGTGCT CGAATTGGCA  
 10081 TAGCAGATGA AGTGAAATCT TGCTTCCGTG TTGGGTGGTC TGATGATGGC AGCCCTGAAC  
 10141 GTGGGTTTCA GTACATTTAT CTAAGCGAAG AAGACTATGC TCGTATTGGC ACTTCTGTCA  
 10201 TAGCACATAA GATGCAGCTA GACAGTGGTG AAATTAGGTG GGTATTGAT TCTGTTGTGG  
 10261 GCAAGGAAGA TGGACTTGGT GTGGAGAATA TACATGGAAG TGCTGCTATT GCCAGTGCTT  
 10321 ATTCTAGGGC ATATAAGGAG ACATTTACAC TTACATTTGT GACTGGAAGA ACTGTTGGAA  
 10381 TAGGAGCTTA TCTTGCTCGA CTTGGCATCC GGTGCATACA GCCTCTTGAC CAGCCTATTA  
 10441 TTCTTACAGG CTATTCTGCA CTGAACAAGC TTCTTGGGCG GGAAGTGTAC AGCTCCCACA  
 10501 TGCAGTTGGG TGGTCCCAA ATCATGGCAA CTAATGGTGT TGTCCATCTT ACTGTTTCAG  
 10561 ATGACCTTGA AGGCGTTTCT AATATAATTGA GGTGGCTCAG TTATGTTTCT GCCTACATTG  
 10621 GTGGACCACT TCCAGTAACA ACACCGTTGG ACCCACCGBA CAGACCTGTT GCATACATTC  
 10681 CTGAGAACTC GTGTGATCCT CGAGCGGCTA TCCGTGGTGT TGATGACAGC CAAGGGAAAT  
 10741 GGTTAGGTGG TATGTTTGAT AAAGACAGCT TTGTGGAAC ATTTGAAGGT TGGGCTAAGA  
 10801 CAGTGGTTAC TGGCAGAGCA AAGCTTGGTG GAATTCAGT GGGTGTGATA GCTGTGGAGA  
 10861 CTCAGACCAT GATGCAAAC ATCCCTGCTG ACCCTGGTCA GCTTGATTCC CGTGAGCAAT  
 10921 CTGTTCTCG TGCTGGACAA GTGTGGTTTC CAGATTCTGC AACCAAGACT GCGCAGGCAT  
 10981 TGCTGGACTT CAACCGTGAA GGATTACCTC TGTTATCCT CGCTAACTGG AGAGGCTTCT  
 11041 CTGGTGGACA AAGAGATCTT TTTGAAGGAA TTCTTCAGGC TGGCTCGACT ATTTGTTGAGA  
 11101 ACCTTAGGAC ATACAATCAG CCTGCCTTTG TCTACATTCC CATGGCTGCA GAGCTACGAG  
 11161 GAGGGGCTTG GGTGTGGTT GATAGCAAGA TAAACCCAGA CCGCATTGAG TGCTATGCTG  
 11221 AGAGGACTGC AAAAGGCAAT GTTCTGGAAC CGCAAGGGTT AATTGAGATC AAGTTCAGGT  
 11281 CAGAGGAACT CCAGGATTGC ATGAGTCGGC TTGACCCAAC ATTAATTGAT CTGAAAGCAA  
 11341 AACTCGAAGT AGCAAATAAA AATGGAAGTG CTGACACAAA ATCGTTCAA GAAAATATAG  
 11401 AAGCTCGAAC AAAACAGTTG ATGCCTCTAT AACTCAGAT TGCATACGG TTTGCTGAAT  
 11461 TGCATGATAC ATCCCTCAGA ATGGCTGCGA AAGGTGTGAT TAAGAAAGTT GTGGACTGGG  
 11521 AAGAATCACG ATCTTTCTTC TATAAGAGAT TACGGAGGAG GATCTCTGAG GATGTTCTTG  
 11581 CAAAAGAAAT TAGAGCTGTA GCAGGTGAGC AGTTTTCCA CCAACCAGCA ATCGAGCTGA  
 11641 TCAAGAAATG GTATTGAGCT TCACATGCAG CTGAATGGGA TGATGACGAT GCTTTTGTG  
 11701 CTTGGATGGA TAACCTGAA AACTACAAGG ATTATATTCA ATATCTTAAG GCTCAAAGAG  
 11761 TATCCCAATC CCTCTCAAGT CTTTCAGATT CCAGCTCAGA TTTGCAAGCC CTGCCACAGG  
 11821 GTCTTTCCAT GTTACTAGAT AAGGTAATTA GCTTACTGAT GCTTATATAA ATCTTTTTTC  
 11881 ATTACATATG GCTGGAGAAC TATCTAATCA AATAATGATT ATAATTCAA TCGTTCTTTT  
 11941 TATGCCATTA TGATCTTCTG AAATTTCCCT CTTTGGACAC TTATTCAGAT GGATCCCTCT  
 12001 AGAAGAGCTC AACTTGTTGA AGAAATCAGG AAGGTCCTTG GTTGAATCAT ATGATGCCAA  
 12061 AACTATTATT GGAGGCACAA ATAGCTTGTG GACCCTGTCG GATTGTTGGT GAGTGTATAT  
 12121 TGGATTTGTT AGTTCTGCCA GATGAAAGTG CAAGTCTGAT GATTCATGAT ACCGTCAGTT  
 12181 GGCAAGAACA CCGGTTAACC TGAGTGCTTG TTTACAAATG GTCCTTTATG ACAATCGTTG  
 12241 TTTCCGCTA GTTCCGTGAT CTAATATCAT CTGTTAGACG CTGTAATTAG TGAGTCTCCG  
 12301 CGGATCCACA GTATACGTT GAGCTGTTGA TTCAATTTTG GACACGAATA ATATGATTTT  
 12361 GTAGGCATAA ATGCGTCTGT ATGTGAAATA AATTGTCTGT TGAGTTAACA CACAAGATGA  
 12421 CAATATGTTT GTGCTCTACT GCTATTGTCC ATGAATACTG ATTGCGGAAT CAACCACATG  
 12481 CATTAT

FIG. 26E-5

EAZ33685 - *Oryza sativa* plastidal ACCase Os05g22940 Japonica rice translated protein

1	MTSTHVATLG	VGAQAPPRHQ	KKSAGTAFVS	SGSSRPSYRK	NGQRTRSLRE
51	ESNGGVSDSK	KLNHSIRQGL	AGIIDLPNDA	ASEVDISHGS	EDPRGPTVPG
101	SYQMNGIINE	THNGRHASVS	KVVEFCTALG	GKTPIHSLV	ANNGMAAAKF
151	MRSVRTWAND	TFGSEKAIQL	IAMATPEDLR	INAEHIRIAD	QFVEVPGGTN
201	NNNYANVQLI	VEIAERTGVS	AVWPGWGHAS	ENPELDPALT	AKGIVFLGPP
251	ASSMHALGDK	VGSALIAQAA	GVPTLAWSGS	HVEVPLECCL	DSIPDEMYRK
301	ACVTTTTEEAV	ASCQVVGYP	MIKASWGGGG	KGIRKVHND	EVRTLKQVQ
351	GEVPGSPIFI	MRLAAQSRHL	EVQLLCDQYG	NVAALHSRDC	SVQRRHQKII
401	EEGPVTVAPR	ETVKELEQAA	RRLAKAVGYV	GAATVEYLYS	METGEYFFLE
451	LNPRLQVEHP	VTEWIAEVNL	PAAQVAVGMG	IPLWQIPEIR	RFYGMNHGGG
501	YDLWRKTAAL	ATPFNFDEVD	SKWPKGHCVA	VRITSEDPDD	GFKPTGGKVK
551	EISFKSKPNV	WAYFSVKS	GIHEFADSQF	GHVFAYGTR	SAAITTMALA
601	LKEVQIRGEI	HSNVDYTVDL	LNASDFRENK	IHTGWLDTRI	AMRVQAERPP
651	WYISVVGAL	YKTVTANTAT	VSDYVGYLTK	GQIPPKHISL	VYTTVALNID
701	GKKYTIDTVR	SGHGSRRLRM	NGSTVDANVQ	ILCDGGLLMQ	LDGNSHVIYA
751	EEEASGTRLL	IDGKTCMLQN	DHDPSKLLAE	TPCKLLRFLV	ADGAHVADAV
801	PYAEVEVMKM	CMPLLSPASG	VIHVVMSEGO	AMQAGDLIAR	LDLDDPSAVK
851	RAEPFEDTFP	QMGLPIAASG	QVHKLCAASL	NACRMILAGY	EHDIDKVVPE
901	LVYCLDTPEL	PFLQWEELMS	VLATRLPRNL	KSELEGKYEE	YKVKFDSGII
951	NDFPANMLRV	IIEENLACGS	EKEKATNERL	VEPLMSLLKS	YEGGRESHAH
1001	FVVKSLFEEY	LYVEELFSDG	IQSDVIERLR	LQHSKDLOKV	VDIVLSHQSV
1051	RNKTKLILKL	MESLVYPNPA	AYRDQLIRFS	SLNHKAYYKL	ALKASELLEQ
1101	TKLSELRARI	ARSLSELEMF	TEESKGLSMH	KREIAIKESM	EDLVTAPLPV
1151	EDALISLFDC	SDTTVQQRVI	ETYIARLYQP	HLVKDSIKMK	WIESGVIALW
1201	EFPEGHFDAR	NGGAVLGDKR	WGAMVIVKSL	ESLSMAIRFA	LKETSHTYSS
1251	EGNMMHIALL	GADNMHIIQ	ESGDDADRIA	KLPLILKDNV	TDLHASGVKT
1301	ISFIVQRDEA	RMTMRRTFLW	SDEKLSYEEE	PILRHVEPPL	SALLELDKLK
1351	VKGYNEMKYT	PSRDRQWHIY	TLRNTENPKM	LHRVFFRTL	RQPSVSNKFS
1401	SGQIGDMEVG	SAEPLSFTS	TSILRSLMTA	IEEELHAIR	TGHSHMYLHV
1451	LKEQKLLDLV	PVSGNTVLDV	GQDEATAYSL	LKEMAMKIH	LVGARMHLS
1501	VCQWEVKLKL	DCDGPASGTW	RIVTTNVTSH	TCTVDIYREM	EDKESRKL
1551	HPATPAAGPL	HGVALNPNYQ	PLSVIDLKRC	SARNNRTTYC	YDFPLAFETA
1601	VRKSWSSSTS	GASKGVENAO	CYVKATELVQ	MDRLAGLNDI	GMVAWTLKMS
1651	TPEFLSGREI	IVVANDITFR	AGSFGPREDA	FFEAVTNLAC	EKKLPLIYLA
1701	ANSGARIGIA	DEVKSCFRVG	WSDDGSPERG	FQYIYLSEED	YARIGTSVIA
1751	HKMQLDSGEI	RWVIDSVVGK	EDGLGVENIH	GSAAIASAYS	RAYKETFTLT
1801	FVTGRTVGIG	AYLARLGIRC	IQRLDQPIIL	TGYSALNKLL	GREVYSSHM
1851	LGGPKIMATN	GVVHLTVSDD	LEGVSNILRW	LSYVPAYIGG	PLPVTPLDP
1901	PDRPVAYIPE	NSCDPRAAIR	GVDDSQGKWL	GGMFDKDSFV	ETFEGWAKTV
1951	VTGRAKGGI	PVGVIIVETQ	TMMQTIPADP	GQLDSREQSV	PRAGQVWFPD
2001	SATKTAQALL	DFNREGLPLF	ILANWRGFSG	GQRDLFEGIL	QAGSTIVENL
2051	RTYNQPAFVY	IPMAAELRGG	AWVVVDSKIN	PDRIEYAEER	TAKGNVLEPQ
2101	GLIEIKFRSE	ELQDCMSRLD	PTLIDLKAKL	EVANKNGSAD	TKSLQENIEA
2151	RTKQLMPLYT	QIAIRFAELH	DTSLRMAAKG	VIKKVVDWEE	SRSFFYKRLR
2201	RRISEDVLAK	EIRAVAGEQF	SHQPAIELIK	KWYSASHAAE	WDDDDAFVAW
2251	MDNPENYKDY	IQYLKAQRVS	QSLSSLS	SDLQALPQGL	SMLLDKMDPS
2301	RRAQLVEEIR	KVLG			

FIG. 26F

B9FK36 - *Oryza sativa* ACCase protein

10	20	30	40	50	60
MTSTHVATLG	VGAQAPPRHQ	KKSAGTAFVS	SGSSRPSYRK	NGQRTRSLRE	ESNGGVSDSK
70	80	90	100	110	120
KLNHSIRQGL	AGIIDLPNDA	ASEVDISHGS	EDPRGPTVPG	SYQMNGIINE	THNGRHASVS
130	140	150	160	170	180
KVVEFCTALG	GKTPIHSVLV	ANNGMAAAKF	MRSVRTWAND	TFGSEKAIQL	IAMATPEDLR
190	200	210	220	230	240
INAEHIRIAD	QFVEVPGGTN	NNNYANVQLI	VEIAERTGVS	AVWPGWGHAS	ENPELDPALT
250	260	270	280	290	300
AKGIVFLGPP	ASSMHALGDK	VGSALIAQAA	GVPTLAWSGS	HVEVPLECCL	DSIPDEMYRK
310	320	330	340	350	360
ACVTTTTEEAV	ASCQVVGYPV	MIKASWGGGG	KGIRKVNDD	EVRTLKQVQ	GEVPGSPIFI
370	380	390	400	410	420
MRLAAQSRHL	EVQLLCDQYG	NVAALHSRDC	SVQRRHQKII	EEGPVTVAPR	ETVKELEQAA
430	440	450	460	470	480
RRLAKAVGYV	GAATVEYLYS	METGEYFLE	LNRLQVEHP	VTEWIAEVNL	PAAQVAVGMG
490	500	510	520	530	540
IPLWQIPEIR	RFYGMNHGGG	YDLWRKTAAL	ATPFNFDEVD	SKWPKGHCVA	VRITSEDPDD
550	560	570	580	590	600
GFKPTGGKVK	EISFKSKPNV	WAYFSVKSGG	GIHEFADSQF	GHVFAVGTRR	SAAITTMALA
610	620	630	640	650	660
LKEVQIRGEI	HSNVDYTVDL	LNASDFRENK	IHTGWLDTRI	AMRVQAERPP	WYISVVGAL
670	680	690	700	710	720
YKTVTANTAT	VSDYVGYLTK	GQIPPKHISL	VYTTVALNID	GKKYTIDTVR	SGHGYSRLRM
730	740	750	760	770	780
NGSTVDANVQ	ILCDGGLLMQ	LDGNSHVIYA	EEEASGTRLL	IDGKTCMLQN	DHDPSKLLAE

FIG. 26G-1

790	800	810	820	830	840
TPCKLLRFLV	ADGAHVDA DV	PYAEVEVMKM	CMPLLSPASG	VIHVVMSEGO	AMQAGDLIAR
850	860	870	880	890	900
LDLDDPSAVK	RAEPFEDTFP	QMGLPIAASG	QVHKLCAASL	NACRMILAGY	EHDIDKVVPE
910	920	930	940	950	960
LVYCLDTPEL	PFLQWEELMS	VLATRLPRNL	KSELEGKYEE	YKVKFDSGII	NDFPANMLRV
970	980	990	1000	1010	1020
IIEENLACGS	EKEKATNERL	VEPLMSLLKS	YEGGRESHAH	FVVKSLFEEY	LYVEELFSDG
1030	1040	1050	1060	1070	1080
IQSDVIERLR	LQHSKDLQKV	VDIVLSHQSV	RNKTCLILKL	MESLVYPNPA	AYRDQLIRFS
1090	1100	1110	1120	1130	1140
SLNHKAYYKL	ALKASELLEQ	TKLSELRARI	ARSLSELEMF	TEESKGLSMH	KREIAIKESM
1150	1160	1170	1180	1190	1200
EDLVTAPLPV	EDALISLFDC	SDTTVQQRVI	ETYIARLYQP	HLVKDSIKMK	WIESGVIALW
1210	1220	1230	1240	1250	1260
EFPEGHFDAR	NGGAVLGDKR	WGAMVIVKSL	ESLSMAIRFA	LKETSHYTSS	EGNMMHIALL
1270	1280	1290	1300	1310	1320
GADNKMHI IQ	ESGDDADRIA	KLPLILKDNV	TDLHASGVKT	ISFIVQRDEA	RMTMRRTFLW
1330	1340	1350	1360	1370	1380
SDEKLSYEEE	PILRHVEPPL	SALLELDKIK	VKGYNEMKYT	PSRDRQWHIY	TLRNTENPKM
1390	1400	1410	1420	1430	1440
LHRVFFRTL V	RQPSVSNKFS	SGQIGDMEVG	SAEPLSFTS	TSILRSLMTA	IEEELHAIR
1450	1460	1470	1480	1490	1500
TGSHMYLHV	LKEQKLLDLV	PVSGNTVLDV	GQDEATAYSL	LKEMAMKIHE	LVGARMHHS
1510	1520	1530	1540	1550	1560
VCQWEVKLKL	DCDGPASGTW	RIVTTNVTSH	TCTVDIYREM	EDKESRKL VY	HPATPAAGPL
1570	1580	1590	1600	1610	1620
HGVALNNPYQ	PLSVIDLKRC	SARNNRTTYC	YDFPLAFETA	VRKSWSSSTS	GASKGVENAO

## FIG.26G-2

1630            1640            1650            1660            1670            1680  
 CYVKATELVF ADKHGSWGTP LVQMDRPA GL NDIGMVAWTL KMSTPEFPSG REIIVVANDI

1690            1700            1710            1720            1730            1740  
 TFRAGSFGPR EDAFFEAVTN LACEKKLPLI YLAANS GARI GIADEVKSCF RVGWSDDGSP

1750            1760            1770            1780            1790            1800  
 ERGFQYIYLS EEDYARIGTS VIAHKMQLDS GEIRWVIDSV VGKEDGLGVE NIHGSAAIAS

1810            1820            1830            1840            1850            1860  
 AYSRAYKETF TLTFVTGRTV GIGAYLARLG IRCIQRLDQP IILTGYSALN KLLGREVYSS

1870            1880            1890            1900            1910            1920  
 HMQLGGPKIM ATNGVVHLTV SDDLEGVSNI LRWLSYVPAY IGGPLPV TTP LDPPDRPVAY

1930            1940            1950            1960            1970            1980  
 IPENSCDPRA AIRGVDDSQ G KWLGGMFDKD SFVETFEGWA KTVVTGRAKL GGIPVGVIAV

1990            2000            2010            2020            2030            2040  
 ETQTMQTIP ADPGQLDSRE QSVPRAGQVW FPDSATKTAQ ALLDFNREGL PLFILANWRG

2050            2060            2070            2080            2090            2100  
 FSGGQRDLFE GILQAGSTIV ENLRTYNQPA FVYIPMAAEL RGGAWVVVDS KINPDRIECY

2110            2120            2130            2140            2150            2160  
 AERTAKGNVL EPQGLIEIKF RSEELQDCMS RLDPTLIDLK AKLEVANKNG SADTKSLQEN

2170            2180            2190            2200            2210            2220  
 IEARTKQLMP LYTQIAIRFA ELHDTSLRMA AKGVIKKVVD WEESRSFFYK RLRRRISE DV

2230            2240            2250            2260            2270            2280  
 LAKEIRAVAG EQFSHQPAIE LIKKWYSASH AAEWDDDDAF VAWMDNPENY KDYIQYLKAQ

2290            2300            2310            2320  
 RVSQSLSSLS DSSSDLQALP QGLSMLLDKM DPSRRAQLVE EIRKVLG

## FIG.26G-3

