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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 May 2013 (10.05.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/066423 A2

(51) **International Patent Classification**:
C12N 15/82 (2006.01)

(21) **International Application Number**:
PCT/US2012/043082

(22) **International Filing Date**:
19 June 2012 (19.06.2012)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
61/499,441 21 June 2011 (21.06.2011) US

(71) **Applicants** *(for all designated States except US)*: **PIONEER HI-BRED INTERNATIONAL, INC.** [US/US]; 7100 NW 62nd Avenue, PO Box 1000, Johnston, IA 50131-1000 (US). **E.I. DU PONT DE NEMOURS AND COMPANY** [US/US]; 1007 Market Street, Wilmington, DE 50131-1000 (US).

(72) **Inventors; and**
(71) **Applicants** : **BIDNEY, Dennis, L.** [US/US]; 8385 Plum Drive, Urbandale, IA 50322 (US). **CIGAN, Mark** [US/US]; 5764 Chatham Circle, Johnston, IA 50131 (US). **FALCO, Carl** [US/US]; 1902 Millers Road, Wilmington, DE 79810 (US). **GAO, Huirong** [US/US]; 7068 Ridgedale Court, Johnston, IA 50131 (US). **JANTZ, Derek** [US/US]; 215 Hillview Drive, Durham, NC 27703 (US). **LASSNER, Mike** [US/US]; 2935-133rd Way, Urbandale, IA 50323 (US). **LOWE, Keith** [US/US]; 6407 Harbor Oaks Drive, Johnston, IA 50131 (US). **LYZNIK, Leszek, Aleksander** [US/US]; 8208 Chambery Blvd, Johnston, IA 50131 (US). **SMITH, James, Jefferson** [US/US]; 1808 Carnation Drive, Durham, NC 27703 (US).

(74) **Agents**: **SUNMAN, Jeffrey, A.** et al.; Alston & Bird LLP, Bank of America Plaza, 101 South Tryon Street, Suite 4000, Charlotte, NC 28280-4000 (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, 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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, 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, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2013/066423 A2

(54) **Title**: METHODS AND COMPOSITIONS FOR PRODUCING MALE STERILE PLANTS

(57) **Abstract**: Methods of making a targeted modification in a male fertility gene in the genome of a plant are disclosed. The methods involve contacting a plant cell with an engineered double-strand-break-inducing agent capable of inducing a double-strand break in a target sequence in the male fertility gene and identifying a eel! comprising an alteration in the target sequence. Also disclosed are plants, plant cells, plant parts, and seeds comprising a male fertility gene with an alteration in a male fertility gene. Nucleic acid molecules comprising male fertility genes with at least one targeted modification therein, optimized nucleic acid molecules encoding endonucleases that are engineered double-strand- break-inducing agents and expression cassettes, host cells, and plants comprising one or more of the nucleic acid molecules are further disclosed.

METHODS AND COMPOSITIONS FOR PRODUCING MALE STERILE PLANTS

FIELD OF INVENTION

The invention relates to the field of plant molecular biology, particularly to methods for making targeted mutations in male fertility genes in plants.

REFERENCE TO A SEQUENCE LISTING SUBMITTED AS A TEXT FILE VIA EFS-WEB

The official copy of the sequence listing is submitted electronically via EFS-Web as an ASCII formatted sequence listing with a file named 418871SEQLIST.TXT, created on June 12, 2012, and having a size of 322 kilobytes, and is filed concurrently with the specification. The sequence listing contained in this ASCII formatted document is part of the specification and is herein incorporated by reference in its entirety.

BACKGROUND

Development of hybrid plant breeding has made possible considerable advances in quality and quantity of crops produced. Increased yield and combination of desirable characteristics, such as resistance to disease and insects, heat and drought tolerance, along with variations in plant composition are all possible because of hybridization procedures. These procedures frequently rely heavily on providing for a male parent contributing pollen to a female parent to produce the resulting hybrid.

Field crops are bred through techniques that take advantage of the plant's method of pollination. A plant is self-pollinating if pollen from one flower is transferred to the same or another flower of the same plant or a genetically identical plant. A plant is cross-pollinated if the pollen comes from a flower on a different plant.

In certain species, such as *Brassica campestris*, the plant is normally self-sterile and can only be cross-pollinated. In self-pollinating species, such as soybeans and cotton, the male and female plants are anatomically juxtaposed. During natural pollination, the male reproductive organs of a given flower pollinate the female reproductive organs of the same flower.

Maize plants (*Zea mays* L.) can be bred by both self-pollination and cross-pollination techniques. Maize has male flowers, located on the tassel, and female flowers, located on the ear, on the same plant. It can self or cross pollinate. Natural pollination occurs in maize when wind blows pollen from the tassels to the silks that protrude from the tops of the incipient ears.

The development of maize hybrids requires the development of homozygous inbred lines, the crossing of these lines, and the evaluation of the crosses. Pedigree breeding and recurrent selection are two of the breeding methods used to develop inbred lines from populations. Breeding programs combine desirable traits from two or more inbred lines or various broad-based sources into breeding pools from which new inbred lines are developed by selfing and selection of desired phenotypes. A hybrid maize variety is the cross of two such inbred lines, each of which may have one or more desirable characteristics lacked by the other or which complement the other. The new inbreds are crossed with other inbred lines and the hybrids from these crosses are evaluated to determine which have commercial potential. The hybrid progeny of the first generation is designated F_1. In the development of hybrids only the F_1 hybrid plants are sought. The F_1 hybrid is more vigorous than its inbred parents. This hybrid vigor, or heterosis, can be manifested in many ways, including increased vegetative growth and increased yield.

Hybrid maize seed can be produced by a male sterility system incorporating manual detasseling. To produce hybrid seed, the male tassel is removed from the growing female inbred parent, which can be planted in various alternating row patterns with the male inbred parent. Consequently, providing that there is sufficient isolation from sources of foreign maize pollen, the ears of the female inbred will be fertilized only with pollen from the male inbred. The resulting seed is therefore hybrid (F_1) and will form hybrid plants.

Field variation impacting plant development can result in plants tasseling after manual detasseling of the female parent is completed. Or, a female inbred plant tassel may not be completely removed during the detasseling process. In any event, the result is that the female plant will successfully shed pollen and some female plants will be self-pollinated. This will result in seed of the female inbred being harvested along with the hybrid seed which is normally produced. Female inbred seed does not exhibit heterosis and therefore is not as productive as F_1 seed. In addition, the presence of female inbred seed can represent a germplasm security risk for the company producing the hybrid.

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

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

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

BRIEF SUMMARY OF THE INVENTION

The present invention provides methods for making a targeted modification in a male fertility gene in the genome of a plant. The methods

involve contacting at least one plant cell comprising a target sequence in a male fertility gene with an engineered double-strand-break-inducing agent that is capable of inducing a double-strand break at the target sequence. The methods further involve identifying at least one cell comprising an alteration in its genome at the target sequence. If desired, the methods can further comprise regenerating a fertile plant comprising the alteration. The alterations include, but are not limited to, the replacement of at least one nucleotide in the target sequence, the deletion of at least one nucleotide in the target sequence, the insertion of at least one nucleotide in the target sequence or any combination thereof. For example, the alteration can be the insertion of a transgene in the target sequence of the male fertility gene or a null mutation, wherein a progeny plant that is homozygous for the null mutation is male sterile. In an embodiment of the invention, the insertion of a transgene in the target sequence of the male fertility gene is a null mutation in the male fertility gene.

In a first embodiment of methods for making a targeted modification in a male fertility gene in the genome of a plant, the male fertility gene is selected from the group consisting of *MS26, MS45, BS92-7, 5126* and *Msca1*.

In a second embodiment, the methods further comprise regenerating a plant, particularly a fertile plant, comprising the alteration.

In a third embodiment, the engineered double-strand-break-inducing agent is an endonuclease, a zinc finger nuclease, a TAL effector nuclease, a transposase, or a site-specific recombinase. Preferably, the endonuclease is modified to specifically cut at the target sequence and no longer cuts at its wild-type endonuclease target sequence.

In a fourth embodiment, the methods further comprise selfing the fertile plant and selecting a progeny plant resulting therefrom, wherein said progeny plant is homozygous for the alteration.

In a fifth embodiment, the methods further comprise crossing the fertile plant with a second fertile plant comprising a null mutation in the male fertility gene and selecting a progeny plant resulting therefrom, wherein said progeny plant is male sterile.

In a sixth embodiment, the alteration comprises insertion of a transgene comprising a polynucleotide of interest. The transgene can further comprise a promoter operably linked to the polynucleotide of interest, wherein the promoter

is capable of driving the expression of the polynucleotide of interest in a plant. For example, the polynucleotide of interest can encode a phenotypic marker or an RNA or protein providing an agronomic advantage to the plant.

In a seventh embodiment, the plant is selected from the group consisting of maize, sorghum, rice, wheat, rye, barley, millet, and oat.

In an eighth embodiment, the male fertility gene is *MS26*. For example, the target sequence for this embodiment can comprise the nucleotide sequence set forth in SEQ ID NO: 1. The engineered double-strand-break-inducing agent can, for example, be derived from I-CreI.

In a ninth embodiment, the step of contacting at least one plant cell comprising a target sequence in *MS26* with the engineered double-strand-break-inducing agent comprises introducing into the at least one plant cell a nucleic acid construct comprising a nucleotide sequence encoding the engineered double-strand-break-inducing agent. The nucleotide sequence can be selected, for example, from the group consisting of the nucleotide sequences set forth in SEQ ID NO: 4, 5, 6, and 7; and a nucleotide sequence having at least 80% nucleotide sequence identity to at least one nucleotide sequence selected from the group consisting of the nucleotide sequences set forth in SEQ ID NOS: 4, 5, 6, and 7, wherein the nucleotide sequence encodes a polypeptide comprising endonuclease activity. If desired, the nucleic acid construct can further comprise a promoter operably linked to the nucleotide sequence encoding the engineered double-strand-break-inducing agent, wherein the promoter is capable of driving expression of the nucleotide sequence in a plant cell. For example, the promoter can be a maize ubiquitin promoter. Additionally, the nucleic acid construct can further comprise an operably linked coding sequence for a nuclear localization signal. Such nuclear localization signals can comprise, for example, an amino acid sequence selected from the group consisting of SEQ ID NOS: 2, 3, and 21.

In a tenth embodiment, the male fertility gene is *MS45*. For example, the target sequence for this embodiment can comprise the nucleotide sequence set forth in SEQ ID NO: 20. The engineered double-strand-break-inducing agent can, for example, be derived from I-CreI.

In an eleventh embodiment, the step of contacting at least one plant cell comprising a target sequence in *MS45* with the engineered double-strand-break-inducing agent comprises introducing into the at least one plant cell a nucleic

acid construct comprising a nucleotide sequence encoding the engineered double-strand-break-inducing agent. The nucleotide sequence can be selected, for example, from the group consisting of the nucleotide sequence set forth in SEQ ID NO: 22, 23, or 34; and a nucleotide sequence having at least 80% nucleotide sequence identity to at least one nucleotide sequence selected from the group consisting of the nucleotide sequences set forth in SEQ ID NOS: 22, 23, or 34, wherein the nucleotide sequence encodes a polypeptide comprising endonuclease activity. If desired, the nucleic acid construct can further comprise a promoter operably linked to the nucleotide sequence encoding the engineered double-strand-break-inducing agent, wherein the promoter is capable of driving expression of the nucleotide sequence in a plant cell. For example, the promoter can be a maize ubiquitin promoter. Additionally, the nucleic acid construct can further comprise an operably linked coding sequence for a nuclear localization signal. Such nuclear localization signals can comprise, for example, an amino acid sequence selected from the group consisting of SEQ ID NOS: 2, 3, and 21.

The present invention further provides isolated nucleic acid molecules comprising at least one male fertility gene with a targeted modification or alteration and plants, plant parts, plant cells, and seeds comprising at least one male fertility gene with a targeted modification or alteration. The plants of invention include, but are not limited to, a plant produced by any of the methods disclosed herein and a descendant of any plant produced by any of such methods, wherein the descendant comprises the alteration.

In one embodiment, the plant comprises a targeted modification in a male fertility gene in its genome, wherein the targeted modification is the insertion of a transgene, and wherein the male fertility gene is selected from the group consisting of *MS26, MS45, BS92-7, 5126 and Msca1.* For example, the insertion of a transgene can cause a null mutation in the male fertility gene, and a plant that is homozygous for the alteration is male sterile.

In another embodiment, the plant is selected from the group consisting of maize, sorghum, rice, wheat, rye, barley, millet, and oat.

In a further embodiment, the plant is a sorghum plant comprising a targeted modification in the male fertility gene *MS26*, wherein the *MS26* gene comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77 and 78.

Additionally provided are isolated nucleic acid molecules encoding engineered double-strand-break-inducing agents that are capable of inducing a double-strand break in DNA comprising a target sequence of the invention. Expression cassettes comprising at least one isolated nucleic acid molecule encoding an engineered double-strand-break-inducing agent, and host cells, and plants comprising at least one of the expression cassettes are further provided.

In one embodiment of the invention, the expression cassettes comprise a promoter operably linked to a nucleotide sequence selected from the group consisting of SEQ ID NOS: 4, 5, 6, 7, 22, 23, and 34.

In another embodiment, the present invention provides a plant comprising an expression construct, which comprises a promoter operably linked to a nucleotide sequence encoding an endonuclease. The endonuclease is capable of specifically binding to and creating a double strand break in a target sequence selected from the group consisting of SEQ ID NOS: 1 and 20, wherein the promoter is capable of driving expression of an operably linked nucleotide sequence in a plant cell. The nucleotide sequence encoding the endonuclease can comprise a coding sequence of a DNA binding domain of an endonuclease, wherein the coding sequence is selected from the group consisting of:

(a) nucleotides 100-261 and nucleotides 661-822 of SEQ ID NO: 4;

(b) nucleotides 70-231 and nucleotides 631-792 of SEQ ID NO: 5;

(c) nucleotides 70-231 and nucleotides 820-981 of SEQ ID NO: 6, 7 or 34; and

(d) a degenerate coding sequence of (a), (b), or (c).

Preferably, the nucleotide sequence encoding the endonuclease is a nucleotide sequence selected from the group consisting of SEQ ID NOS: 4, 5, 6, 7, 22, 23, and 34.

BRIEF DESCRIPTION OF THE DRAWINGS AND THE SEQUENCE LISTING

The invention can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing, which form a part of this application. The sequence descriptions and sequence listing attached hereto comply with the rules governing nucleotide and amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §§1.821-1.825. The sequence descriptions contain the three letter

codes for amino acids as defined in 37 C.F.R. §§ 1.821-1.825, which are incorporated herein by reference.

Figures

Figure 1. DNA double-strand-break-induced DNA alteration of an endogenous target site. (A) A generalized endogenous target site with flanking genomic DNA sequences designated as DNA 1 and DNA 2 which can be used as DNA exchange regions by homologous recombination. (B) A generalized DNA construct that can be used to express a DNA endonuclease (nuclease gene) to recognize and cleave the endogenous target site. The DNA endonuclease gene can be physically linked to the donor DNA described in (C) or (D), or substituted by other double-strand-break-inducing agents. (C) A generalized donor DNA construct having two regions DNA1 and DNA 2 of homology to the genomic target, which flank a polynucleotide of interest and/or marker gene. (D) A generalized donor DNA construct that does not have regions of homology to the genomic target to flank a polynucleotide of interest and/or marker gene. Insertion of the DNA fragment will produce an insertion of the polynucleotide of interest at or near the recognition site. (E) One expected outcome when the polynucleotide of interest and/or marker gene of donor construct described in (C) or (D) is inserted at the endogenous target site by homologous recombination or non-homologous recombination, respectively. (F) Another outcome when the endogenous target site is altered by a deletion during the repair of the DNA double-strand break generated by the DNA endonuclease. The polynucleotide of interest and/or marker gene of donor construct described in (C) or (D) can be inserted at unrelated sites by random DNA integration. (G) Another outcome when the endogenous target site is altered by the insertion of an unrelated DNA during the repair of the DNA double-strand breaks cleaved by the DNA endonuclease. The polynucleotide of interest and/or marker gene of donor construct described in (C) or (D) can be inserted at unrelated sites by random DNA integration.

Figure 2. Mutated alleles of the maize TS-MS26 target site. The mutated alleles found in the first generation maize transformants (T0 plants) were centered on the apparent 3' end GTAC overhang produced by the engineered MS26 endonuclease. Wild type (SEQ ID NO: 35), 3281 (SEQ ID NO: 36), 2963

(SEQ ID NO: 37), 2980 (SEQ ID NO: 38), 3861 (SEQ ID NO: 39), 3956 (SEQ ID NO: 40), 3990 (SEQ ID NO: 41), 6227 (SEQ ID NO: 42).

Figure 3. Sequence homology across the TS-MS26 target site (MS26 TS, SEQ ID NO:1) between genomic regions of *MS26* genes from maize (maize MS26, SEQ ID NO:13), rice (rice MS26, SEQ ID NO:14), sorghum (sorghum MS26, SEQ ID NO:15), and rye (rye MS26, SEQ ID NO:16).

Figure 4. Vectors for the biolistic transformation of rice.

Figure 5. Mutations at the rice *MS26* gene introduced by biolistic transformation.

Figure 6. A) Plasmid fragment of PHP40827 used for rice transformation. This plasmid contains a tetracycline repressor under the control of the maize Ubiquitin promoter, and a blue-fluorescence gene (CFP) regulated by the ZmEND2 promoter. In addition, this plasmid fragment contains a copy of a red fluorescence gene regulated by the maize Histone 2B promoter. A portion of the red fluorescence gene in this construct was duplicated in a direct orientation, consisting of two fragments of the RFP gene with 369 bp of overlap. The two fragments are separated by a 136-bp spacer which contains the TS-MS26 target site (Figure 6A). B) PCR analysis for mutations at the TS-MS26 target site of TET treated events (1, 2, 3) compared to the PCR products of these same events not exposed to tetracycline (control).

Figure 7. Mutations at the rice *MS26* gene identified in PHP40827 callus events. Highlighted in gray is the wild-type TS-MS26 from rice. Wild-type rice *MS26* (SEQ ID NO: 49), ms26.1 (SEQ ID NO: 50), ms26.2 (SEQ ID NO: 51), ms26.3 (SEQ ID NO: 52), ms26.4 (SEQ ID NO: 53), ms26.5 (SEQ ID NO 54), ms26.6 (SEQ ID NO: 55), ms26.7 (SEQ ID NO: 56), ms26.8 (SEQ ID NO: 57)

Figure 8. Maize T0 plants at time of flowering. There was no obvious difference in the growth and development of T0 plants containing one mutated *ms26* allele (two outside plants) as compared to the T0 biallelic event (the tagged plant) produced by the engineered MS26++ endonuclease (A). The biallelic event was sterile (the tassel at anthesis shown between two tassels from monoallelic events) (B).

Figure 9. Maize T1 progeny (A, B, C) plants at time of flowering. T1 progeny plants heterozygous for the *ms26-Td* or *ms26-Ci* mutant alleles (two plants at the left side) and two homozygous, sterile T1 plants at the right side are

shown (A). There were no pleiotropic effects of the *ms26* gene mutations on the growth and development of T1 progeny plants. Both mutant alleles (*ms26-Td* and *ms26-Ci)* produced a sterile phenotype only when in homozygous state in the T1 progeny plants (B and C).

Figure 10. Panicles and anthers from male sterile (*ms26/ms26*) and male fertile (*MS26/ms26*) rice plants. (A) Rice panicles showing male sterile homozygous *ms26/ms26* plants on the left and male fertile heterozygous *Ms26/ms26* plants on the right. Anther squashes from male sterile *ms26/ms26* (B) and male fertile *Ms26/ms26* (C) panicles shown in Panel A.

Figure 11A-E. Alignment of fragments from the plant-optimized nucleotide sequence of meganucleases comprising the nucleotides 170-231 and nucleotides 820-981 of SEQ ID NO:6, 7 or 34, the nucleotides 70-231 and nucleotides 631-792 of SEQ ID NO: 5, and the nucleotides 100-261 and nucleotides 661-822 of SEQ ID NO: 4.

Figure 12. Alignment of the MS26 recognition sequence and DNA sequences from different sorghum plants containing mutations and deletions at the TS-MS26 target site. NOs:1 and 62-78 correspond to SEQ ID NOs:1 and 62-78, respectively. SEQ ID NO:62 represents the wild-type sorgham *MS26* nucleotide sequence.

Figure 13. (A) panicles of MS26/ms26.78Δ and (B) panicles of ms26.78Δ / ms26.78Δ sorghum plants .

Figure 14. Stigma, anthers and pollen from MS26/ms26.78Δ plants (Figure 14A) and ms26.78Δ /ms26.78Δ plants (Figure 14B). Pollen was easily detected in MS26/ms26.78Δ anthers (Figure 14C), however pollen was not observed from anthers from ms26.78Δ /ms26.78Δ plants (Figure 14D).

Sequences

SEQ ID NO: 1 is the nucleotide sequence of the TS-MS26 target site recognized by the engineered MS26 endonuclease that is capable of inducing a double-strand break at that target sequence.

SEQ ID NO: 2 is a nuclear localization signal SV40 NLS-1.

SEQ ID NO: 3 is a nuclear localization signal SV40 NLS-2.

SEQ ID NO: 4 is the plant optimized nucleotide sequence (without an intron) encoding the engineered MS26 endonuclease.

SEQ ID NO: 5 is the plant optimized nucleotide sequence (without an intron) encoding the engineered MS26+ endonuclease.

SEQ ID NO: 6 is the plant optimized nucleotide sequence encoding the engineered MS26++ endonuclease. This nucleotide sequence has a GC content adjusted to less than 60% and contains an intron.

SEQ ID NO: 7 is the plant optimized nucleotide sequence encoding the engineered MS26+ endonuclease. This nucleotide sequence has a GC content adjusted to less than 60 and contains an intron.

SEQ ID NO: 8 is the nucleotide sequence of a male fertility gene encoding a cytochrome P450 (MS26) in maize (AF366297)

SEQ ID NO: 9 is the nucleotide sequence of a male fertility gene encoding a cytochrome P450 (MS26) in rice (LOC_Os03g07250)

SEQ ID NO: 10 is the nucleotide sequence of a male fertility gene encoding a cytochrome P450 (MS26) in sorghum.

SEQ ID NO: 11 is the nucleotide sequence of a male fertility gene encoding a cytochrome P450 (MS26) in rye (Secale cereal, FJ539083).

SEQ ID NO: 12 is the amino acid sequence of a male fertility gene (*MS26*) encoding a cytochrome P450 in maize (AAK52956.1).

SEQ ID NO: 13 is the maize genomic region comprising the maize TS-MS26 target site shown in Figure 3.

SEQ ID NO: 14 is the rice genomic region comprising the rice TS-MS26 target site shown in Figure 3.

SEQ ID NO: 15 is the sorghum genomic region comprising the sorghum TS-MS26 target site shown in Figure 3.

SEQ ID NO: 16 is the rye genomic region comprising the rye TS-MS26 target site shown in Figure 3.

SEQ ID NO: 17 is primer UNIMS26 5'-2.

SEQ ID NO: 18 is primer UNIMS26 3'-1.

SEQ ID NO: 19 is the maize genomic region comprising the maize-TS-MS26 target sequence.

SEQ ID NO: 20 is the TS-MS45 target sequence from maize.

SEQ ID NO: 21 is the nuclear localization amino acid sequence used in MAY1/MAY fusions.

SEQ ID NO: 22 is the plant optimized nucleotide sequence encoding MAY1.

SEQ ID NO: 23 is the plant optimized nucleotide sequence encoding MAY2.

SEQ ID NO: 24 is the nucleotide sequence of a male fertility gene encoding a chalcone and stilbene synthase (5126) in maize (AX060770).

SEQ ID NO: 25 is the nucleotide sequence of a male fertility gene encoding a chalcone and stilbene synthase (5126) in rice (LOC_Os07g22850).

SEQ ID NO: 26 is the nucleotide sequence of a male fertility gene encoding a dihydroflavonol 4-reductase (BS7) in maize (AF366295).

SEQ ID NO: 27 is the nucleotide sequence of a male fertility gene encoding a a dihydroflavonol 4-reductase (BS7) in rice (LOC_Os08g40440).

SEQ ID NO: 28 is the nucleotide sequence of a male fertility gene encoding a strictosidine synthase (MS45) in maize (AF360356).

SEQ ID NO: 29 is the nucleotide sequence of male fertility gene encoding a strictosidine synthase (MS45) in rice (LOC_Os03g15710).

SEQ ID NO: 30 is plasmid PHP31457.

SEQ ID NO: 31 is plasmid PHP31459.

SEQ ID NO: 32 is the nucleotide sequence of a male fertility gene encoding a MS22 protein in maize.

SEQ ID NO: 33 is a DNA sequence encoding the nuclear localization amino acid sequence used in MAY1/MAY fusions.

SEQ ID NO: 34 is a plant optimized gene encoding a MAY1-linker-MAY2 protein.

SEQ ID NO: 35 is the wild-type TS-MS26 DNA fragment shown in Figure 2.

SEQ ID NO: 36 is the 3281 TS-MS26 DNA fragment shown in Figure 2.

SEQ ID NO: 37 is the 2963 TS-MS26 DNA fragment shown in Figure 2.

SEQ ID NO: 38 is the 2980 TS-MS26 DNA fragment shown in Figure 2.

SEQ ID NO: 39 is the 3861TS-MS26 DNA fragment shown in Figure 2.

SEQ ID NO: 40 is the 3956TS-MS26 DNA fragment shown in Figure 2.

SEQ ID NO: 41 is the 3990TS-MS26 DNA fragment shown in Figure 2.

SEQ ID NO: 42 is the 6227 TS-MS26 DNA fragment shown in Figure 2.

SEQ ID NO: 43 is the wild-type TS-MS26 DNA fragment shown in Figure 5.

SEQ ID NO: 44 is the Ev48 TS-MS26 DNA fragment shown in Figure 5.

SEQ ID NO: 45 is Ev62.1 TS-MS26 DNA fragment shown in Figure 5.

SEQ ID NO: 46 is the Ev62.13 TS-MS26 DNA fragment shown in Figure 5.

SEQ ID NO: 47 is the Ev62.14 TS-MS26 DNA fragment shown in Figure 5.

SEQ ID NO: 48 is the Ev67 TS-MS26 DNA fragment shown in Figure 5.

SEQ ID NO: 49 is the wild-type TS-MS26 DNA fragment shown in Figure 7.

SEQ ID NO: 50 is the ms26.1 TS-MS26 DNA fragment shown in Figure 7.

SEQ ID NO: 51 is the ms26.2 TS-MS26 DNA fragment shown in Figure 7.

SEQ ID NO: 52 is the ms26.3 TS-MS26 DNA fragment shown in Figure 7.

SEQ ID NO: 53 is the ms26.4 TS-MS26 DNA fragment shown in Figure 7.

SEQ ID NO: 54 is the ms26.5 TS-MS26 DNA fragment shown in Figure 7.

SEQ ID NO: 55 is the ms26.6 TS-MS26 DNA fragment shown in Figure 7.

SEQ ID NO: 56 is the ms26.7 TS-MS26 DNA fragment shown in Figure 7.

SEQ ID NO: 57 is the ms26.8 TS-MS26 DNA fragment shown in Figure 7.

SEQ ID NO: 58 is the nucleotide sequence of plasmidPHP40082.

SEQ ID NO: 59 is the nucleotide sequence of plasmid PHP40126.

SEQ ID NO: 60 is the nucleotide sequence of plasmid PHP40827.

SEQ ID NO: 61 is the nucleotide sequence of plasmid PHP42063.

SEQ ID NO: 62-78 are the DNA fragments shown in Figure 12. SEQ ID NO: 62 is the nucleotide sequence of a portion of the wild-type sorghum *MS26* gene. SEQ ID NOs: 63-78 set forth modifications to the nucleotide sequence of the wild-type sorghum *MS26* gene set forth in SEQ ID NO: 62.

DETAILED DESCRIPTION OF THE INVENTION

All publications and patent applications mentioned in the specification are indicative of the level of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

As used herein and in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a plant" includes a plurality of such plants; reference to "a cell" includes one or more cells and equivalents thereof known to those skilled in the art, and so forth.

A reliable system of producing genetic male sterility would provide advantages to develop hybrid plants. The laborious detasseling process can be avoided in some genotypes by using cytoplasmic male-sterile (CMS) inbreds. In the absence of a fertility restorer gene, plants of a CMS inbred are male sterile as a result of factors resulting from the cytoplasmic, as opposed to the nuclear, genome. Thus, this characteristic is inherited exclusively through the female parent in maize plants, since only the female provides cytoplasm to the fertilized seed. CMS plants are fertilized with pollen from another inbred that is not male-sterile. Pollen from the second inbred may or may not contribute genes that make the hybrid plants male-fertile. Usually seed from detasseled normal maize and CMS produced seed of the same hybrid must be blended to insure that adequate pollen loads are available for fertilization when the hybrid plants are grown and to insure cytoplasmic diversity.

Nuclear (genic) sterility can be either dominant or recessive. Dominant sterility can only be used for hybrid seed formation if propagation of the female line is possible (for example, via *in vitro* clonal propagation). Recessive sterility can be used if sterile and fertile plants are easily discriminated. Commercial utility of genic sterility systems is limited however by the expense of clonal propagation and roguing the female rows of self-fertile plants.

One type of genetic sterility is disclosed in U.S. Patents 4,654,465 and 4,727,219 to Brar, et al. However, this form of genetic male sterility requires maintenance of multiple mutant genes at separate locations within the genome and requires a complex marker system to track the genes and make use of the system convenient. Patterson also described a genic system of chromosomal

translocations which can be effective, but which are complicated. (See, U.S. Patents No. 3,861,709 and 3,710,511.)

Many other attempts have been made to improve on these systems. For example, Fabijanski, et al., developed several methods of causing male sterility in plants (see EPO 89/3010153.8 publication no. 329,308 and PCT application PCT/CA90/00037 published as WO 90/08828). One method includes delivering into the plant a gene encoding a cytotoxic substance associated with a male tissue specific promoter. Another involves an antisense system in which a gene critical to fertility is identified and an antisense to the gene inserted in the plant. Fabijanski, et al. also shows several cytotoxic antisense systems. See EP0329308. Still other systems use "repressor" genes which inhibit the expression of another gene critical to male sterility. See PCT/GB90/00102, published as WO 90/08829. For yet another example see US Patent No. 6,281,348.

A still further improvement of this system is one described at U.S. Patent No. 5,478,369 in which a method of imparting controllable male sterility is achieved by inactivating or otherwise silencing a gene native to the plant that is critical for male fertility and transforming that plant with the gene critical to male fertility linked to an inducible promoter controlling expression of the gene. That is, the expression of the endogenous sequence is prevented, by any of the methods known to a skilled person in the art for preventing expression of a sequence (such an antisense methods, cosuppression, mutation, use of ribozymes or hairpins, various repression systems and the like, discussed *infra*.) The plant is thus constitutively sterile, becoming fertile only when the promoter is induced and its linked male fertility gene is expressed.

In a number of circumstances, a male sterility plant trait is expressed by maintenance of a homozygous recessive condition. Difficulties arise in maintaining the homozygous condition, when a restoration gene must be used for maintenance. For example, a natural mutation in a gene critical to male fertility can impart a male sterility phenotype to plants when this mutant allele is in the homozygous state. But because this homozygosity results in male sterility, the homozygous male-sterile line cannot be maintained. Fertility is restored when the non-mutant form of the gene is introduced into the plant.

However, this form of line maintenance removes the desired homozygous recessive condition, restores full male fertility in half of the resulting progeny, and prevents maintenance of pure male sterile maternal lines. These issues can be avoided where production of pollen containing the restoration gene is eliminated, thus providing a maintainer plant producing only pollen not containing the restoration gene, and the progeny retain their homozygous condition when fertilized by such pollen. An example of one approach is shown in Dellaporta et al., 6,743,968, in which a plant is produced having a hemizygotic construct comprising a gene that produces a product fatal to a cell, linked with a pollen-specific promoter, and the restoration gene. When crossed with the homozygous recessive male sterile plant, the progeny thus retains the homozygous recessive condition. Other approaches have been described, for example in U.S. 7,696,405.

As noted, an important aspect of much of the work underway with male sterility systems is the identification of genes impacting male fertility. Such a gene can be used in a variety of systems to control male fertility including those described above.

As used herein "genetic male sterility" results from a mutation, suppression, or other impact to one of the genes critical to a specific step in microsporogenesis, the term applied to the entire process of pollen formation. These genes can be collectively referred to as male fertility genes (or, alternatively, male sterility genes). There are many steps in the overall pathway where gene function impacts fertility, as demonstrated by the frequency of genetic male sterility in maize. New alleles of male sterility mutants are uncovered in materials that range from elite inbreds to unadapted populations.

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

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

Table 1 lists a number of known male fertility mutants or genes from *Zea mays*.

TABLE 1

Male fertility mutants or genes from *Zea mays*.

GENE NAME	ALTERNATE NAME	REFERENCE
ms1 male sterile1	*male sterile1*, *ms1*	Singleton, WR and Jones, DF. 1930. J Hered 21:266-268
ms10 male sterile10	*male sterile10*, *ms10*	Beadle, GW. 1932. Genetics 17:413-431
ms11 male sterile11	*ms11*, *male sterile11*	Beadle, GW. 1932. Genetics 17:413-431
ms12 male sterile12	*ms12*, *male sterile12*	Beadle, GW. 1932. Genetics 17:413-431
ms13 male sterile13	*ms*-6060*, *male sterile13*, *ms13*	Beadle, GW. 1932. Genetics 17:413-431
ms14 male sterile14	*ms14*, *male sterile14*	Beadle, GW. 1932. Genetics 17:413-431
ms17 male sterile17	*ms17*, *male sterile17*	Emerson, RA. 1932. Science 75:566
ms2 male sterile2	*male sterile2*, *ms2*	Eyster, WH. 1931. J Hered 22:99-102
ms20 male sterile20	*ms20*, *male sterile20*	Eyster, WH. 1934. Genetics of Zea mays. Bibliographia Genetica 11:187-392
ms23 male sterile23	: *ms*-6059*, *ms*-6031*, *ms*-6027*, *ms*-6018*,	West, DP and Albertsen, MC. 1985. MNL 59:87

	ms-6011*, *ms35*, *male sterile23*, *ms*-Bear7*, *ms23*	
ms24 male sterile24	*ms24*, *male sterile24*	West, DP and Albertsen, MC. 1985. MNL 59:87
ms25 male sterile25	*ms*-6065*, *ms*-6057*, *ms25*, *male sterile25*, *ms*-6022*	Loukides, CA; Broadwater, AH; Bedinger, PA. 1995. Am J Bot 82:1017-1023
ms27 male sterile27	*ms27*, *male sterile27*	Albertsen, MC. 1996. MNL 70:30-31
ms28 male sterile28	*ms28*, *male sterile28*	Golubovskaya, IN. 1979. MNL 53:66-70
ms29 male sterile29	*male sterile29*, *ms*-JH84A*, *ms29*	Trimnell, MR et al. 1998. MNL 72:37-38
ms3 male sterile3	*Group 3*, *ms3*, *male sterile3*	Eyster, WH. 1931. J Hered 22:99-102
ms30 male sterile30	*ms30*, *msx*, *ms*-6028*, *ms*-Li89*, *male sterile30*, *ms*-LI89*	Albertsen, MC et al. 1999. MNL 73:48
ms31 male sterile31	*ms*-CG889D*, *ms31*, *male sterile31*	Trimnell, MR et al. 1998. MNL 72:38
ms32 male sterile32	*male sterile32*, *ms32*	Trimnell, MR et al. 1999. MNL 73:48-49
ms33 male sterile33	: *ms*-6054*, *ms*-6024*, *ms33*, *ms*-GC89A*, *ms*-6029*, *male sterile6019*, *Group 7*, *ms*-6038*, *ms*-Stan1*, *ms*-6041*, *ms*-6019*, *male sterile33*	Patterson, EB. 1995. MNL 69:126-128
ms34 male sterile34	*Group 1*, *ms*-6014*, *ms*-6010*, *male sterile34*, *ms34*, *ms*-6013*, *ms*-6004*, *male sterile6004*	Patterson, EB. 1995. MNL 69:126-128
ms36 male sterile36	*male sterile36*, *ms*-MS85A*, *ms36*	Trimnell, MR et al. 1999. MNL 73:49-50
ms37 male sterile 37	*ms*-SB177*, *ms37*, *male sterile 37*	Trimnell, MR et al. 1999. MNL 73:48
ms38 male sterile38	*ms30*, *ms38*, *ms*-WL87A*, *male sterile38*	Albertsen, MC et al. 1996. MNL 70:30
ms43 male sterile43	*ms43*, *male sterile43*, *ms29*	Golubovskaya, IN. 1979. Int Rev Cytol 58:247-290
ms45 male sterile45	*Group 6*, *male sterile45*, *ms*-6006*, *ms*-6040*, *ms*-BS1*, *ms*-BS2*, *ms*-BS3*, *ms45*, *ms45'-9301*	Albertsen, MC; Fox, TW; Trimnell, MR. 1993. Proc Annu Corn Sorghum Ind Res Conf 48:224-233
ms48 male sterile48	*male sterile48*, *ms*-6049*, *ms48*	Trimnell, M et al. 2002. MNL 76:38
ms5 male sterile5	: *ms*-6061*, *ms*-6048*,	Beadle, GW. 1932.

	ms-6062*, *male sterile5*, *ms5*	Genetics 17:413-431
ms50 male sterile50	*ms50*, *male sterile50*, *ms*-6055*, *ms*-6026*	Trimnell, M et al. 2002. MNL 76:39
ms7 male sterile7	*ms7*, *male sterile7*	Beadle, GW. 1932. Genetics 17:413-431
ms8 male sterile8	*male sterile8*, *ms8*	Beadle, GW. 1932. Genetics 17:413-431
ms9 male sterile9	*Group 5*, *male sterile9*, *ms9*	Beadle, GW. 1932. Genetics 17:413-431
ms49 male sterile49	*ms*-MB92*, *ms49*, *male sterile49*	Trimnell, M et al. 2002. MNL 76:38-39

U.S. Patent publication US 2008-0086783 A1 describes a male fertile gene referred to as "*BS92-7*" or "*BS7*" that is located on maize chromosome 7. *BS92-7* can be used in the systems described above, and other systems impacting male fertility.

U.S. Patent No. 5,750,868, issued 05-12-1998 describes a male fertile gene referred to as "*5126*" (SEQ ID NO: 24).

U.S. Patent No. 5,478,369 issued December 26, 1995 describes a male fertile gene referred to as "*MS45*".

U.S. Patent No. 7517975, issued April 14, 2009, describes a male fertile gene referred to as "*MS26*" (also known as SB200 or SBMu200) that is located on maize chromosome 1. MS26 can be used in the systems described above, and other systems impacting male fertility.

U.S. Patent publication US 2009-0038026 A1, published 02-05-2009, describes a male fertile gene referred to as "*Msca1*" or "*MS22*" that is located on maize chromosome 7 and encodes a protein critical to male fertility. Mutations referred to as *ms22* or *msca1* were first noted as phenotypically male sterile with anthers which did not extrude from the tassel and lacked sporogenous tissue. West and Albertsen (1985) *Maize Newsletter* 59:87; Neuffer et al. (1977) Mutants of maize. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. The mutant locus was originally referred to as *ms22* but was later changed to *msca1*, or male sterile converted anther. See Chaubal et al. "The transformation of anthers in the *msca1* mutant of maize" *Planta* (2003) 216:778-788.

In the context of this disclosure, a number of terms and abbreviations are used. The following definitions are provided.

The term "recognition sequence" or "recognition site" as used herein refers to a DNA sequence at which a double-strand break is induced in the plant cell genome by a double-strand-break-inducing agent. The terms "recognition sequence" and "recognition site" are used interchangeably herein.

The terms "target site", "target sequence", "target locus", "genomic target site", "genomic target sequence" as used interchangeably herein refer to a polynucleotide sequence in the genome of a plant cell that comprises a recognition sequence for a double-strand-break-inducing agent.

An "artificial target sequence" is a target sequence that has been introduced into the genome of a plant. Such an artificial target sequence can be identical in sequence to an endogenous or native target sequence in the genome of a plant but be located in a different position (*i.e.*, a non-endogenous or non-native position) in the genome of a plant.

An "endogenous target sequence" or "native target sequence" are used interchangeably herein to refer to a target sequence that is endogenous or native to the genome of a plant and is at the endogenous or native position of that target sequence in the genome of the plant.

An "altered target sequence" refers to a target sequence as disclosed herein that comprises at least one alteration of the invention when compared to non-altered target sequence. Such "alterations" of the invention include, for example: (i) replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, or (iv) any combination of (i) – (iii).

The term "double-strand-break-inducing agent" as used herein refers to any enzyme which produces a double-strand break in the target sequence. Producing the double-strand break in a target sequence or other DNA can be referred to herein as "cutting" or "cleaving" the target sequence or other DNA. In some embodiments of the invention, the double-strand-break-inducing agent has been engineered (or modified) to cut a specific endogenous target sequence, wherein the endogenous target sequence prior to being cut by the engineered double-strand-break-inducing agent was not a sequence that would have been

recognized by a native (non-engineered or non-modified) double-strand-break-inducing agent.

An "engineered double-strand-break-inducing agent" is any double-strand-break-inducing agent, including but not limited to, native or wild-type double-strand-break-inducing agents and previously engineered double-strand-break-inducing agent that has been modified to produce a double-strand break at a target sequence of interest which has a different nucleotide sequence than the original target sequence of the double-strand-break-inducing agent prior to its modification. Preferably, an engineered double-strand-break-inducing agent of the invention is no longer capable of making a double-strand break at the original target sequence.

The term "endonuclease" refers to any enzyme that cleaves the phosphodiester bond within a polynucleotide chain, and includes restriction endonucleases that cleave DNA as specific sites without damaging the bases. Restriction endonucleases include Type I, Type II, Type III, and Type IV endonucleases, which further include subtypes. In the Type I and Type III systems, both the methylase and restriction activities are contained in a single complex.

Type I and Type III restriction endonucleases recognize specific recognition sites, but typically cleave at a variable position from the recognition site, which can be hundreds of base pairs away from the recognition site. In Type II systems the restriction activity is independent of any methylase activity, and cleavage typically occurs at specific sites within or near to the recognition site. Most Type II enzymes cut palindromic sequences, however Type IIa enzymes recognize non-palindromic recognition sites and cleave outside of the recognition site, Type IIb enzymes cut sequences twice with both sites outside of the recognition site, and Type IIs enzymes recognize an asymmetric recognition site and cleave on one side and at a defined distance of about 1-20 nucleotides from the recognition site.

Type IV restriction enzymes target methylated DNA. Restriction enzymes are further described and classified, for example in the REBASE database (webpage at rebase.neb.com; Roberts, *et al.*, (2003) *Nucleic Acids Res* 31:418-20), Roberts, *et al.*, (2003) *Nucleic Acids Res* 31:1805-12, and Belfort, *et al.*,

(2002) in *Mobile DNA II*, pp. 761-783, Eds. Craigie, *et al.*, (ASM Press, Washington, DC).

Endonucleases also include meganucleases, also known as homing endonucleases (HEases), which like restriction endonucleases, bind and cut at a specific recognition sequence, however the recognition sites for meganucleases are typically longer, about 18 bp or more. Meganucleases have been classified into four families based on conserved sequence motifs; the families are the LAGLIDADG, GIY-YIG, H-N-H, and His-Cys box families. These motifs participate in the coordination of metal ions and hydrolysis of phosphodiester bonds. HEases are notable for their long recognition sites, and for tolerating some sequence polymorphisms in their DNA substrates. The naming convention for meganuclease is similar to the convention for other restriction endonuclease. Meganucleases are also characterized by prefix F-, I-, or PI- for enzymes encoded by free-standing ORFs, introns, and inteins, respectively. For example, intron-, intein-, and freestanding gene encoded meganuclease from *Saccharomyces cerevisiae* are denoted I-SceI, PI-SceI, and F-SceII, respectively. Meganuclease domains, structure and function are known, see for example, Guhan and Muniyappa (2003) *Crit Rev Biochem Mol Biol* 38:199-248; Lucas, *et al.*, (2001) *Nucleic Acids Res* 29:960-9; Jurica and Stoddard, (1999) *Cell Mol Life Sci* 55:1304-26; Stoddard, (2006) *Q Rev Biophys* 38:49-95; and Moure, *et al.*, (2002) *Nat Struct Biol* 9:764. In some examples an engineered meganuclease is used. Methods for modifying the kinetics, cofactor interactions, expression, optimal conditions, and/or recognition site specificity, and screening for activity are known. See, for example, Epinat, *et al.*, (2003) *Nucleic Acids Res* 31:2952-62; Chevalier, *et al.*, (2002) *Mol Cell* 10:895-905; Gimble, *et al.*, (2003) *Mol Biol* 334:993-1008; Seligman, *et al.*, (2002) *Nucleic Acids Res* 30:3870-9; Sussman, *et al.*, (2004) *J Mol Biol* 342:31-41; Rosen, *et al.*, (2006) *Nucleic Acids Res* 34:4791-800; Chames, *et al.*, (2005) *Nucleic Acids Res* 33:e178; Smith, *et al.*, (2006) *Nucleic Acids Res* 34:e149; Gruen, *et al.*, (2002) *Nucleic Acids Res* 30:e29; Chen and Zhao, (2005) *Nucleic Acids Res* 33:e154; WO2005105989; WO2003078619; WO2006097854; WO2006097853; WO2006097784; and WO2004031346.

The endonuclease can be a modified endonuclease that binds a non-native or exogenous recognition sequence and does not bind a native or

endogenous recognition sequence. Modification of the endonuclease can be as little as one nucleotide. A modified endonuclease is not capable of making a double-strand break within a wild-type target sequence. A wild-type (*i.e.*, prior to being modified) endonuclease is capable of making a double-strand break within the wild-type target sequence.

The endonuclease can be provided via a polynucleotide encoding the endonuclease. Such a polynucleotide encoding an endonuclease can be modified to substitute codons having a higher frequency of usage in a plant, as compared to the naturally occurring polynucleotide sequence. For example the polynucleotide encoding the endonuclease can be modified to substitute codons having a higher frequency of usage in a maize or soybean plant, as compared to the naturally occurring polynucleotide sequence.

The term "engineered endonuclease" is any endonuclease that has been engineered (or modified) to cut a specific endogenous target sequence, wherein the endogenous target sequence prior to being cut by the engineered endonuclease was not a sequence that would have been recognized by a native (non-engineered or non-modified) endonuclease.

In some embodiments of the invention, the engineered endonuclease is an engineered MS26 endonuclease, an engineered MS26+ endonuclease, engineered MS26++ endonuclease or an engineered MS45 endonuclease .

As used herein, "physically linked," and "in physical linkage", and "genetically linked" are used to refer to any two or more genes, transgenes, native genes, mutated genes, alterations, target sites, markers, and the like that are part of the same DNA molecule or chromosome.

As used herein, a "polynucleotide of interest" within a genomic region of interest is any coding and/or non-coding portion of the genomic region of interest including, but not limited to, a transgene, a native gene, a mutated gene, and a genetic marker such as, for example, a single nucleotide polymorphism (SNP) marker and a simple sequence repeat (SSR) marker.

"Open reading frame" is abbreviated ORF.

As used herein, "nucleic acid" means a polynucleotide and includes a single or a double-stranded polymer of deoxyribonucleotide or ribonucleotide bases. Nucleic acids may also include fragments and modified nucleotides. Thus, the terms "polynucleotide", "nucleic acid sequence", "nucleotide sequence"

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

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

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

Polynucleotide and polypeptide sequences, variants thereof, and the structural relationships of these sequences can be described by the terms "homology", "homologous", "substantially identical", "substantially similar" and "corresponding substantially" which are used interchangeably herein. These refer to polypeptide or nucleic acid fragments wherein changes in one or more amino acids or nucleotide bases do not affect the function of the molecule, such as the ability to mediate gene expression or to produce a certain phenotype.

These terms also refer to modification(s) of nucleic acid fragments that do not substantially alter the functional properties of the resulting nucleic acid fragment relative to the initial, unmodified fragment. These modifications include deletion, substitution, and/or insertion of one or more nucleotides in the nucleic acid fragment.

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

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

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

Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion

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

Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth, *et al.*, (1984) *Anal Biochem* 138:267-284: $T_m = 81.5°C + 16.6 (\log M) + 0.41 (\%GC) - 0.61 (\%\ form) - 500/L$; where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The T_m is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe. T_m is reduced by about 1°C for each 1% of mismatching; thus, T_m, hybridization and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with $\geq$90% identity are sought, the T_m can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3 or 4°C lower than the thermal melting point (T_m); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9 or 10°C lower than the thermal melting point (T_m); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15 or 20°C lower than the thermal melting point (T_m). Using the equation, hybridization and wash compositions, and desired T_m, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently

described. If the desired degree of mismatching results in a T_m of less than 45°C (aqueous solution) or 32°C (formamide solution) it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes,* Part I, Chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York (1993); and *Current Protocols in Molecular Biology*, Chapter 2, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995). Hybridization and/or wash conditions can be applied for at least 10, 30, 60, 90, 120 or 240 minutes.

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

The term "percentage of sequence identity" refers to the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the results by 100 to yield the percentage of sequence identity. Useful examples of percent sequence identities include, but are not limited to, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%, or any integer percentage from 50% to 100%. These identities can be determined using any of the programs described herein.

Sequence alignments and percent identity or similarity calculations may be determined using a variety of comparison methods designed to detect homologous sequences including, but not limited to, the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI).

Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" will mean any set of values or parameters that originally load with the software when first initialized.

The "Clustal V method of alignment" corresponds to the alignment method labeled Clustal V (described by Higgins and Sharp, (1989) *CABIOS* 5:151-153; Higgins, *et al.*, (1992) *Comput Appl Biosci* 8:189-191) and found in the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). For multiple alignments, the default values correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10. Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal method are KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids these parameters are KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. After alignment of the sequences using the Clustal V program, it is possible to obtain a "percent identity" by viewing the "sequence distances" table in the same program.

The "Clustal W method of alignment" corresponds to the alignment method labeled Clustal W (described by Higgins and Sharp, (1989) *CABIOS* 5:151-153; Higgins, *et al.*, (1992) *Comput Appl Biosci* 8:189-191) and found in the MegAlign™ v6.1 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Default parameters for multiple alignment (GAP PENALTY=10, GAP LENGTH PENALTY=0.2, Delay Divergen Seqs (%)=30, DNA Transition Weight=0.5, Protein Weight Matrix=Gonnet Series, DNA Weight Matrix=IUB). After alignment of the sequences using the Clustal W program, it is possible to obtain a "percent identity" by viewing the "sequence distances" table in the same program.

Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using GAP Version 10 (GCG, Accelrys, San Diego, CA) using the following parameters: % identity and % similarity for a nucleotide sequence using a gap creation penalty weight of 50 and a gap length extension penalty weight of 3, and the nwsgapdna.cmp scoring matrix; % identity

and % similarity for an amino acid sequence using a GAP creation penalty weight of 8 and a gap length extension penalty of 2, and the BLOSUM62 scoring matrix (Henikoff and Henikoff, (1989) *Proc. Natl. Acad. Sci. USA* 89:10915). GAP uses the algorithm of Needleman and Wunsch, (1970) *J Mol Biol* 48:443-53, to find an alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP considers all possible alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps, using a gap creation penalty and a gap extension penalty in units of matched bases.

"BLAST" is a searching algorithm provided by the National Center for Biotechnology Information (NCBI) used to find regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches to identify sequences having sufficient similarity to a query sequence such that the similarity would not be predicted to have occurred randomly. BLAST reports the identified sequences and their local alignment to the query sequence.

It is well understood by one skilled in the art that many levels of sequence identity are useful in identifying polypeptides from other species or modified naturally or synthetically wherein such polypeptides have the same or similar function or activity. Useful examples of percent identities include, but are not limited to, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% or 95%, or any integer percentage from 50% to 100%. Indeed, any integer amino acid identity from 50% to 100% may be useful in describing the present invention, such as 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived

from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature, or at a different genetic locus than that found in nature. A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes.

A "mutated gene" is a native gene that has been altered through human intervention. Such a "mutated gene" has a sequence that differs from the sequence of the corresponding native gene by at least one nucleotide addition, deletion, or substitution. In certain embodiments of the invention, the mutated gene comprises an alteration that results from a double-strand-break-inducing agent as disclosed herein.

A "transgene" is a gene that has been introduced into the genome by a transformation procedure. A transgene can, for example encode one or more proteins or RNA that is not translated into protein. However, a transgene of the invention need not encode a protein and/or non-translated RNA. In certain embodiments of the invention, the transgene comprises one or more chimeric genes, including chimeric genes comprising, for example, a gene of interest, phenotypic marker, a selectable marker, and a DNA for gene silencing.

As used herein, a "targeted modification" is a modification in a target sequence in the genome of an organism that was made by altering a target sequence within the native gene using a method involving a double-strand-break-inducing agent that is capable of inducing a double-strand break in the DNA of the target sequence as disclosed herein or known in the art. A "targeted mutation" is a mutation in a native gene that was made by altering a target sequence within the native gene using a method involving a double-strand-break-inducing agent that is capable of inducing a double-strand break in the DNA of the target sequence as disclosed herein or known in the art. A "targeted mutation" is one type of "targeted modification."

When used herein with respect to DNA, genes, and other nucleic acids, the terms "alteration," modification," and "mutation" are to be considered equivalent terms unless it is apparent from the context that a different meaning is intended for any one or more of these terms.

A "null mutation" is a mutation in a gene that leads to it not being transcribed into RNA and/or translated into a functional protein product. An allele that comprises the null mutation is referred to a "null allele." A null mutation in a gene can be caused, for example, by an alteration in the gene including (i) replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, or (iv) any combination of (i) – (iii).

As used herein, a "male fertility gene" is a gene that is critical to steps leading to and including microsporogenesis, the term applied to the entire process of pollen formation. These genes can be collectively referred to as male fertility genes (or, alternatively, male sterility genes). The terms "male fertility gene", "male fertile gene", "male sterility gene" and "male sterile gene" are used interchangeably.

A "fertile plant" is a plant that is capable of producing a progeny plant. In certain embodiments of the invention, a fertile plant is a plant that produces viable male and female gametes and is self fertile. Such a self-fertile plant can produce a progeny plant without the contribution from any other plant of a gamete and the genetic material contained therein. Other embodiments of the invention can involve the use of a plant that is not self fertile because the plant does not produce male or female gametes that are viable or otherwise capable of fertilization. As used herein, a "male sterile plant" is a plant that does not produce male gametes that are viable or otherwise capable of fertilization. As used herein, a "female sterile plant" is a plant that does not produce female gametes that are viable or otherwise capable of fertilization. It is recognized that male-sterile and female-sterile plants can be female-fertile and male-fertile, respectively. It is further recognized that a male-fertile (but female-sterile) plant can produce viable progeny when crossed with a female-fertile plant, and that a female-fertile (but male-sterile) plant can produce viable progeny when crossed with a male fertile plant.

The term "genome" as it applies to a plant cell encompasses not only chromosomal DNA found within the nucleus, but organelle DNA found within subcellular components (e.g., mitochondria, or plastid) of the cell.

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

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

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

"Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. An "enhancer" is a DNA sequence that can stimulate promoter activity, and may be an innate element of the promoter or a heterologous element inserted to enhance the activity or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, and/or comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of some variation may have identical promoter activity. Promoters that cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are constantly being discovered; numerous examples

may be found in the compilation by Okamuro and Goldberg, (1989) In *The Biochemistry of Plants*, Vol. 115, Stumpf and Conn, eds (New York, NY: Academic Press), pp. 1-82.

"Translation leader sequence" refers to a polynucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (e.g., Turner and Foster, (1995) *Mol Biotechnol* 3:225-236).

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

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

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

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

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

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

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

The terms "plasmid", "vector" and "cassette" refer to an extra chromosomal element often carrying genes that are not part of the central metabolism of the cell, and usually in the form of double-stranded DNA. Such elements may be autonomously replicating sequences, genome integrating

sequences, phage, or nucleotide sequences, in linear or circular form, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a polynucleotide of interest into a cell. "Transformation cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that facilitates transformation of a particular host cell. "Expression cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that allow for expression of that gene in a foreign host.

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

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

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

"Mature" protein refers to a post-translationally processed polypeptide (i.e., one from which any pre- or propeptides present in the primary translation product have been removed). "Precursor" protein refers to the primary product of translation of mRNA (i.e., with pre- and propeptides still present). Pre- and propeptides may be but are not limited to intracellular localization signals.

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

As used herein, "transgenic" refers to a plant or a cell which comprises within its genome a heterologous polynucleotide. Typically, the heterologous polynucleotide is stably integrated within the genome such that the polynucleotide is passed on to successive generations. The heterologous

polynucleotide may be integrated into the genome alone or as part of an expression construct. Transgenic is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic. The term "transgenic" as used herein does not encompass the alteration of the genome (chromosomal or extra-chromosomal) by conventional plant breeding methods or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation.

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

The present invention finds use in the production of hybrid plants. Mutations that cause male sterility in plants are useful in hybrid seed production methods for crop plants such as, for example, maize. The use of male sterile plants in hybrid maize seed production eliminates the need for the labor-intensive removal of male flowers (also known as de-tasseling when maize is the plant) from the maternal parent plants used to produce the hybrid seed. Mutations that cause male sterility in maize have been produced by a variety of methods such as X-rays or UV-irradiations, chemical treatments, or transposable element insertions (*ms23, ms25, ms26, ms32*) (Chaubal *et al.* (2000) *Am J Bot*

87:1193-1201). However, such methods are random mutagenesis methods that induce mutations randomly throughout the genome and not just in the gene of interest. Typically, with such random mutagenesis methods, it requires considerable effort to identify a plant that contains a mutation in the gene of interest and it is by no means certain that such a plant will be identified. Furthermore, with random mutagenesis methods, each plant tested is likely to carry multiple mutations. Therefore, a plant that is identified with the mutation in the gene of interest must be backcrossed for several or more generations to eliminate the undesired mutations that are not within the gene of interest.

In contrast to such random mutagenesis methods, the present invention provides improved methods for producing male sterile plants by making targeted mutations or alterations in a male fertility gene of interest in any plant, particularly any crop plant. Because the mutations are targeted to the male fertility gene of interest, it is not necessary to screen a population of thousands of plants carrying random mutations in order to identify a plant with a mutation in the male fertility gene of interest. Furthermore, undesired mutations outside of the gene of interest are rare, and if they occur are all in a particular plant produced by the methods of the present invention. Therefore, the need to backcross a plant to remove undesired mutations that are not in the gene of interest is eliminated or at least reduced.

In a first aspect, the present invention provides methods for making a targeted modification in a male fertility gene in the genome of a plant. The methods involve contacting at least one plant cell comprising a target sequence in a male fertility gene with an engineered double-strand-break-inducing agent capable of inducing a double-strand break at the target sequence and then identifying at least one cell comprising an alteration in its genome at the target sequence. The methods can further comprise regenerating a fertile plant or a male sterile plant comprising the alteration.

The methods involve the use of an engineered double-strand-break-inducing agent that is capable of inducing a double-strand break in DNA comprising the target sequence in a male fertility gene of interest. The methods of the invention do not depend on a particular engineered double-strand-break-inducing agent but only that the engineered double-strand-break-inducing agent is capable of inducing a double-strand break in DNA in a target sequence of the

invention. Any such engineered double-strand-break-inducing agent that is disclosed herein or known in the art can be used in the methods of the present invention. Furthermore, the invention encompasses the use of any engineered double-strand-break-inducing agent that is made by methods disclosed herein or known in the art.

The methods of the invention comprise contacting at least one plant cell comprising a target sequence in a male fertility gene with an engineered double-strand-break-inducing agent capable of inducing a double-strand break at the target sequence. Such contacting can involve, for example, introducing a polypeptide comprising the double-strand-break-inducing agent directly into the plant cell or introducing into the plant cell a nucleic acid construct comprising a nucleotide sequence encoding the engineered double-strand-break-inducing agent, whereby the engineered double-strand-break-inducing agent is produced in the cell. The nucleic acid construct can comprise, for example, a promoter operably linked to a nucleotide sequence encoding an engineered double-strand-break-inducing agent of the invention. Any promoter disclosed herein or known in the art that can drive the expression of the operably linked nucleotide sequence in the plant cell can be used in the methods of the present invention.

If desired or necessary to achieve nuclear localization of the engineered double-strand-break-inducing agent, the nucleotide construct can further comprise an operably linked nucleotide sequence encoding a nuclear localization signal. Any nuclear localization signal that can facilitate nuclear localization of the engineered double-strand-break-inducing agent that is disclosed herein or known in the art can be used in the methods of the present invention. Such nuclear localization signals include, but are not limited to, a nuclear localization signal comprising an amino acid sequence set forth in SEQ ID NO: 2, 3 or 21.

The methods of the invention involve making an alteration at the target sequence. Such an alteration includes, for example, a replacement of at least one nucleotide, a deletion of at least one nucleotide, an insertion of at least one nucleotide, and any combination of one or more replacements, deletions, and insertions.

In one embodiment of the invention, the alteration is an insertion of a transgene. Such a transgene can comprise, for example, one, two, three, four,

or more polynucleotides of interest. If desired, a polynucleotide of interest can be operably linked to promoter that is capable of driving the expression of the polynucleotide of interest in a plant. Polynucleotides of interest include, but are not limited to, a phenotypic marker and an RNA or protein providing an agronomic advantage to the plant.

In another embodiment of the invention, the alteration in the target sequence of the male fertility gene is a null mutation. When a plant is homozygous for such a null mutation (*i.e.*, has two null alleles at the male fertility gene of interest), the plant is male sterile. Such a null mutation can result from any of the alterations disclosed hereinabove including, for example, the insertion of a transgene. In certain embodiments of the invention, the transgene comprises a phenotypic marker, particularly a selectable marker. It is recognized that when the null mutation is caused by the insertion of a transgene comprising a phenotypic marker, particularly a selectable marker, identifying plants comprising at least one null allele at the male fertility gene of interest can comprise identifying a plant comprising the phenotypic maker, particularly the selectable marker.

The methods of the invention can further comprise selfing the fertile plant comprising the alteration in the male fertility gene and selecting a progeny plant resulting therefrom, wherein said progeny plant is homozygous for the alteration. In an embodiment of the invention, the methods further comprise selfing the fertile plant comprising an alteration that is a null mutation in the male fertility gene and selecting a progeny plant resulting therefrom, wherein said progeny plant is homozygous for the alteration and is male sterile.

In another embodiment of the invention, the methods of the invention further comprise crossing a first fertile plant comprising a null mutation in the male fertility gene with a second fertile plant comprising a null mutation in the male fertility gene and selecting a progeny plant resulting therefrom, wherein the progeny plant is male sterile. Both the first and second male sterile plants can be produced by the methods as disclosed herein or can be descendants of a fertile plant that is produced by the methods as disclosed herein. The first and second male sterile plants can comprise the same null mutation in the male fertility gene. Alternatively, the first male sterile plant can comprise a first null mutation in the male fertility gene, and the second male sterile plant can

comprise a second null mutation in the male fertility gene wherein the first null mutation is not identical to the second null mutation. In one embodiment of the invention, the first null mutation comprises the insertion of a first transgene comprising a first phenotypic marker, particularly a first selectable marker, and the second null mutation comprises the insertion of a second transgene comprising a second phenotypic marker, particularly a second selectable marker. Thus, when the first fertile plant is crossed to the second fertile plant, male sterile progeny which comprise both the first null mutation and the second null mutation can be identified as those progeny plants comprising both the first and second phenotypic markers.

The methods of the invention can be employed to make targeted modifications in any male fertility gene in a plant and thus provide for the production of male sterile plants in any plant comprising a male fertility gene. Male fertility genes of interest include, but are not limited to, the genes disclosed in Table1 and *MS26, MS45, BS92-7, 5126* and *Msca1*.

In one embodiment of the invention, the methods of the invention involve making a targeted modification in the male fertility gene, *MS26*, in the genome of a plant, such as for example, a maize plant or a sorghum plant. The methods involve contacting at least one plant cell comprising a target sequence in the *MS26* gene with an engineered double-strand-break-inducing agent capable of inducing a double-strand break at the target sequence and then identifying at least one cell comprising an alteration, particularly a null mutation, in its genome at the target sequence. The methods further comprise regenerating a fertile plant comprising the alteration.

An example of a target sequence in the *MS26* gene that can be used in this embodiment is set forth in TS-MS26 (SEQ ID NO: 1). In this example, any double-strand-break-inducing agent capable of inducing a double-strand break at this target sequence can be used.

The engineered double-strand-break-inducing agent that is cable of inducing a double strand break in the TS-MS26 target sequence comprising SEQ ID NO: 1 can be introduced into a plant as a nucleotide construct comprising a promoter operably linked to a nucleotide sequence encoding the engineered double-strand-break-inducing agent. Any promoter disclosed herein or known in the art that can drive expression of the operably linked nucleotide

sequence encoding the engineered double-strand-break-inducing agent in the plant cell can be used in the methods of the present invention. The nucleotide sequence encoding the engineered double-strand-break-inducing agent can be selected from the group consisting of: the nucleotide sequences set forth in SEQ ID NOS: 4 to 7; and a nucleotide sequence having at least 80% nucleotide sequence identity to at least one nucleotide sequence selected from the group consisting of the nucleotide sequences set forth in SEQ ID NOS: 4 to 7. The nucleic acid construct can further comprise an operably linked nuclear localization signal.

In one embodiment, the plant produced by the methods of the invention is a sorghum plant comprising a targeted modification in the male fertility gene *MS26*, wherein the *MS26* gene comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77 and 78.

In another embodiment of the invention, the methods of the invention involve making a targeted modification in the male fertility gene, *MS45*, in the genome of a plant, such as for example, a maize plant. The methods involve contacting at least one plant cell comprising a target sequence in the *MS45* gene with an engineered double-strand-break-inducing agent capable of inducing a double-strand break at the target sequence and then identifying at least one cell comprising an alteration, particularly a null mutation, in its genome at the target sequence. The methods further comprise regenerating a fertile plant comprising the alteration.

An example of a target sequence in the *MS45* gene that can be used in this embodiment is set forth in the TS-MS45 target site (SEQ ID NO: 20). The nucleotide sequence encoding the engineered double-strand-break-inducing agent can be selected from the group consisting of: the nucleotide sequences set forth in SEQ ID NOS: 22, 23, and 34; and a nucleotide sequence having at least 80% nucleotide sequence identity to at least one nucleotide sequence selected from the group consisting of the nucleotide sequences set forth in SEQ ID NOS: 22, 23, and 34. The nucleic acid construct can further comprise an operably linked nuclear localization signal.

In this example, any double-strand-break-inducing agent capable of inducing a double-strand break at this target sequence can be used : in one

embodiment a first polypeptide encoded by the nucleotide sequence set forth in SEQ ID NO: 22 and a second polypeptide encoded by the nucleotide sequence set forth in SEQ ID NO: 23 can be used. The engineered double-strand-break-inducing agent that is capable of inducing double strand breaks in the TS-MS45 target site (SEQ ID NO: 20) can be introduced into a plant as a nucleotide construct comprising a promoter operably linked to a nucleotide sequence encoding the engineered double-strand-break-inducing agent. Any promoter disclosed herein or known in the art that can drive expression of the operably linked nucleotide sequence encoding the engineered double-strand-break-inducing agent in the plant cell can be used in the methods of the present invention. The nucleotide sequence encoding the engineered double-strand-break-inducing agents includes, but is not limited to, the nucleotide sequence set forth in SEQ ID NO: 22, 23, or 34. The nucleic acid construct can further comprise an operably linked nuclear localization signal.

In a second aspect, the present invention provides a plant comprising in its genome at least one male fertility gene with a targeted modification and descendants thereof that comprise at least one of the male fertility genes with a targeted modification. Such targeted modifications comprise the alterations in a male fertility gene as disclosed hereinabove. The plants of the invention can be made by the methods disclosed herein for making a targeted modification in the genome of a plant including, but not limited to, fertile plants that are heterozygous for a null mutation in a male fertility gene, male sterile plants that are homozygous for a null mutation in the male fertility gene, and plants comprising an alteration in a male fertility gene, wherein the alteration comprises the insertion of a transgene. In one embodiment of the invention, a plant of the invention comprises in its genome the insertion of a transgene in the male fertility gene and such insertion is a null mutation.

In a third aspect, the present invention provides isolated nucleic acid molecules comprising a male fertility gene with at least one targeted modification. Such targeted modifications comprise one or more alterations in a male fertility gene as disclosed hereinabove.

In a fourth aspect, the present invention provides isolated plant-optimized nucleic acid molecules encoding engineered double-strand-break-inducing agents, particularly an engineered double-strand-break-inducing agent derived

from I-CreI, more particularly an engineered double-strand-break-inducing agent derived from I-CreI that is capable of inducing double-strand breaks in DNA in a TS-MS26 or TS-MS45 target sequence, most particularly an engineered double-strand-break-inducing agent derived from I-CreI that encodes an engineered MS26 endonuclease or an engineered MS45 endonuclease. Nucleic acid molecules of the invention include, but are not limited to, nucleic acid molecules comprising the nucleotide sequence set forth in SEQ ID NO: 4, 5, 6, 7, 22, 23, or 34, nucleotide sequences and fragments and variants thereof that encode an engineered MS26 endonuclease, an engineered MS26+ endonuclease, an engineered MS26++ endonuclease or an engineered MS45 endonuclease. In one embodiment of the invention, the nucleic acid molecules comprise nucleotide sequences that have been optimized for expression in a plant of interest.

Compositions of the invention include endonucleases that are double-strand-break-inducing agents capable of inducing a double-strand break in a specific recognition or target sequence in a DNA molecule. In particular, the present invention provides for isolated polynucleotides comprising nucleotide sequences encoding endonculeases. The invention encompasses isolated or substantially purified polynucleotide or protein compositions. An "isolated" or "purified" polynucleotide or protein, or biologically active portion thereof, is substantially or essentially free from components that normally accompany or interact with the polynucleotide or protein as found in its naturally occurring environment. Thus, an isolated or purified polynucleotide or protein is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. Optimally, an "isolated" polynucleotide is free of sequences (optimally protein encoding sequences) that naturally flank the polynucleotide (i.e., sequences located at the 5' and 3' ends of the polynucleotide) in the genomic DNA of the organism from which the polynucleotide is derived. For example, in various embodiments, the isolated polynucleotide can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotide sequence that naturally flank the polynucleotide in genomic DNA of the cell from which the polynucleotide is derived. A protein that is substantially free of cellular material includes preparations of protein having less

than about 30%, 20%, 10%, 5%, or 1% (by dry weight) of contaminating protein. When the protein of the invention or biologically active portion thereof is recombinantly produced, optimally culture medium represents less than about 30%, 20%, 10%, 5%, or 1% (by dry weight) of chemical precursors or non-protein-of-interest chemicals.

Fragments and variants of the disclosed polynucleotides and proteins encoded thereby are also encompassed by the present invention. By "fragment" is intended a portion of the polynucleotide or a portion of the amino acid sequence and hence protein encoded thereby. Fragments of a polynucleotide may encode protein fragments that retain biological activity of the exemplified protein and hence comprise target-sequence-specific endonuclease active, particularly nuclease activity at the TS-MS26 or TS-MS45 target site, as described herein. Thus, fragments of a nucleotide sequence may range from at least about 200 nucleotides, about 400 nucleotides, and up to the full-length polynucleotide encoding the proteins of the invention.

"Variants" is intended to mean substantially similar sequences. For polynucleotides, a variant comprises a polynucleotide having deletions (i.e., truncations) at the 5' and/or 3' end; deletion and/or addition of one or more nucleotides at one or more internal sites in the native polynucleotide; and/or substitution of one or more nucleotides at one or more sites in a polynucleotide sequence disclosed herein. For polynucleotides, conservative variants include those sequences that, because of the degeneracy of the genetic code, encode the amino acid sequence of one of the endonuclease polypeptides of the invention. Variant polynucleotides also include synthetically derived polynucleotides, such as those generated, for example, by using site-directed mutagenesis but which still encode an endonuclease protein of the invention. Generally, variants of a particular polynucleotide of the invention will have at least about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to that particular polynucleotide as determined by sequence alignment programs and parameters as described elsewhere herein.

Variants of a particular polynucleotide of the invention (i.e., the reference polynucleotide) can also be evaluated by comparison of the percent sequence identity between the polypeptide encoded by a variant polynucleotide and the polypeptide encoded by the reference polynucleotide. Percent sequence identity

between any two polypeptides can be calculated using sequence alignment programs and parameters described elsewhere herein. Where any given pair of polynucleotides of the invention is evaluated by comparison of the percent sequence identity shared by the two polypeptides they encode, the percent sequence identity between the two encoded polypeptides is at least about 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity.

"Variant" protein is intended to mean a protein derived from the native protein by deletion (so-called truncation) of one or more amino acids at the N-terminal and/or C-terminal end of the native protein; deletion and/or addition of one or more amino acids at one or more internal sites in the native protein; or substitution of one or more amino acids at one or more sites in the native protein. Variant proteins encompassed by the present invention are biologically active, that is they continue to possess the desired biological activity of the native protein, that is nuclease activity at the TS-MS26 or TS-MS45 target site, as described herein. Biologically active variants of an endonuclease protein of the invention will have at least about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the amino acid sequence for the native protein as determined by sequence alignment programs and parameters described elsewhere herein. A biologically active variant of a protein of the invention may differ from that protein by as few as 1-15 amino acid residues, as few as 1-10, such as 6-10, as few as 5, as few as 4, 3, 2, or even 1 amino acid residue.

The proteins of the invention may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants and fragments of the endonuclease proteins can be prepared by mutations in the DNA. Methods for mutagenesis and polynucleotide alterations are well known in the art. See, for example, Kunkel (1985) *Proc. Natl. Acad. Sci. USA* 82:488-492; Kunkel *et al.* (1987) *Methods in Enzymol.* 154:367-382; U.S. Patent No. 4,873,192; Walker and Gaastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan Publishing Company, New York) and the references cited therein. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found in the

model of Dayhoff *et al.* (1978) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, D.C.), herein incorporated by reference. Conservative substitutions, such as exchanging one amino acid with another having similar properties, may be optimal.

Variant polypeptides will continue to possess the desired nuclease activity at the TS-MS26 or TS-MS45 target site. Obviously, the mutations that will be made in the DNA encoding the variant must not place the sequence out of reading frame and optimally will not create complementary regions that could produce secondary mRNA structure. See, EP Patent Application Publication No. 75,444.

The deletions, insertions, and substitutions of the protein sequences encompassed herein are not expected to produce radical changes in the characteristics of the protein. However, when it is difficult to predict the exact effect of the substitution, deletion, or insertion in advance of doing so, one skilled in the art will appreciate that the effect will be evaluated by routine screening assays (Lucas *et al.* 2001 (*Nucl. Acids Res.* 29: 960-969).

Variant polynucleotides and proteins also encompass sequences and proteins derived from a mutagenic and recombinogenic procedure such as DNA shuffling. With such a procedure, one or more different endonuclease sequences can be manipulated to create a new endonuclease possessing the desired properties. In this manner, libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined *in vitro* or *in vivo*. Strategies for such DNA shuffling are known in the art. See, for example, Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91:10747-10751; Stemmer (1994) *Nature* 370:389-391; Crameri *et al.* (1997) *Nature Biotech.* 15:436-438; Moore *et al.* (1997) *J. Mol. Biol.* 272:336-347; Zhang *et al. (1997) Proc. Natl. Acad. Sci. USA* 94:4504-4509; Crameri *et al.* (1998) *Nature* 391:288-291; and U.S. Patent Nos. 5,605,793 and 5,837,458.

The methods of the present invention involve the use of one or more double-strand-break-inducing agents. A double-strand-break-inducing agent of the present invention is any agent that recognizes and/or binds to a specific polynucleotide recognition sequence to produce a break in the target sequence at or near the recognition sequence. Examples of double-strand-break-inducing

agents include, but are not limited to, endonucleases, site-specific recombinases, transposases, topoisomerases, TAL effector nucleases, and zinc finger nucleases, and include modified derivatives, variants, and fragments thereof.

A recognition sequence is any polynucleotide sequence that is specifically recognized and/or bound by a double-strand-break-inducing agent. The length of the recognition site sequence can vary, and includes, for example, sequences that are at least 4, 6, 8, 10, 12, 14, 16, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70 or more nucleotides in length.

It is possible that the recognition site could be palindromic, that is, the sequence on one strand reads the same in the opposite direction on the complementary strand. The nick/cleavage site could be within the recognition sequence or the nick/cleavage site could be outside of the recognition sequence. In another variation, the cleavage could occur at nucleotide positions immediately opposite each other to produce a blunt end cut or, in other cases, the incisions could be staggered to produce single-stranded overhangs, also called "sticky ends", which can be either 5' overhangs, or 3' overhangs. The recognition sequence can be endogenous or exogenous. When the recognition site is an endogenous sequence, it may be a recognition sequence recognized by a naturally-occurring, or native double-strand-break-inducing agent. Alternatively, an endogenous recognition site could be recognized and/or bound by a modified or engineered double-strand-break-inducing agent designed or selected to specifically recognize the endogenous recognition sequence to produce a double-strand break. A modified double-strand–break-inducing agent can be derived from a native, naturally-occurring double-strand-break-inducing agent or it could be artificially created or synthesized.

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

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

Any meganuclease can be used as a double-strand break inducing agent including, but not limited to, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TliI, I-PpoI, PI-PspI, F-SceI, F-SceII, F-SuvI, F-TevI, F-TevII, I-AmaI, I-AniI, I-ChuI, I-CmoeI, I-CpaI, I-CpaII, I-CsmI, I-CvuI, I-CvuAIP, I-DdiI, I-DdiII, I-DirI, I-DmoI, I-HmuI, I-HmuII, I-HsNIP, I-LlaI, I-MsoI, I-NaaI, I-NanI, I-NclIP, I-NgrIP, I-NitI, I-NjaI, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbpIP, I-SpBetaIP, I-ScaI, I-SexIP, I-SneIP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MtuI, PI-MtuHIP PI-MtuHIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-TfuI, PI-TfuII, PI-ThyI, PI-TliI, PI-TliII, or any variant or derivative thereof.

A site-specific recombinase, also referred to as a recombinase, is a polypeptide that catalyzes conservative site-specific recombination between its compatible recombination sites, and includes native polypeptides as well as derivatives, variants and/or fragments that retain activity, and native polynucleotides, derivatives, variants, and/or fragments that encode a recombinase that retains activity.

One step in the recombination process involves polynucleotide cleavage at or near the recognition site. This cleaving activity can be used to produce a double-strand break. For reviews of site-specific recombinases and their recognition sites, see, Sauer (1994) *Curr Op Biotechnol* 5:521-7; and Sadowski, (1993) *FASEB* 7:760-7.

The Integrase family of recombinases has over one hundred members and includes, for example, FLP, Cre, lambda integrase, and R. The Integrase family has been grouped into two classes based on the structure of the active sites, serine recombinases and tyrosine recombinases. The tyrosine family, which includes Cre, FLP, SSV1, and lambda (λ) integrase, uses the catalytic tyrosine's hydroxyl group for a nucleophilic attack on the phosphodiester bond of the DNA. Typically, members of the tyrosine family initially nick the DNA, which later forms a double-strand break. In the serine recombinase family, which includes phiC31 (ΦC31) integrase, a conserved serine residue forms a covalent link to the DNA target site (Grindley *et al.*, (2006) *Ann Rev Biochem* 16:16). For other members of the Integrase family, see for example, Esposito *et al.*, (1997) *Nucleic Acids Res* 25:3605-14 and Abremski *et al.*, (1992) *Protein Eng* 5:87-91.

Other recombination systems include, for example, the *streptomycete* bacteriophage phiC31 (Kuhstoss, *et al.*, (1991) *J Mol Biol* 20:897-908); the SSV1 site-specific recombination system from *Sulfolobus shibatae* (Maskhelishvili, *et al.*, (1993) *Mol Gen Genet* 237:334-42); and a retroviral integrase-based integration system (Tanaka, *et al.*, (1998) *Gene* 17:67-76).

Sometimes the recombinase is one that does not require cofactors or a supercoiled substrate, including but not limited to Cre, FLP, and active derivatives, variants or fragments thereof. FLP recombinase catalyzes a site-specific reaction during DNA replication and amplification of the two-micron plasmid of *S. cerevisiae*. FLP recombinase catalyzes site-specific recombination

between two FRT sites. The FLP protein has been cloned and expressed (Cox, (1993) *Proc. Natl. Acad. Sci. USA* 80:4223-7). Functional derivatives, variants, and fragments of FLP are known (Buchholz, *et al.*, (1998) *Nat Biotechnol* 16:617-8, Hartung, *et al.*, (1998) *J Biol Chem* 273:22884-91, Saxena, *et al.*, (1997) *Biochim Biophys Acta* 1340:187-204, and Hartley, *et al.*, (1980) *Nature* 286:860-4).

The bacteriophage recombinase Cre catalyzes site-specific recombination between two lox sites (Guo, *et al.*, (1997) *Nature* 389:40-6; Abremski, *et al.*, (1984) *J Biol Chem* 259:1509-14; Chen, *et al.*, (1996) *Somat Cell Mol Genet* 22:477-88; Shaikh, *et al.*, (1977) *J Biol Chem* 272:5695-702; and, Buchholz, *et al.*, (1998) *Nat Biotechnol* 16:617-8). Examples of site-specific recombinases that can be used to produce a double-strand break at a recognition sequence, including for example FLP, Cre, SSV1, lambda Int, phi C31, HK022, and R. Examples of site-specific recombination systems used in plants can be found in U.S. Patent No. 5,929,301; U.S. Patent No. 6,175,056; WO99/25821; U.S. Patent No. 6,331,661; WO99/25855; WO99/25841, and WO99/25840, the contents of each are herein incorporated by reference.

Methods for modifying the kinetics, cofactor interaction and requirements, expression, optimal conditions, and/or recognition site specificity, and screening for activity of recombinases and variants are known, see for example Miller, *et al.*, (1980) *Cell* 20:721-9; Lange-Gustafson and Nash, (1984) *J Biol Chem* 259:12724-32; Christ, *et al.*, (1998) *J Mol Biol* 288:825-36; Lorbach, *et al.*, (2000) *J Mol Biol* 296:1175-81; Vergunst, *et al.*, (2000) *Science* 290:979-82; Dorgai, *et al.*, (1995) *J Mol Biol* 252:178-88; Dorgai, *et al.*, (1998) *J Mol Biol* 277:1059-70; Yagu, *et al.*, (1995) *J Mol Biol* 252:163-7; Sclimente, *et al.*, (2001) *Nucleic Acids Res* 29:5044-51; Santoro and Schultze, (2002) *Proc. Natl. Acad. Sci. USA* 99:4185-90; Buchholz and Stewart, (2001) *Nat Biotechnol* 19:1047-52; Voziyanov, *et al.*, (2002) *Nucleic Acids Res* 30:1656-63; Voziyanov, *et al.*, (2003) *J Mol Biol* 326:65-76; Klippel, *et al.*, (1988) *EMBO J* 7:3983-9; Arnold, *et al.*, (1999) *EMBO J* 18:1407-14; WO03/08045; WO99/25840; and WO99/25841. The recognition sites range from about 30 nucleotide minimal sites to a few hundred nucleotides.

Any recognition site for a recombinase can be used, including naturally occurring sites, and variants. Variant recognition sites are known, see for

example Hoess, *et al.*, (1986) *Nucleic Acids Res* 14:2287-300; Albert, *et al.*, (1995) *Plant J* 7:649-59; Thomson, *et al.*, (2003) *Genesis* 36:162-7; Huang, *et al.*, (1991) *Nucleic Acids Res* 19:443-8; Siebler and Bode, (1997) *Biochemistry* 36:1740-7; Schlake and Bode, (1994) *Biochemistry* 33:12746-51; Thygarajan, *et al.*, (2001) *Mol Cell Biol* 21:3926-34; Umlauf and Cox, (1988) *EMBO J* 7:1845-52; Lee and Saito, (1998) *Gene* 216:55-65; WO01/23545; WO99/25821; WO99/25851; WO01/11058; WO01/07572 and U.S. Patent No. 5,888,732.

A recombinase can be provided via a polynucleotide that encodes the recombinase or it can be provided via a modified polynucleotide encoding the recombinase. For example, the polynucleotide (encoding a recombinase) can be modified to substitute codons having a higher frequency of usage in a plant, as compared to the naturally occurring polynucleotide sequence, or it can be modified to substitute codons having a higher frequency of usage in a maize or soybean plant, as compared to the naturally occurring polynucleotide sequence.

TAL effector nucleases are a new class of sequence-specific nucleases that can be used to make double-strand breaks at specific target sequences in the genome of a plant or other organism. TAL effector nucleases are created by fusing a native or engineered transcription activator-like (TAL) effector, or functional part thereof, to the catalytic domain of an endonuclease, such as, for example, *FokI*. The unique, modular TAL effector DNA binding domain allows for the design of proteins with potentially any given DNA recognition specificity. Thus, the DNA binding domains of the TAL effector nucleases can be engineered to recognize specific DNA target sites and thus, used to make double-strand breaks at desired target sequences. *See*, WO 2010/079430; Morbitzer *et al.* (2010) *PNAS* 10.1073/pnas.1013133107; Scholze & Boch (2010) *Virulence* 1:428-432; Christian *et al. Genetics* (2010) 186:757-761; Li *et al.* (2010) *Nuc. Acids Res.* (2010) doi:10.1093/nar/gkq704; and Miller *et al.* (2011) *Nature Biotechnology* 29:143–148; all of which are herein incorporated by reference.

Transposases are polypeptides that mediate transposition of a transposon from one location in the genome to another. Transposases typically induce double-strand breaks to excise the transposon, recognize subterminal repeats, and bring together the ends of the excised transposon; in some systems other proteins are also required to bring together the ends during transposition.

Examples of transposons and transposases include, but are not limited to, the Ac/Ds, Dt/rdt, Mu-M1/Mn, and Spm(En)/dSpm elements from maize, the Tam elements from snapdragon, the Mu transposon from bacteriophage, bacterial transposons (Tn) and insertion sequences (IS), Ty elements of yeast (retrotransposon), Ta1 elements from *Arabidopsis* (retrotransposon), the P element transposon from *Drosophila* (Gloor, *et al.*, (1991) *Science* 253:1110-1117), the Copia, Mariner and Minos elements from *Drosophila*, the Hermes elements from the housefly, the PiggyBack elements from *Trichplusia ni*, Tc1 elements from *C. elegans*, and IAP elements from mice (retrotransposon). In some examples the transposase is provided via a polynucleotide that encodes the transposase.

It is possible to modify the polynucleotide encoding the transposase by substituting codons having a higher frequency of usage in a plant, as compared to the naturally occurring polynucleotide sequence, by substituting codons having a higher frequency of usage in a maize or soybean plant, as compared to the naturally occurring polynucleotide sequence.

DNA topoisomerases modulate DNA secondary and higher order structures and functions related primarily to replication, transcription, recombination and repair. Topoisomerases share two characteristics: (i) the ability to cleave and reseal the phosphodiester backbone of DNA in two successive transesterification reactions; and (ii) once a topoisomerase-cleaved DNA intermediate is formed, the enzyme allows the severed DNA ends to come apart, allowing the passage of another single- or double-stranded DNA segment. DNA topoisomerases can be classified into three evolutionarily independent families: type IA, type IB and type II.

Those that cleave one strand of DNA and allow single step changes in the linking number of circular DNA are defined as type I DNA topoisomerases. The *Escherichia coli* topoisomerase I and topoisomerase III, *Saccharomyces cerevisiae* topoisomerase III and reverse gyrase belong to the type IA or type I-5' subfamily as the protein link is to a 5' phosphate in the DNA. The prototype of type IB or I-3' enzymes are found in all eukaryotes and also in vaccinia virus topoisomerase I where the protein is attached to a 3' phosphate. Despite differences in mechanism and specificity between the bacterial and eukaryotic enzymes, yeast DNA topoisomerase I can complement a bacterial DNA

topoisomerase I mutant (Bjornsti, *et al.*, (1987) *Proc. Natl. Acad. Sci. USA* 84:8971-5). Type IA topoisomerases relax negatively supercoiled DNA and require magnesium and a single-stranded region of DNA. Topoisomerases IB relax both positively and negatively supercoiled DNA with equal efficiency and do not require a single-stranded region of DNA or metal ions for function.

The type II family includes *E. coli* DNA gyrase, *E. coli* topoisomerase IV (par E), eukaryotic type II topoisomerases, and archaic topoisomerase VI. Type II enzymes are homodimeric (eukaryotic topoisomerase II) or tetrameric (gyrase), cleaving both strands of a duplex. Preferred cutting sites are known for available topoisomerases.

Zinc finger nucleases (ZFNs) are engineered double-strand-break-inducing agents comprised of a zinc finger DNA binding domain and a double-strand-break-inducing agent domain. Recognition site specificity is conferred by the zinc finger domain, typically comprising two, three, or four zinc fingers, for example having a C2H2 structure, however other zinc finger structures are known and have been engineered. Zinc finger domains are amenable for designing polypeptides which specifically bind a selected polynucleotide recognition sequence. ZFNs consist of an engineered DNA-binding zinc finger domain linked to a non-specific endonuclease domain, for example nuclease domain from a Type IIs endonuclease such as FokI. Additional functionalities can be fused to the zinc-finger binding domain, including transcriptional activator domains, transcription repressor domains, and methylases. In some examples, dimerization of nuclease domain is required for cleavage activity. Each zinc finger recognizes three consecutive base pairs in the target DNA. For example, where a 3 finger domain recognizes a sequence of 9 contiguous nucleotides, with a dimerization requirement of the nuclease, two sets of zinc finger triplets are used to bind an 18 nucleotide recognition sequence. A recognition sequence of 18 nucleotides is long enough to be unique in a mammalian genome ($4^{18} = 6.9 \times 10^{10}$).

To date, designer zinc finger modules predominantly recognize GNN and ANN triplets (Dreier, *et al.*, (2001) *J Biol Chem* 276:29466-78; Dreier, *et al.*, (2000) *J Mol Biol* 303:489-502; Liu, *et al.*, (2002) *J Biol Chem* 277:3850-6), but examples using CNN or TNN triplets are also known (Dreier, *et al.*, (2005) *J Biol Chem* 280:35588-97; Jamieson, *et al.*, (2003) *Nature Rev Drug Discov* 2:361-8).

See also, Durai, *et al.*, (2005) *Nucleic Acids Res* 33:5978-90; Segal, (2002) *Methods* 26:76-83; Porteus and Carroll, (2005) *Nat Biotechnol* 23:967-73; zinc-finger consortium (website at www.zincfinger.org); Pabo, *et al.*, (2001) *Ann Rev Biochem* 70:313-40; Wolfe, *et al.*, (2000) *Ann Rev Biophys Biomol Struct* 29:183-212; Segal and Barbas, (2001) *Curr Opin Biotechnol* 12:632-7; Segal, *et al.*, (2003) *Biochemistry* 42:2137-48; Beerli and Barbas, (2002) *Nat Biotechnol* 20:135-41; Carroll, *et al.*, (2006) *Nature Protocols* 1:1329; Ordiz, *et al.*, (2002) *Proc. Natl. Acad. Sci. USA* 99:13290-5; Guan, *et al.*, (2002) *Proc. Natl. Acad. Sci. USA* 99:13296-301; WO2002099084; WO00/42219; WO02/42459; WO2003062455; U.S. Patent Application Publication No. 20030059767; U.S. Patent Application Publication No. 2003/0108880; U.S. Patent Nos. 6,140,466, 6,511,808 and 6,453,242.

Alternatively, engineered zinc finger DNA binding domains can be fused to other double-strand-break-inducing agents or derivatives thereof that retain DNA nicking/cleaving activity. For example, this type of fusion can be used to direct the double-strand-break-inducing agent to a different target site, to alter the location of the nick or cleavage site, to direct the inducing agent to a shorter target site, or to direct the inducing agent to a longer target site. In some examples a zinc finger DNA binding domain is fused to a site-specific recombinase, transposase, topoisomerase, or a derivative thereof that retains DNA nicking and/or cleaving activity.

It is possible to provide a zinc-finger nuclease via a polynucleotide that encodes the zinc-finger nuclease. This polynucleotide encoding the zinc-finger nuclease can be modified by substituting codons having a higher frequency of usage in a plant, as compared to the naturally occurring polynucleotide sequence or by substituting codons having a higher frequency of usage in a maize or soybean plant, as compared to the naturally occurring polynucleotide sequence.

Sufficient homology or sequence identity indicates that two polynucleotide sequences have sufficient structural similarity to act as substrates for a homologous recombination reaction. The structural similarity includes overall length of each polynucleotide fragment, as well as the sequence similarity of the polynucleotides. Sequence similarity can be described by the percent sequence identity over the whole length of the sequences, and/or by conserved regions

comprising localized similarities such as contiguous nucleotides having 100% sequence identity, and percent sequence identity over a portion of the length of the sequences.

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

Any means can be used to bring together the various components needed to alter the genome of a dicot plant cell. For example, in *in vitro* systems, the double-strand-break-inducing agent and the polynucleotide(s) comprising the recognition site(s) can be provided by contacting the components under the appropriate conditions for DNA cleavage.

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

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

Protocols for introducing polynucleotides and polypeptides into plants may vary depending on the type of plant or plant cell targeted for transformation, such as monocot or dicot. Suitable methods of introducing polynucleotides and polypeptides into plant cells and subsequent insertion into the plant genome include microinjection (Crossway, *et al.*, (1986) *Biotechniques* 4:320-34 and U.S. Patent No. 6,300,543), meristem transformation (U.S. Patent No. 5,736,369), electroporation (Riggs, *et al.*, (1986) *Proc. Natl. Acad. Sci. USA* 83:5602-6, *Agrobacterium*-mediated transformation (U.S. Patent Nos. 5,563,055 and 5,981,840), direct gene transfer (Paszkowski, *et al.*, (1984) *EMBO J* 3:2717-22), and ballistic particle acceleration (U.S. Patent Nos. 4,945,050; 5,879,918; 5,886,244; 5,932,782; Tomes, *et al.*, (1995) "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment" in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods,* ed. Gamborg & Phillips (Springer-Verlag, Berlin); McCabe, *et al.*, (1988) *Biotechnology* 6:923-6; Weissinger, *et al.*, (1988) *Ann Rev Genet* 22:421-77; Sanford, *et al.*, (1987) *Particulate Science and Technology* 5:27-37 (onion); Christou, *et al.*, (1988) *Plant Physiol* 87:671-4 (soybean); Finer and McMullen, (1991) *In Vitro Cell Dev Biol* 27P:175-82 (soybean); Singh, *et al.*, (1998) *Theor Appl Genet* 96:319-24 (soybean); Datta, *et al.*, (1990) *Biotechnology* 8:736-40 (rice); Klein, *et al.*, (1988) *Proc. Natl. Acad. Sci. USA* 85:4305-9 (maize); Klein, *et al.*, (1988) *Biotechnology* 6:559-63 (maize); U.S. Patent Nos. 5,240,855; 5,322,783 and 5,324,646; Klein, *et al.*, (1988) *Plant Physiol* 91:440-4 (maize); Fromm, *et al.*, (1990) *Biotechnology*

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

Alternatively, polynucleotides may be introduced into plants by contacting plants with a virus or viral nucleic acids. Generally, such methods involve incorporating a polynucleotide within a viral DNA or RNA molecule. In some examples a polypeptide of interest may be initially synthesized as part of a viral polyprotein, which is later processed by proteolysis *in vivo* or *in vitro* to produce the desired recombinant protein. Methods for introducing polynucleotides into plants and expressing a protein encoded therein, involving viral DNA or RNA molecules, are known, see, for example, U.S. Patent Nos. 5,889,191, 5,889,190, 5,866,785, 5,589,367 and 5,316,931. Transient transformation methods include, but are not limited to, the introduction of polypeptides, such as a double-strand-break-inducing agent, directly into the organism, the introduction of polynucleotides such as DNA and/or RNA polynucleotides, and the introduction of the RNA transcript, such as an mRNA encoding a double-strand-break-inducing agent, into the organism. Such methods include, for example, microinjection or particle bombardment. See, for example Crossway, *et al.*, (1986) *Mol Gen Genet* 202:179-85; Nomura, *et al.*, (1986) *Plant Sci* 44:53-8; Hepler, *et al.*, (1994) *Proc. Natl. Acad. Sci. USA* 91:2176-80; and, Hush, *et al.*, (1994) *J Cell Sci* 107:775-84.

Standard DNA isolation, purification, molecular cloning, vector construction, and verification/characterization methods are well established, see, for example Sambrook, *et al.*, (1989) *Molecular Cloning: A Laboratory Manual,* (Cold Spring Harbor Laboratory Press, NY). Vectors and constructs include circular plasmids, and linear polynucleotides, comprising a polynucleotide of interest and optionally other components including linkers, adapters, regulatory

regions, introns, restriction sites, enhancers, insulators, selectable markers, nucleotide sequences of interest, promoters, and/or other sites that aid in vector construction or analysis. In some examples a recognition site and/or target site can be contained within an intron, coding sequence, 5' UTRs, 3' UTRs, and/or regulatory regions.

Any promoter can be used, and can be selected based on the desired outcome. A promoter is a region of DNA involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. A plant promoter is a promoter capable of initiating transcription in a plant cell; for a review of plant promoters, see, Potenza, *et al.*, (2004) *In Vitro Cell Dev Biol* 40:1-22. Constitutive promoters include, for example, the core promoter of the Rsyn7 promoter and other constitutive promoters disclosed in WO99/43838 and U.S. Patent No. 6,072,050; the core CaMV 35S promoter (Odell, *et al.*, (1985) *Nature* 313:810-2); rice actin (McElroy, *et al.*, (1990) *Plant Cell* 2:163-71); ubiquitin (Christensen, *et al.*, (1989) *Plant Mol Biol* 12:619-32; Christensen, *et al.*, (1992) *Plant Mol Biol* 18:675-89); pEMU (Last, *et al.*, (1991) *Theor Appl Genet* 81:581-8); MAS (Velten, *et al.*, (1984) *EMBO J* 3:2723-30); ALS promoter (U.S. Patent No. 5,659,026), and the like. Other constitutive promoters are described in, 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. In some examples an inducible promoter may be used. Pathogen-inducible promoters induced following infection by a pathogen include, but are not limited to those regulating expression of PR proteins, SAR proteins, beta-1,3-glucanase, chitinase, *etc.*

Chemical-regulated promoters can be used to modulate the expression of a gene in a plant through the application of an exogenous chemical regulator. The promoter may be a chemical-inducible promoter, where application of the chemical induces gene expression, or a chemical-repressible promoter, where application of the chemical represses gene expression. Chemical-inducible promoters include, but are not limited to, the maize In2-2 promoter, activated by benzene sulfonamide herbicide safeners (De Veylder, *et al.*, (1997) *Plant Cell Physiol* 38:568-77), the maize GST promoter (GST-II-27, WO93/01294), activated by hydrophobic electrophilic compounds used as pre-emergent herbicides, and the tobacco PR-1a promoter (Ono, *et al.*, (2004) *Biosci Biotechnol Biochem* 68:803-7) activated by salicylic acid. Other chemical-

regulated promoters include steroid-responsive promoters (see, for example, the glucocorticoid-inducible promoter (Schena, *et al.*, (1991) *Proc. Natl. Acad. Sci. USA* 88:10421-5; McNellis, *et al.*, (1998) *Plant J* 14:247-257); tetracycline-inducible and tetracycline-repressible promoters (Gatz, *et al.*, (1991) *Mol Gen Genet* 227:229-37; U.S. Patent Nos. 5,814,618 and 5,789,156).

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

Seed-preferred promoters include both seed-preferred promoters active during seed development, as well as seed-germinating promoters active during seed germination. See, Thompson, *et al.*, (1989) *BioEssays* 10:108. Seed-preferred promoters include, but are not limited to, Cim1 (cytokinin-induced message); cZ19B1 (maize 19 kDa zein); and milps (myo-inositol-1-phosphate synthase) (WO00/11177; and U.S. Patent 6,225,529). For dicots, seed-preferred promoters include, but are not limited to, bean β-phaseolin, napin, β-conglycinin, soybean lectin, cruciferin, and the like. For monocots, seed-preferred promoters include, but are not limited to, maize 15 kDa zein, 22 kDa zein, 27 kDa gamma zein, waxy, shrunken 1, shrunken 2, globulin 1, oleosin, and nuc1. See also, WO00/12733, where seed-preferred promoters from *END1* and *END2* genes are disclosed.

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

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

enzyme, chemical, etc.; and, the inclusion of a DNA sequence required for a specific modification (e.g., methylation) that allows its identification.

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

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

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

miliaceum), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), wheat (*Triticum aestivum*), sugarcane (*Saccharum spp.*), oats (*Avena*), barley (*Hordeum*), pineapple (*Ananas comosus*), banana (*Musa spp.*), palm, ornamentals, and grasses. Examples of dicot plants that can be used include, but are not limited to, soybean (*Glycine max*), canola (*Brassica napus* and *B. campestris*), alfalfa (*Medicago sativa*), tobacco (*Nicotiana tabacum*), *Arabidopsis* (*Arabidopsis thaliana*), sunflower (*Helianthus annuus*), cotton (*Gossypium arboreum*), and peanut (*Arachis hypogaea*), tomato (*Solanum lycopersicum*), potato (*Solanum tuberosum*) *etc.*

The transgenes, recombinant DNA molecules, DNA sequences of interest, and polynucleotides of interest can comprise one or more genes of interest. Such genes of interest can encode, for example, a protein that provides agronomic advantage to the plant. Genes of interest can be reflective of the commercial markets and interests of those involved in the development of the crop. Crops and markets of interest change, and as developing nations open up world markets, new crops and technologies will emerge also. In addition, as understanding of agronomic traits and characteristics such as yield and heterosis increases, the choice of genes for transformation will change accordingly. General categories of genes of interest include, for example, those genes involved in information, such as zinc fingers, those involved in communication, such as kinases, and those involved in housekeeping, such as heat shock proteins. More specific categories of transgenes, for example, include genes encoding important traits for agronomics, insect resistance, disease resistance, herbicide resistance, sterility, grain characteristics, and commercial products. Genes of interest include, generally, those involved in oil, starch, carbohydrate, or nutrient metabolism as well as those affecting kernel size, sucrose loading, and the like.

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

protein encoded by the soybean 2S albumin described in U.S. Patent No. 5,850,016, and the chymotrypsin inhibitor from barley, described in Williamson *et al.* (1987) *Eur. J. Biochem.* 165:99-106, the disclosures of which are herein incorporated by reference.

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

Insect resistance genes may encode resistance to pests that have great yield drag such as rootworm, cutworm, European Corn Borer, and the like. Such genes include, for example, *Bacillus thuringiensis* toxic protein genes (U.S. Patent Nos. 5,366,892; 5,747,450; 5,736,514; 5,723,756; 5,593,881; and Geiser *et al.* (1986) *Gene* 48:109); and the like.

Genes encoding disease resistance traits include detoxification genes, such as against fumonosin (U.S. Patent No. 5,792,931); avirulence (avr) and disease resistance (R) genes (Jones *et al.* (1994) *Science* 266:789; Martin *et al.* (1993) *Science* 262:1432; and Mindrinos *et al.* (1994) *Cell* 78:1089); and the like.

Herbicide resistance traits may include genes coding for resistance to herbicides that act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides (e.g., the acetolactate synthase (ALS) gene containing mutations leading to such resistance, in particular the S4 and/or Hra mutations), genes coding for resistance to herbicides that act to inhibit action of glutamine synthase, such as phosphinothricin or basta (e.g., the bar gene);

glyphosate (e.g., the EPSPS gene and the GAT gene; see, for example, U.S. Publication No. 20040082770 and WO 03/092360); or other such genes known in the art. The *bar* gene encodes resistance to the herbicide basta, the *nptII* gene encodes resistance to the antibiotics kanamycin and geneticin, and the ALS-gene mutants encode resistance to the herbicide chlorsulfuron.

Sterility genes can also be encoded in an expression cassette and provide an alternative to physical detasseling. Examples of genes used in such ways include male tissue-preferred genes and genes with male sterility phenotypes such as QM, described in U.S. Patent No. 5,583,210. Other genes include kinases and those encoding compounds toxic to either male or female gametophytic development. Interference with pollen formation, function, or dispersal may be accomplished by disrupting starch accumulation as described in US 7,969,405 and US 7,612,251.

The quality of grain is reflected in traits such as levels and types of oils, saturated and unsaturated, quality and quantity of essential amino acids, and levels of cellulose. In corn, modified hordothionin proteins are described in U.S. Patent Nos. 5,703,049, 5,885,801, 5,885,802, and 5,990,389.

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

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

The transgenes, recombinant DNA molecules, DNA sequences of interest, and polynucleotides of interest can be comprise one or more DNA sequences for gene silencing. Methods for gene silencing involving the expression of DNA sequences in plant are known in the art include, but are not

limited to, cosuppression, antisense suppression, double-stranded RNA (dsRNA) interference, hairpin RNA (hpRNA) interference, intron-containing hairpin RNA (ihpRNA) interference, transcriptional gene silencing, and micro RNA (miRNA) interference

Cosuppression may be used to inhibit the expression of plant genes to produce plants having undetectable protein levels for the proteins encoded by these genes. See, for example, Broin *et al.* (2002) *Plant Cell* 14:1417-1432. Cosuppression may also be used to inhibit the expression of multiple proteins in the same plant. See, for example, U.S. Patent No. 5,942,657. Methods for using cosuppression to inhibit the expression of endogenous genes in plants are described in Flavell *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3490-3496; Jorgensen *et al.* (1996) *Plant Mol. Biol.* 31:957-973; Johansen and Carrington (2001) *Plant Physiol.* 126:930-938; Broin *et al.* (2002) *Plant Cell* 14:1417-1432; Stoutjesdijk *et al* (2002) *Plant Physiol.* 129:1723-1731; Yu *et al.* (2003) *Phytochemistry* 63:753-763; and U.S. Patent Nos. 5,034,323, 5,283,184, and 5,942,657; each of which is herein incorporated by reference. The efficiency of cosuppression may be increased by including a poly-dT region in the expression cassette at a position 3' to the sense sequence and 5' of the polyadenylation signal. See, U.S. Patent Publication No. 20020048814, herein incorporated by reference. Typically, such a nucleotide sequence has substantial sequence identity to the sequence of the transcript of the endogenous gene, optimally greater than about 65% sequence identity, more optimally greater than about 85% sequence identity, most optimally greater than about 95% sequence identity. See, U.S. Patent Nos. 5,283,184 and 5,034,323; herein incorporated by reference.

Antisense suppression may be used to inhibit the expression of multiple proteins in the same plant. See, for example, U.S. Patent No. 5,942,657. Furthermore, portions of the antisense nucleotides may be used to disrupt the expression of the target gene. Generally, sequences of at least 50 nucleotides, 100 nucleotides, 200 nucleotides, 300, 400, 450, 500, 550, or greater may be used. Methods for using antisense suppression to inhibit the expression of endogenous genes in plants are described, for example, in Liu *et al* (2002) *Plant Physiol.* 129:1732-1743 and U.S. Patent Nos. 5,759,829 and 5,942,657, each of which is herein incorporated by reference. Efficiency of antisense suppression

may be increased by including a poly-dT region in the expression cassette at a position 3' to the antisense sequence and 5' of the polyadenylation signal. See, U.S. Patent Publication No. 20020048814, herein incorporated by reference.

Methods for using dsRNA interference to inhibit the expression of endogenous plant genes are described in Waterhouse *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95:13959-13964, Liu *et al.* (2002) *Plant Physiol.* 129:1732-1743, and WO 99/49029, WO 99/53050, WO 99/61631, and WO 00/49035; each of which is herein incorporated by reference.

Methods of hpRNA interference are described in Waterhouse and Helliwell (2003) *Nat. Rev. Genet.* 4:29-38 and the references cited therein. These methods are highly efficient at inhibiting the expression of endogenous genes. See, for example, Chuang and Meyerowitz (2000) *Proc. Natl. Acad. Sci. USA* 97:4985-4990; Stoutjesdijk *et al.* (2002) *Plant Physiol.* 129:1723-1731; and Waterhouse and Helliwell (2003) *Nat. Rev. Genet.* 4:29-38. Methods for using hpRNA interference to inhibit or silence the expression of genes are described, for example, in Chuang and Meyerowitz (2000) *Proc. Natl. Acad. Sci. USA* 97:4985-4990; Stoutjesdijk *et al.* (2002) *Plant Physiol.* 129:1723-1731; Waterhouse and Helliwell (2003) *Nat. Rev. Genet.* 4:29-38; Pandolfini *et al. BMC Biotechnology* 3:7, and U.S. Patent Publication No. 20030175965; each of which is herein incorporated by reference. A transient assay for the efficiency of hpRNA constructs to silence gene expression *in vivo* has been described by Panstruga *et al.* (2003) *Mol. Biol. Rep.* 30:135-140, herein incorporated by reference.

For ihpRNA, the interfering molecules have the same general structure as for hpRNA, but the RNA molecule additionally comprises an intron that is capable of being spliced in the cell in which the ihpRNA is expressed. The use of an intron minimizes the size of the loop in the hairpin RNA molecule following splicing, and this increases the efficiency of interference. See, for example, Smith *et al.* (2000) *Nature* 407:319-320. In fact, Smith *et al.* show 100% suppression of endogenous gene expression using ihpRNA-mediated interference. Methods for using ihpRNA interference to inhibit the expression of endogenous plant genes are described, for example, in Smith *et al.* (2000) *Nature* 407:319-320; Wesley *et al.* (2001) *Plant J.* 27:581-590; Wang and Waterhouse (2001) *Curr. Opin. Plant Biol.* 5:146-150; Waterhouse and Helliwell

(2003) *Nat. Rev. Genet.* 4:29-38; Helliwell and Waterhouse (2003) *Methods* 30:289-295, and U.S. Patent Publication No. 20030180945, each of which is herein incorporated by reference.

Transcriptional gene silencing (TGS) may be accomplished through use of hpRNA constructs wherein the inverted repeat of the hairpin shares sequence identity with the promoter region of a gene to be silenced. Processing of the hpRNA into short RNAs which can interact with the homologous promoter region may trigger degradation or methylation to result in silencing (Aufsatz *et al.* (2002) *PNAS* 99 (Suppl. 4):16499-16506; Mette *et al.* (2000) *EMBO J* 19(19):5194-5201).

The inhibition of the expression of a target protein may be obtained by RNA interference by expression of a gene encoding a micro RNA (miRNA). miRNAs are regulatory agents consisting of about 22 ribonucleotides. miRNA are highly efficient at inhibiting the expression of endogenous genes. See, for example Javier *et al.* (2003) *Nature* 425: 257-263, herein incorporated by reference. For miRNA interference, the expression cassette is designed to express an RNA molecule that is modeled on an endogenous miRNA gene. The miRNA gene encodes an RNA that forms a hairpin structure containing a 22-nucleotide sequence that is complementary to another endogenous gene (target sequence). miRNA molecules are highly efficient at inhibiting the expression of endogenous genes, and the RNA interference they induce is inherited by subsequent generations of plants.

The frequency of homologous recombination is influenced by a number of factors. Different organisms vary with respect to the amount of homologous recombination and the relative proportion of homologous to non-homologous recombination. Generally, the length of the region of homology affects the frequency of homologous recombination events, the longer the region of homology, the greater the frequency. The length of the homology region needed to observe homologous recombination is also species-variable. In many cases, at least 5 kb of homology has been utilized, but homologous recombination has been observed with as little as 25-50 bp of homology. The minimum length of homology needed has been estimated at 20-50 bp in *E. coli* (Singer, *et al.*, (1982) *Cell* 31:25-33; Shen and Huang, (1986) *Genetics* 112:441-57; Watt, *et al.*, (1985) *Proc. Natl. Acad. Sci. USA* 82:4768-72), 63-89 bp in *Sacchromyces.*

cerevisaie (Sugawara and Haber, (1992) *Mol Cell Biol* 12:563-75), and 163-300 bp in mammalian cells (Rubnitz and Subramani, (1984) *Mol Cell Biol* 4:2253-8; Ayares, *et al.*, (1986) *Proc. Natl. Acad. Sci. USA* 83:5199-203; Liskay, *et al.*, (1987) *Genetics* 115:161-7).

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

Homologous recombination has also been accomplished in other organisms. For example, at least 150-200 bp of homology was required for homologous recombination in the parasitic protozoan *Leishmania* (Papadopoulou and Dumas, (1997) *Nucleic Acids Res* 25:4278-86). In the filamentous fungus *Aspergillus nidulans*, gene replacement has been accomplished with as little as 50 bp flanking homology (Chaveroche, *et al.*, (2000) *Nucleic Acids Res* 28:e97). Targeted gene replacement has also been demonstrated in the ciliate *Tetrahymena thermophila* (Gaertig, *et al.*, (1994) *Nucleic Acids Res* 22:5391-8). In mammals, homologous recombination has been most successful in the mouse using pluripotent embryonic stem cell lines (ES) that can be grown in culture, transformed, selected and introduced into a mouse embryo. Embryos bearing inserted transgenic ES cells develop as genetically chimeric offspring. By interbreeding siblings, homozygous mice carrying the selected genes can be obtained. An overview of the process is provided in Watson, *et al.*, (1992) *Recombinant DNA*, 2nd Ed., (Scientific American Books distributed by WH Freeman & Co.); Capecchi, (1989) *Trends Genet* 5:70-6; and Bronson, (1994) *J Biol Chem* 269:27155-8. Homologous recombination in mammals other than mouse has been limited by the lack of

stem cells capable of being transplanted to oocytes or developing embryos. However, McCreath, *et al.*, *Nature* 405:1066-9 (2000) reported successful homologous recombination in sheep by transformation and selection in primary embryo fibroblast cells.

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

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

Alteration of the genome of a plant cell, for example, through homologous recombination (HR), is a powerful tool for genetic engineering. Despite the low frequency of homologous recombination in higher plants, there are a few examples of successful homologous recombination of plant endogenous genes. The parameters for homologous recombination in plants have primarily been investigated by rescuing introduced truncated selectable marker genes. In these experiments, the homologous DNA fragments were typically between 0.3 kb to 2 kb. Observed frequencies for homologous recombination were on the order of

10^{-4} to 10^{-5}. See, for example, Halfter, *et al.*, (1992) *Mol Gen Genet* 231:186-93; Offringa, *et al.*, (1990) *EMBO J* 9:3077-84; Offringa, *et al.*, (1993) *Proc. Natl. Acad. Sci. USA* 90:7346-50; Paszkowski, *et al.*, (1988) *EMBO J* 7:4021-6; Hourda and Paszkowski, (1994) *Mol Gen Genet* 243:106-11; and Risseeuw, *et al.*, (1995) *Plant J* 7:109-19.

An endogenous, non-selectable gene was targeted in *Arabidopsis* using a targeting vector containing a region of about 7 kb homologous to the target gene and the targeting frequency was estimated to be at least 3.9 X 10^{-4} (Maio and Lam, (1995) *Plant J* 7:359-65). In another example, using a positive-negative selection scheme and a targeting vector containing up to 22.9 kb of sequence homologous to the target, homologous recombination was detected with a frequency less than 5.3 X 10^{-5}, despite the large flanking sequences available for recombination (Thykjær, *et al.*, (1997) *Plant Mol Biol* 35:523-30). In *Arabidopsis*, the AGL5 MADS-box gene was knocked out by homologous recombination using a targeting construct consisting of a kanamycin-resistance cassette inserted into the AGL5 sequence roughly 3 kb from the 5' end and 2 kb from the 3' end. Of the 750 kanamycin-resistant transgenic lines that were generated, one line contained the anticipated insertion (Kempin, *et al.*, (1997) *Nature* 389:802-3). Hanin, *et al.*, obtained homologous recombination events at a basal frequency of $7x10^{-4}$ using 3 kb 5'-end and 2kb 3'-end homology to the *Arabidopsis* PPO gene encoding protoporphyrinogen oxidase (Hanin, *et al.*, (2001) *Plant J* 28:671-7). Terada, *et al.*, targeted the Waxy locus in rice using an *Agrobacterium*-mediated transformation procedure. Negative selection, in the form of two copies of the diphteria toxin gene placed at both ends of T-DNA, was used to eliminate random integration of T-DNAs, allowing for enrichment of rare homologous recombination events in the selected material, and their transformation system generated thousands of events from just 150 rice seeds. The reported frequency of homologous recombination of the waxy gene in rice was $0.65x10^{-3}$, without inclusion of elements to enhance homologous recombination (Terada, *et al.*, (2002) *Nat Biotech* 20:1030-4).

DNA double-strand breaks (DSBs) appear to be an effective factor to stimulate homologous recombination pathways in every organism tested to date (Puchta, *et al.*, (1995) *Plant Mol Biol* 28:281-92; Tzfira and White, (2005) *Trends Biotechnol* 23:567-9; Puchta, (2005) *J Exp Bot* 56:1-14). Using DNA-breaking

agents, two- to nine-fold increase of homologous recombination was observed between artificially constructed homologous DNA repeats in plants (Puchta, *et al.*, (1995) *Plant Mol Biol* 28:281-92). In maize protoplasts, experiments with linear DNA molecules demonstrated enhanced homologous recombination between plasmids (Lyznik, *et al.*, (1991) *Mol Gen Genet* 230:209-18).

The effects of DSBs on homologous recombination have been investigated by using rare-cutting as well as transposons such as Ac and Mutator (Chiurazzi, *et al.*, (1996) *Plant Cell* 8:2057-66; Puchta, *et al.*, (1996) *Proc. Natl. Acad. Sci. USA* 93:5055-60; Xiao and Peterson, (2000) *Mol Gen Genet* 263:22-9; and Shalev and Levy (1997) *Genetics* 146:1143-51). Chiurazzi, *et al.*, (1996) *Plant Cell* 8:2057-66) introduced DSBs into an *Arabidopsis* chromosome using HO-endonuclease and observed 10-fold increase in the frequency of homologous recombination between repeats flanking the HO recognition site. Excision of Ac transposable elements also stimulated homologous recombination between repeats flanking the elements at an even higher frequency (Xiao and Peterson (2000) *Mol Gen Genet* 263:22-9).

Puchta *et al.* reported that homologous recombination frequency at an artificial target locus was increased by up to two orders of magnitude when DSBs were generated using I-SceI (Puchta, *et al.*, (1996) *Proc. Natl. Acad. Sci. USA* 93:5055-60). In experiment of Puchta *et al.*, I-SceI expression cassette was introduced into transgenic tobacco target lines together with targeting construct by co-inoculation with the two respective *Agrobacterium* strains. Homologous recombination between T-DNA containing the targeting construct and the target site reconstituted the kanamycin-resistance gene (nptII). There was an apparent correlation between frequency of homologous recombination and the amount of I-SceI expression cassette, suggesting that more DSBs yielded higher homologous recombination frequency.

High frequency of homologous recombination at a pre-introduced artificial target site was obtained using a zinc-finger nuclease (ZFN) in tobacco (Wright, *et al.*, (2005) *Plant J* 44:693-705). The zinc-finger nuclease expression cassette and donor DNA were introduced into protoplasts by co-electroporation and targeted modification was monitored by kanamycin resistance and GUS activity. One modified event was observed in approximately every 10 transformants,

however, only 20% of the modified events contained the desired homologous recombination products as indicated by Southern blot analysis.

Zinc finger nucleases are engineered endonucleases with altered specificities, for example by fusion of an engineered DNA binding domain to an endonuclease, for example, FokI (Durai, *et al.*, (2005) *Nucleic Acids Res* 33:5978-90; Mani, *et al.*, (2005) *Biochem Biophys Res Comm* 335:447-57). Wright, *et al.*, and Lloyd, *et al.*, reported a high frequency mutagenesis at a DNA target site integrated into tobacco or *Arabidopsis* chromosomal DNA using zinc-finger nucleases (Wright, *et al.*, (2005) *Plant J* 44:693-705; Lloyd, *et al.*, (2005) *Proc. Natl. Acad. Sci. USA* 102:2232-7). Using a designed zinc-finger nuclease recognizing a tobacco endogenous acetolactate synthase (ALS) gene locus, a mutated *ALS* gene known to confer resistance to imidazolinone and sulphonylurea herbicides was introduced to replace the endogenous *ALS* gene at frequencies exceeding 2% of transformed cells (Townsend, *et al.*, (2009) *Nature* 459:442-5). The knock-out of an endogenous gene and the expression of a transgene can be achieved simultaneously by gene targeting. The *IPK1* gene, which encodes inositol-1,3,4,5,6-pentakisphosphate 2-kinase needed in the final step of phytate biosythesis in maize seeds, was targeted using a designed zinc-finger nuclease to insert via homologous recombination a *PAT* gene, which encodes phosphinothricin acetyl transferase tolerance to glufosinate ammonium herbicides such as bialaphos. The disruption of the *IPK1* gene with the insertion of the *PAT* gene resulted in both herbicide tolerance and the expected alteration of the inositol phosphate profile in developing seeds (Shukla, *et al.*, (2009) *Nature* 459:437-41).

Members of the serine family of recombinases produce double-strand breaks at the recombination sites as a part of their catalytic activities (Grindley, *et al.*, (2006) *Ann Rev Biochem* 16:16). The R/RS system in sweet orange appeared to induce mutations of RS sites leading to chromosomal deletions not associated with site-specific recombination reactions per se (Ballester, *et al.*, (2006) *Plant Cell Rep* 26:39-45).

Another approach uses protein engineering of existing homing endonucleases to alter their target specificities. Homing endonucleases, such as I-SceI or I-CreI, bind to and cleave relatively long DNA recognition sequences (18 bp and 22 bp, respectively). These sequences are predicted to naturally

occur infrequently in a genome, typically only 1 or 2 sites/genome. The cleavage specificity of a homing endonuclease can be changed by rational design of amino acid substitutions at the DNA binding domain and/or combinatorial assembly and selection of mutated monomers (see, for example, Arnould, *et al.*, (2006) *J Mol Biol* 355:443-58; Ashworth, *et al.*, (2006) *Nature* 441:656-9; Doyon, *et al.*, (2006) *J Am Chem Soc* 128:2477-84; Rosen, *et al.*, (2006) *Nucleic Acids Res* 34:4791-800; and Smith, *et al.*, (2006) *Nucleic Acids Res* 34:e149; Lyznik, *et al.*, (2009) U.S. Patent Application Publication No. 20090133152A1; Smith, *et al.*, (2007) U.S. Patent Application Publication No. 20070117128A1). Engineered meganucleases have been demonstrated that can cleave cognate mutant sites without broadening their specificity. An artificial recognition site specific to the wild type yeast I-SceI homing nuclease was introduced in maize genome and mutations of the recognition sequence were detected in 1% of analyzed F1 plants when a transgenic I-SceI was introduced by crossing and activated by gene excision (Yang, *et al.*, (2009) *Plant Mol Biol* 70:669-79). More practically, the maize liguleless locus was targeted using an engineered single-chain endonuclease designed based on the I-CreI meganuclease sequence. Mutations of the selected liguleless locus recognition sequence were detected in 3% of the T0 transgenic plants when the designed homing nuclease was introduced by *Agrobacterium*-mediated transformation of immature embryos (Gao, *et al.*, (2010) *Plant J* 61:176-87).

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

EXAMPLES

The present invention is further defined in the following Examples, in which parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating

embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions. Such modifications are also intended to fall within the scope of the appended claims.

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

EXAMPLE 1

DNA Double-Strand-Break-Induced Alteration of an Endogenous Target Site

When a DNA double-strand-break-inducing agent recognizes and cleaves the specific recognition sequence at a target site in the genome, a DNA double-strand break is formed, triggering the cell DNA repair mechanisms to mobilize to repair the damage that could be fatal to the cell. The process can be utilized in plant transformation to introduce mutations specifically at the target site to knock out the gene residing at the target site or to insert a donor DNA of interest at the target site. Once the DNA double-strand break is formed, depending on the designs of the DNA constructs involved and the actual processes of DNA repair, different outcomes can be obtained serving different transformation purposes.

For simple site-specific gene mutations, a target site containing a recognition sequence (Figure 1A) and a DNA double-strand break agent such as a endonuclease (Figure 1B) that recognizes specifically the recognition sequence have to be present in the same cell. After the endonuclease recognizes and cuts the DNA, the two free ends can be repaired through end joining by the cell DNA repair machinery without the intervention of any external factors. The two ends can be repaired to the original state so no change can be detected, or they can be altered before being repaired resulting in detectable

changes after they are connected again such as the deletion of one or more nucleotides of the recognition sequence and possibly extra surrounding sequences (Figure 1F). Mutations are introduced at the target site by the latter process.

To achieve site-specific DNA insertions, a donor DNA containing the DNA of interest has to be simultaneously present in the cell in addition to the target site and the endonuclease. The donor DNA can contain the same DNA sequences that flank the target site to flank the DNA of interest, i.e., the homologous sequences (Figure 1C). The DNA of interest can be inserted at the target site by homologous recombination (Figure 1E), a process that is stimulated by the DNA double-strand break at the target site. The donor DNA can also contain only the DNA of interest without any flanking homologous sequences (Figure 1D). The DNA of interest can still be inserted at the target site, though in a less predictable fashion, through non-homologous recombination. Similarly, any unrelated DNA that happens to be present when the DNA ends are repaired can be inserted at the target site (Figure 1G). The different outcomes (Figures 1E-G) can be obtained simultaneously in the same transformation experiment.

Any means to make a DNA double-strand break *in vivo* can be used as the DNA double-strand-break-inducing agent such as the most commonly used meganucleases which recognize >18 bp sequences long enough to be unique in most genomes. Numerous meganucleases have been found and characterized to recognize many different sequences, but such sequences are often not naturally present in important crops such as soybean or maize. Even if similar sequences can be found in crop genomes, the limited numbers of these sequences are still too small to be useful. Certain meganucleases such as I-CreI can be modified by protein engineering in such a way that it will no longer preferentially recognize the recognition sequence of wild type I-CreI and instead will preferentially recognize specifically selected sequences of interest. Taking advantage of the flexibility of the I-CreI endonuclease, one can design and make a modified I-CreI to cleave a target site of choice in the genome and subsequently introduce mutations or insert genes of interest at the selected target site. The precise genetic engineering that this methodology provides will solve many problems that traditional plant transformation methods such as

Agrobacterium infection and biolistic bombardment currently face, such as unpredictable integration, unwanted endogenous gene interruption, unpredicted transgene expression, etc.

EXAMPLE 2

Male-Sterile Maize Plants Produced by the Targeted Mutagenesis of a Cytochrome P450-Like Gene, *MS26*, Using an Engineered MS26 Endonuclease

ZmMS26 is a locus of interest for making a male sterile mutation in maize located on the short arm of chromosome 1. The maize *MS26* gene (SEQ ID NO: 8; AF366297) consists of five exons and it encodes an amino acid sequence (SEQ ID NO 12; AAK52956.1) that shows substantial homology to the *CYP704B1-Zm* gene – a member of the extensive (over 26 genes) family of the maize cytochrome P450 monooxygenases. The heme domain, essential for catalytic activity, is found in the fifth exon (U.S. Patent No. 7517975). Null mutants of the maize *MS26* gene cause premature termination of microspore development in anther locules as this gene has been implicated in pollen wall formation (Li et al. (2010) Plant Cell; 22:173). Frameshift or premature termination mutations in this region are expected to knockout the maize *MS26* gene function.

An engineered I-CreI-based homing endonuclease, referred to as engineered MS26 endonuclease, was able to produce double-strand breaks in the maize *MS26* gene leading to the introduction of mutations that knockout function of the MS26 protein. The process is advantageous because it does not require a dedicated selection step or a modification of routine transformation protocols. As anticipated, single-nucleotide deletions or insertions at the *MS26* coding sequence produced sterile maize plants.

A. TS-MS26 target site and engineered MS26 endonucleases

A target site designated "TS-MS26" target site (SEQ ID NO: 1) was selected for design of a custom double-strand-break-inducing agent. The TS-MS26 target site is a 22 bp polynucleotide positioned 62 bps from the 5' end of the fifth exon of the maize *MS26* gene and having the following sequence:

gatggtgacgtac^gtgccctac (SEQ ID NO: 1).

The double strand break site and overhang region is underlined; the enzyme cuts after C13, as indicated by the ^. Plant optimized nucleotide sequences for three engineered endonucleases (SEQ ID NO:4 encoding engineered MS26 endonuclease; SEQ ID NO: 5 and 7 encoding engineered MS26+endonuclease; SEQ ID NO: 6 encoding engineered MS26++endonuclease) were designed based on the I-CreI homing endonuclease to bind and make double-strand breaks at the selected TS-MS26 target site (SEQ ID NO: 1).

B. Vector construction for plant expression vectors encoding the engineered MS26 endonucleases and repair DNAs for transgene integration by homologous recombination

Vectors comprising expression cassettes for the appropriate engineered endonuclease were constructed using standard molecular biological techniques.

Plant expression cassettes contained the plant codon-optimized nucleotide sequence encoding the engineered MS26 endonuclease for better performance in maize cells. These plant optimized sequences were also supplemented with DNA sequences encoding nuclear localization signals added to the N-terminus of the protein (SEQ ID NO: 2) for the engineered MS26 endonuclease and SEQ ID NO: 3 for the engineered MS26++ endonuclease. The maize ubiquitin promoter and the potato proteinase inhibitor II gene terminator sequences completed the gene designs. In some cases, the plant optimized nucleotide sequence encoding the engineered MS26+ endonuclease (SEQ ID NO:5) was additionally modified by addition of the ST-LS1 intron to the coding sequence of the first endonuclease monomer in order to eliminate its expression in *E.coli* and *Agrobacterium* (SEQ ID NO 7). The expression cassette containing the plant optimized nucleotide sequence encoding the engineered MS26++ endonuclease (SEQ ID NO:6) also contained the ST-LS1 intron inserted into the coding sequences of the first monomer in order to eliminate its expression in *E.coli* and *Agrobacterium* and its codon sequence was optimized for GC content.

These expression cassettes were inserted into T-DNA molecules that were also equipped with a BAR or a moPAT selectable marker gene allowing for selection of transgenic events on media containing bialaphos. No selection was applied for mutations at the TS-MS26 target site.

C. Production of transgenic plants

Maize (*Zea mays*) immature embryos were transformed by a modified *Agrobacterium*-mediated transformation procedure as described in (Djukanovic *et al.* 2006). Ten to eleven day old immature embryos (1.3-1.8 mm) were dissected from sterilized kernels and placed into 2 ml of liquid medium [4.0 g/L N6 Basal Salts (Sigma C-1416; Sigma-Aldrich Co., St.Louis, MO, USA), 1.0 ml/L Eriksson's Vitamin Mix (Sigma E-1511), 1.0 mg/L thiamine HCl, 1.5 mg/L 2,4-dichlorophenoxyacetic acid (2, 4-D), 0.690 g/L L-proline, 68.5 g/L sucrose, 36.0 g/L glucose, pH 5.2]. The *Agrobacterium* suspension was diluted down to O.D. of 0.175 at 550 nm. The embryo-containing medium was replaced with 1 ml of the *Agrobacterium* suspension and the embryos were allowed to incubate for five minutes at room temperature. After incubating, the embryos (40 embryos/plate) were transferred, embryo axis down, onto a plate containing 4.0 g/L N6 Basal Salts (Sigma C-1416), 1.0 ml/L Eriksson's Vitamin Mix (Sigma E-1511), 1.0 mg/L thiamine HCl, 1.5 mg/L 2, 4-D, 0.690 g/L L-proline, 30.0 g/L sucrose, 0.85 mg/L silver nitrate, 0.1 nM acetosyringone, 3.0 g/L Gelrite, pH 5.8. Embryos were incubated in the dark for 3-4 days at 21°C and then transferred to media containing 4.0 g/L N6 Basal Salts (Sigma C-1416), 1.0 ml/L Eriksson's Vitamin Mix (Sigma E-1511), 0.5 mg/L thiamine HCl, 1.5 mg/L 2, 4-D, 0.690 g/L L-proline, 30.0 g/L sucrose, 0.5 g/L 2-(N-morpholino)ethanesulphonic acid (MES) Buffer, 0.85 mg/L silver nitrate, 100 mg/L carbenicillin, and 8 g/L Sigma Agar for an additional four days of incubation in the dark at 28°C. The embryos were then transferred (19 embryos/plate) onto new plates containing 4.0 g/L N6 Basal Salts (Sigma C-1416), 1.0 ml/L Eriksson's Vitamin Mix (Sigma E-1511), 0.5 mg/L thiamine HCl, 1.5 mg/L 2, 4-D, 0.69 g/L L-proline, 30.0 g/L sucrose, 0.5 g/L MES buffer, 0.85 mg/L silver nitrate, 1.5 mg/L Bialaphos, 100 mg/L carbenicillin, 8.0 g/L agar, pH 5.8, and placed in the dark at 28°C. After three weeks, the responding callus (7 calli/plate) was sub-cultured onto media containing 4.0 g/L N6 Basal Salts (Sigma C-1416), 1.0 ml/L Eriksson's Vitamin Mix (Sigma E-1511),

0.5 mg/L thiamine HCl, 1.5 mg/L 2, 4-D, 0.69 g/L L-proline, 30.0 g/L sucrose, 0.5 g/L MES buffer, 0.85 mg/L silver nitrate, 3.0 mg/L Bialaphos, 100 mg/L carbenicillin, 8.0 g/L agar, pH 5.8. After five weeks in the dark at 28°C somatic embryogenesis was induced by transferring a small amount of selected tissue onto a regeneration medium (1 transgenic event/plate) containing 4.3 g/L Murashige and Skoog (MS) salts (Gibco 11117; Gibco, Grand Island, NY), 5.0 ml/L MS Vitamins Stock Solution (Sigma M3900), 100 mg/L myo-inositol, 0.1 □M abscisic acid (ABA), 1 mg/L indoleacetic acid (IAA), 0.5 mg/L zeatin, 60.0 g/L sucrose, 3.0 mg/L Bialaphos, 100 mg/L carbenicillin, 6.0 g/L Ultrapure Agar, pH 5.6. The plates were incubated in the dark for 2-3 weeks at 28°C. All material with visible shoots and roots was transferred (5-10 shoots per event/plate) onto media containing 4.3 g/L MS salts (Gibco 11117), 5.0 ml/L MS Vitamins Stock Solution (Sigma M3900), 100 mg/L myo-inositol, 40.0 g/L sucrose, 3 mg/L Bialaphos, 100 mg/L Benomyl 6.0 g/L Bacto-Agar, pH 5.6, and incubated under artificial light at 28°C. One week later, plantlets were moved into glass tubes (1 plantlet/tube) containing the same medium minus the Bialaphos and grown under artificial light until they were sampled and/or transplanted into soil.

Four versions of plant optimized sequences encoding the engineered endonuclease targeting the TS-MS26 target site (SEQ. ID NO:4, SEQ. ID NO:5, SEQ ID NO:6 or SEQ. ID NO:7) were delivered by *Agrobacterium*-mediated transformation of immature maize embryos. Over 1,000 T0 plants were produced, each plant regenerated from an independent callus tissue selected on media containing 1.5 mg/L Bialaphos. Transformation efficiency represented the percentage of co-cultivated embryos that produced transformation events. The transformation frequency ranged from 14% to 35% for the engineered MS26 endonuclease; this was within the routine transformation frequencies recorded for other genes used for genetic transformation of the maize embryos with similar *Agrobacterium* strains.

D. Screening for mutations at the TS-MS26 target site and selecting mutant plants.

T0 plants were screened for mutations at the TS-MS26 target site by PCR amplification of the TS-MS26 target site region and subsequent digestion of the PCR product with the BsiWI restriction enzyme, which cuts within TS-MS26

target site , followed by gel electrophoresis of the restriction digestion products. Failure to cut the PCR amplified target site region to completion with BsiWI indicated that a mutation had occurred. Three mutated *MS26* alleles were identified among 300 T0 plants from the transformation experiment using SEQ ID NO:5 (Table 2) and three mutated *MS26* alleles were identified among 257 analyzed T0 plants from the experiment using SEQ ID NO: 7. No mutant *MS26* alleles were found in the transformation experiment using SEQ ID NO: 4. The highest frequency of mutant alleles was observed when SEQ ID NO: 6 was used in the transformation experiment, yielding 15 mutations among 344 analyzed T0 plants (Table 2).

The results are presented in Table 2. Mutation rate represents the percentage of T0 analyzed plants containing a mutation at the TS-MS26 target site. The mutated *ms26* alleles found in the T0 plants were centered on the apparent 3' end GTAC overhang produced by the engineered endonucleases (Figure 2).

TABLE 2.

Mutation rate at the TS-MS26 target site of T0 maize plants expressing engineered endonucleases

Endonuclease (Plant optimized sequence)	# analyzed T0 plants	# mutations	Mutation rate*
MS26 (SEQ ID NO: 4)	229	0	0
MS26+ (SEQ ID NO: 5)	300	3	1%
MS26+ (SEQ ID NO: 7)	257	3	1.2%
MS26++ (SEQ ID NO: 6)	344	15	5.8%

T0 plants at time of flowering are shown in Figure 8. There was no obvious difference in the growth and development of T0 plants containing one mutated *ms26* allele (two outside plants) as compared to the T0 biallelic event (the tagged plant) produced by the engineered MS26++ endonuclease (A). The biallelic event was sterile (the tassel at anthesis shown between two tassels from monoallelic events) (B).

Selected T0 plants containing one nucleotide insertion *(ms26-Ci)* or one nucleotide deletion *(ms26-Td)* in the endogenous *MS26* gene were grown in a greenhouse and self-pollinated to produce T1 progeny plants. The *MS26* alleles present in the T1 plants were determined by PCR assays on leaf DNA samples of T1 plants. Segregation of the *ms26-Ci* (one nucleotide insertion) and *ms26-Td* (one nucleotide deletion) alleles in the T1 progeny is shown in Table 3.

TABLE 3.

Segregation data of progeny from T0 plants that were self pollinated.

T0 allele (mutation)	# T1 seeds planted	# T1seeds germinated	*MS26/MS26*	Genotype *MS26/ms26*	*ms26/ms26*
ms26-Ci	100	69	25	27	17
ms26-Td	100	93	26	50	17

E. Sterility induced by the mutations at the TS-MS26 target site.

T1 sibling maize plants derived from selfing two original T0 plants each carrying a mutation in the *MS26* gene (*ms26-Ci* and *ms26-Td*) were cultivated under standard growing conditions. Figure 9A shows one heterozygous *MS26/ms26-* and one homozygous *ms26-/ms26-* plant derived from each T0 event at flowering time. All plants were similar in stature and reached maturity at similar times. The silks emerged from the husks at the tips of the ears over a period of few days in all. The close-up photos of mature tassel inflorescences at anthesis are shown in Figure 9B and 9C. The plants produced spikelet-containing tassels with central rachis and branches. Only heterozygous *MS26/ms26-Ci* and *MS26/ms26-Td* T1 sibling plants developed anthers that were hanging on elongated filaments. The tassels of the *ms26-Ci/ms26-Ci* and *ms26-Td/ ms26-Td* homozygous sibling plants contained spikelets with no evidence of emerging anthers, indicating male sterility.

EXAMPLE 3

Male-Sterile Rice Plants Produced by the Targeted Mutagenesis of a Cytochrome P450-Like Gene, *MS26*, Using an Engineered MS26 Endonuclease

The maize *MS26* gene (ZmMS26, Accession No. AF366297; SEQ ID NO: 8) and its orthologues in rice (Accession No. LOC_Os03g07250; SEQ ID NO: 9) sorghum (SEQ ID NO: 10) and rye (SEQ ID NO: 11) encode a cytochrome P450, CYP704B family (U.S. Pat. App. Pub. No. 2009/0183284), proposed to catalyze the production of omega-hydroxylated fatty acids with 16 and 18 carbon chains (Li *et al.* (2010) *Plant Cell* 22:173). Homozygous recessive mutations that disrupt the coding frame of the maize (U.S. Patent No. 7517975) or rice *MS26* genes result in the plant's inability to generate functional pollen grains which is likely due to reduced production of fatty acids critical to pollen wall formation (Li *et al.* (2010) *Plant Cell* 22:173). The *MS26* genes from maize, rice, sorghum and rye all contain an identical 22 nucleotide sequence, referred to as the "TS-MS26" target site, within the last exon of the genes (Figure 3).

A. TS-MS26 target site and Engineered MS26 endonucleases

As described in Example 2, the TS-MS26 target site was selected for design of engineered MS26 endonucleases. Both Indica and japonica rice varieties contain this endogenous TS-MS26 target site in their genome. The genomic region comprising the TS-MS26 target site in rice is shown in Figure 3 (SEQ ID NO: 14). Figure 3 also shows the genomic region comprising the TS-MS26 target site in maize (SEQ ID NO: 13), rice (SEQ ID NO: 14), sorghum (SEQ ID NO: 15) and rye (SEQ ID NO: 16) and illustrates that these genomic regions can contain some base pair differences between species and still be a functional target site for an engineered MS26 endonuclease.

B. Vectors and transformation

Young rice callus (*Oryza sativa* ssp. *japonica* cv. Nipponbare or Kitaake) containing an endogenous TS-MS26 target site was used as transformation targets for *Agrobacterium*- or biolistic-mediated DNA delivery. For biolistic

transformation, the vectors PHP40082 and PHP40126 (Figure 4; SEQ ID NOs: 58 and 59, respectively) were co-bombarded into two-week-old, seed-derived callus by modifying a protocol described by Chen *et al.* ((1998) *Plant Cell Rep.* 18:25-31). PHP40082 contains a plant optimized sequence (SEQ ID NO:5) encoding a single-chain engineered MS26+ endonuclease placed under the transcriptional control of the maize Ubiquitin promoter. PHP40126 contains the herbicide resistance selectable marker fused to the Red-fluorescence gene (RFP) and placed under the regulation of the maize END2 promoter.

PHP40827 was used to generate rice events by *Agrobacterium*-mediated transformation. PHP40827 contains a plant optimized nucleotide sequence (SEQ. ID NO:5), encoding a single-chain engineered MS26+ endonuclease , placed under the transcriptional control of a CAMV35S promoter containing 3 copies of the Tet operator. This plasmid also contains the tetracycline repressor under the control of the maize Ubiquitin promoter, and a blue-fluorescence gene (CFP) regulated by the ZmEND2 promoter. In addition, PHP40827 contains a copy of a red fluorescence gene regulated by the maize Histone 2B promoter. A portion of the red fluorescence gene in this construct was duplicated in a direct orientation, consisting of two fragments of the RFP gene with 369 bp of overlap. The two fragments are separated by a 136-bp spacer which contains the TS-MS26 target site (Figure 6A). In the absence of tetracycline, callus fluoresce blue due to the expression of the CFP-marker. In the presence of tetracycline, derepression of the engineered MS26+ meganuclease would lead to double-strand breaks at the TS-MS26 target site between the two overlapping sequences to promote intramolecular recombination and produce a functional RFP gene, which is revealed by the appearance of red fluorescing cells against a background of blue fluorescence. Red fluorescing callus events were selected for additional characterization and plant regeneration.

C. Identification of mutations at the TS-MS26 target site in plant tissues

Bialaphos-resistant red fluorescing callus or blue and red fluorescing callus events generated by biolistic or *Agrobacterium*-mediated transformation, respectively were screened for TS-MS26 target site mutations by amplification of the region by PCR using the primer pair UNIMS26 5'-2

(GACGTGGTGCTCAACTTCGTGAT) (SEQ ID NO: 17) and UNIMS26 3'-1 (GCCATGGAGAGGATGGTCATCAT) (SEQ ID NO: 18) and digestion of the amplified products with the DNA restriction enzyme, *BsiWI*, which recognizes the sequence 5'-CGTACG-3'. Products of these reactions were electrophoresed on 1% agarose gels and screened for *BsiWI* digestion resistant bands indicative of mutations at the TS-MS26 target site.

Twenty two of the 292 bialaphos-resistant events generated by co-bombardment PHP40082 and PHP40126 events contained PCR products resistant to *BsiWI* restriction enzyme digestion indicating mutations at TS-MS26 target site. Subcloning and DNA sequence analysis of these PCR products revealed a variety of mutations across the TS-MS26 target site, including point mutations, as well as deletions and insertions ranging from one to greater than 250 nucleotides. Examples of these mutations are shown in Figure 5. In several cases the insertion at the TS-MS26 target site consisted of fragments of sequences derived from the co-bombarded vector (for example, see, Figure 5; Event 48(Ev.48) contains 54 base pairs of RFP).

Blue fluorescing rice callus events containing PHP40827 were also screened for the presence of mutations at the TS-MS26 target site after treatment with tetracycline (Figure 6B). Eight independent PHP40827 events were placed onto callus maintenance media containing 1mg/liter tetracycline (TET) for 24 hours at 37°C; genomic DNA was isolated and analyzed by PCR for mutations at the TS-MS26 target site and compared to the PCR products of these same events not exposed to tetracycline (control) (Figure 6B). Six of the eight PHP40827 events yielded *BsiWI* resistant PCR products that were dependent upon tetracycline application. PCR products from TET and control treatments were subcloned and subjected to DNA sequence analysis. The majority of the PCR products from the uncut control treatment reactions did not reveal mutations across the TS-MS26 target site. In contrast, the majority of the DNA sequences from BSIWI resistant PCR products revealed a high proportion of deletions and insertions across the TS-MS26 target site (see, examples in Figure 7). Plants were regenerated from callus events containing mutations at the TS-MS26 target site for phenotypic analysis.

D. Phenotypic analysis of Rice plants containing *MS26* mutations

Herbicide resistant plants regenerated from callus events co-bombarded with PHP40082 and PHP40126 as well as plants from blue fluorescing callus containing PHP40827 were grown under greenhouse conditions, analyzed for mutations in *MS26* and allowed to set selfed seed (T1 seed). Male fertility was screened by selecting T1 seed from 6 plants (3 PHP40082/PHP40126 and 3 PHP40827) containing non-identical *MS26* mutations but lacking the vectors used for transformation. Male fertility was determined by examining anthers for the development of starch filling pollen grains coupled with the plant's ability to set self seed. Plantlets were screened by PCR for mutations at the TS-MS26 target site target site; *MS26/ms26* heterozygous and *ms26/ms26* homozygous mutant plants were advanced and scored for their ability to generate functional pollen (Figure 10A). In summary, 34 of 34 of the *MS26/ms26* plants were male fertile, while, with two exceptions, 27 of the 29 *ms26/ms26* plants were male sterile (Table 4). Microscopic examination of anthers derived from *ms26/ms26* plants staged at late uninucleate microspore development revealed a reduced number and abnormally shaped microspores (Figure 10B). In contrast, anthers from *MS26/ms26* plants contained many normal microspores (Figure 10C) similar to observations reported by Li et al. (Plant Cell 2010; 22:173). The male sterile plants were female fertile as demonstrated by their ability to set seed when fertilized with wild-type rice pollen.

TABLE 4.

Fertility scores of rice plants from selfed seed

SOURCE	MUTATION	Num. plants	MS26/ms26		ms26/ms26	
PHP40082/PHP40126			FERTILE	STERILE	FERTILE	STERILE
EVENT 1	66bp INSERTION	8	4		1*	3
EVENT 2	36 bp deletion	23	17			6
EVENT 3	56bp deletion	8	4			4
PHP40827						
EVENT 4	51bp deletion	8	3			5
EVENT 5	LARGE DELETION	8	3		1*	4
EVENT 6	3 bp deletion	8	3			5

*incorrect genotype

EXAMPLE 4

Targeted Mutations in the Maize *MS45* Gene

Maize lines comprising an endogenous target recognition sequence in their genome were contacted with an engineered meganuclease designed to specifically recognize and create a double-strand break in the endogenous target sequence in the *MS45* gene. Immature embryos comprising an endogenous target site were contacted with the components described below, events selected and characterized.

A. Maize TS-MS45 target site and engineered MS45 endonuclease

An endogenous maize genomic target site located in the *MS45* gene and referred to as the TS-MS45 target site (SEQ ID NO: 20), was selected for design of an engineered double-strand-break-inducing agent.

The genomic region comprising the TS-MS45 target site has the following sequence, with the TS-MS45 target site shown underlined:

GGAGTTCTGCGGCCGGCCGCTCGGCCTGAGGTTCCACGGGGAGACCGGC GAGCTCTACGTCGCCGACGCGTACTACGGTCTCATGGTCGT (SEQ ID NO: 19)

The TS-MS45 target site is a 22 bp polynucleotide having the following sequence:

CGGGGAGAC**CGGC**^GAGCTCTAC (SEQ ID NO: 20)

The double strand beak site and overhang region is shown in bold, the enzyme cuts after nucleotide 13, as indicated by the ^.

An engineered MS45 endonuclease designed to recognize the TS-MS45 target site was produced under contract with Precision BioSciences, Inc. (Durham, NC USA). The engineered MS45 endonuclease is a heterodimer. One monomer is designated MAY1 and the other is designated MAY2.

A nucleus localization signal (SEQ ID NO: 21) was added to the amino terminus of each monomer to improve transport of the protein into the nucleus.

Plant optimized nucleotide sequences encoding MAY1 (SEQ ID NO: 22) or MAY2 (SEQ ID NO: 23) were constructed.

B. Vector construction for plant expression vectors encoding the engineered MS45 endonuclease and repair DNAs for transgene integration by homologous recombination

The strategies employed for generating and selecting genomic alterations produced do not employ reconstitution of a selectable marker expression cassette, therefore the double-strand-break-inducing agent vectors do not have a fragment of a selectable marker cassette. In this example, the double-strand-break-inducing agent vectors had a phenotypic marker expression cassette encoding phosphinothricin acetyltransferase, which was used to validate successful delivery of the vector.

Vectors containing the plant optimized coding sequences encoding an MS45 endonuclease were constructed using standard molecular biology techniques. PHP31456 comprises the following operably linked components: Ubi pro::ubi 5' UTR::cMAY1::pinII::Ubi pro::ubi 5' UTR::cMAY2::pinII::35S CaMV pro::BAR::pinII; wherein ubi pro is the maize ubiquitin promoter, ubi 5' UTR is the 5' untranslated region of the maize ubiquitin gene, cMAY1 and cMAY2 are the

DNA sequences encoding the MAY1 and MAY2 monomers, respectively, designed to specifically recognize and induce a double-strand break at the endogenous TS-MS45 maize genome target site, 35S CaMV pro is the 35S Cauliflower Mosaic Virus promoter, BAR encodes phosphinothricin acetyltransferase, and pinII is the transcription termination sequence from potato proteinase inhibitor II. PHP31458 comprises the following operably linked components: Ubi pro::ubi 5' UTR::NLS::cMAY1::pinII::Ubi pro::ubi 5' UTR::NLS::cMAY2::pinII::35S CaMV pro::BAR::pinII; wherein ubi pro is the maize ubiquitin promoter, ubi 5' UTR is the 5' untranslated region of the maize ubiquitin gene, NLS is a DNA fragment encoding an SV40 nuclear localization signal, cMAY1 and cMAY2 are the DNA sequences encoding the MAY1 and MAY2 monomers, respectively, designed to specifically recognize and induce a double strand break at the endogenous TS-MS45 maize genome target site, 35S CaMV pro is the 35S Cauliflower Mosaic Virus promoter, BAR encodes phosphinothricin acetyltransferase, and pinII is the transcription termination sequence from potato proteinase inhibitor II.

These vectors were designed to induce double-strand breaks at the TS-MS45 target site and thereby produce alterations of the TS-MS45 target site. The vector components of PHP31456 and PHP31458 were inserted between the right border and left border of T-DNA for Agrobacterium mediated introduction into plant cells creating vectors PHP31457 AND PHP31459.

Maize immature embryos 9-12 days after pollination (DAP) were transformed with vector PHP31457 (SEQ ID NO: 30) or PHP31459 (SEQ ID NO: 31) using *Agrobacterium*-mediated methods as described in Example 2C.

Bialaphos resistance was used to identify putative transformation events by callus selection on media containing 3 mg/L bialaphos. Callus tissue and/or plants regenerated from stable transformants using standard culture and regeneration conditions were screened for modification(s) of the endogenous target site.

C. Evaluation of transformed maize for modification of TS-MS45 target site

Any standard protocol for isolation, manipulation, and characterization of polynucleotides and or proteins can be used to identify, select, and characterize putative modification events.

PCR products were produced from genomic DNA obtained from transformed maize cells using primers flanking the target site and purified by Qiaquick (Qiagen Inc., Valencia, NM, USA). The double-strand-break-inducing enzyme or a restriction enzyme contained in the target site was added to the purified target site PCR product DNA to test if the target site had been modified. This mixture was digested at 37°C for about 0.5 hr to overnight (approximately 17 hr), the digestion time depending on the enzyme used. Samples with meganuclease enzyme were treated with 0.5 µL proteinase K and 0.2 µL 20%SDS to denature the protein. The digestion products were separated on a 1.5 to 2% agarose gel. Undigested products indicate that the target site was modified.

Bialaphos-resistant callus and/or T0 plant events were screened for mutations at the TS-MS45 target site by PCR amplification of the target site region using a primer pair which produced a 389 bp product. Samples that yielded the 389 bp PCR product were subjected to enzyme digestion with an engineered MS45 endonuclease. In some cases the PCR product was directly cloned and sequenced. No transformation events with modifications of TS-MS45 target site were identified in approximately 300 transformed plants analyzed.

D. Improvements of the engineered MS45 endonuclease

Further evaluation of the engineered MS45 endonuclease indicated that the activity of this nuclease was lower than other nucleases, e.g. MS26+ and MS26++, that were able to produce modifications of endogenous maize target sites.

An improvement in the design of the engineered MS45 endonuclease expected to increase nuclease activity in maize is a single chain protein comprising the MAY1 and MAY2 monomers fused using a linker polypeptide. The MAY1 and MAY2 monomers can be linked to create a single chain protein

of the form MAY1-linker-MAY2 or MAY2-linker-MAY1. A plant optimized gene encoding a MAY1-linker-MAY2 protein is shown in SEQ ID NO: 34. If desired, a nuclear localization signal (e.g., SEQ ID NO: 21) can be added to the amino terminus of this protein.

E. Alternate method for delivery of engineered MS45 endonuclease genes into maize cells

Introduction of an engineered MS45 endonuclease gene via a direct DNA delivery method, e.g. particle bombardment, increases the copy number of the meganuclease gene, as compared to introduction of the meganuclease gene via *Agrobacterium*; the increased meganuclease gene copy number increases frequency of the target site modifications 10-50 fold. Immature maize embryos from greenhouse or field grown High type II (HiII) donor plants are bombarded with at least one polynucleotide construct described above. If the construct does not include a selectable marker, another polynucleotide containing a selectable marker gene can be co-precipitated on the particles used for bombardment.

Ears are harvested 8-12 days after pollination for the isolation of fertilized embryos. The harvested ears are surface sterilized in 50% Clorox® bleach plus 0.5% Micro detergent for 20 minutes, then rinsed twice with sterile water. The immature embryos are isolated and placed embryo axis side down (scutellum side up), 25 embryos per plate. These are cultured on 560L agar medium 4 days in the dark prior to bombardment. Medium 560L is an N6-based medium containing Eriksson's vitamins, thiamine, sucrose, 2,4-D, and silver nitrate. The day of bombardment, the embryos are transferred to 560Y medium for 4 hours and are arranged within the 2.5-cm target zone. Medium 560Y is a high osmoticum medium (560L with high sucrose concentration).

Particles are prepared by precipitating the DNA to be delivered onto 1.0 μm (average diameter) gold pellets using a $CaCl_2$ precipitation procedure as follows: 100 μL prepared gold particles (0.6 mg) in water, 20 μL (2 μg) DNA in TrisEDTA buffer (1 μg total), 100 μL 2.5 M CaC12, 40 μl 0.1 M spermidine. Each reagent is added sequentially to the gold particle suspension. The final mixture is sonicated briefly. After the precipitation period, the particles are centrifuged briefly, washed with 500 μL 100% ethanol, pelleted again and resuspended in 60 μL 100% ethanol to make the final suspension. Macrocarriers are prepared by

briefly sonicating the final preparation, spotting 5 µL onto the center of each macrocarrier, and drying for about 2 minutes before bombardment. The sample plates are bombarded at a distance of 8 cm from the stopping screen to the tissue, using a DuPont biolistics helium particle gun. All samples receive a single shot at 650 PSI, with a total of ten aliquots taken from each tube of prepared particles/DNA.

Alternatively, DNA to be delivered is associated with microparticles using a reagent comprising a cationic lipid solution. For example, DNA solutions are added to 50 µL of a gold-particle stock solution (0.1 µg/µL of 0.6 micron gold particles). A DNA stock, 10 µL of a 0.1 µg/µL plasmid solution, is added to 30 µL of water. To this DNA mixture, 50 µL of the gold stock solution is added and the mixture briefly sonicated. Next 5 µL of TFX-50™ (Promega Corp, Madison WI) is added, and the mixture is placed on a rotary shaker at 100 rpm for 10 minutes. The mixture is briefly centrifuged to pellet the gold particles and remove supernatant. After removal of the excess DNA/TFX solution, 120 µL of absolute EtOH is added, and 10 µL aliquots are dispensed onto the macrocarriers typically used with the DuPont PDS-1000 Helium Particle Gun. The gold particles with adhered DNA are allowed to dry onto the carriers and then these are used for standard particle bombardment.

Four to 12 hours post bombardment, the embryos are moved to a low osmoticum callus initiation medium for 3-7 days at 28°C, then transferred to selection medium and subcultured every 2 weeks. Incubation of the embryos post bombardment for about 48hrs at 32°C increases the frequency of target site modifications 2-4 fold for most meganucleases. After about 10 weeks, embryos are transferred to regeneration media. Following 2-4 weeks of somatic embryo maturation, well-developed somatic embryos are transferred to germination medium in a lighted culture room. Approximately 7-10 days later, developing plantlets are transferred to tubes until plantlets are well established and can be transplanted into flats and/or pots and grown to maturity.

EXAMPLE 5

Male-Sterile Sorghum Plants Produced by the Targeted Mutagenesis of a Cytochrome P450-Like Gene, *MS26*, Using an Engineered MS26 Endonuclease

A. TS-MS26 target site and Engineered MS26 endonucleases

As described in Example 2, the TS-MS26 target site was selected for design of engineered MS26 endonucleases. The genomic region comprising the TS-MS26 target site in sorghum *(Sorghum bicolor)* is shown in Figure 3 (SEQ ID NO: 15).

B. Vectors and transformation

Immature sorghum (*Sorghum bicolor*) embryos containing an endogenous MS26 target site (TS-MS26, SEQ ID NO: 1) were used as transformation targets for Agrobacterium DNA delivery. PHP42063 (SEQ ID NO: 61) was used to generate sorghum callus events by Agrobacterium-meditated transformation. PHP42063 contains a single chain MS26+ endonuclease (described in Example 2) placed under the transcriptional control of the maize CAS1 promoter. The CAS1 promoter has been shown to be transcriptionally induced by either the sulfonylurea-safener, 2-CBSU, or by elevated temperature (U.S. Patent application 61/648,758, filed May 18, 2012). PHP42063 also contains a blue-fluorescence gene (CFP) regulated by the ZmEND2 promoter which is used as visual marker for the selection of integration of the T-DNA into sorghum cells. In addition, PHP42063 also contains a copy of a red fluorescence gene regulated by the maize Histone 2B promoter. A portion of the red fluorescence gene in this construct was duplicated in a direct orientation, consisting of two fragments of the RFP gene with 369 bp of overlap. The two fragments are separated by a 136-bp spacer which contains an MS26 target site as described for PHP40827 (Fig. 6, SEQ ID NO:60). Immature embryos were transformed with PHP42063 according to Zhao et al (*Plant Molecular Biology* 44: 789–798, 2000). Blue fluorescing calli were selected and used for regeneration of plants and grown in the greenhouse to maturity and seed set. Sorghum plants containing DNA insertions of PHP42063 were verified by copy-number analysis. Four independent single or low-copy PHP42063 transformed plants were selected for

additional experimentation. Blue fluorescing immature embryos were harvested 14-20 days after pollination, sterilized, placed on maintenance media (PHI-U without PPT selection) and incubated in the dark at either room temperature (23C-26C) or at the elevated temperature of 37C for 24 to 48 hours. At the end of this period, embryos incubated at the elevated temperature were moved to room temperature (<26C) and embryos were allowed to grow in the dark. As described above, embryos containing PHP42063 and maintained at 26C post harvest only fluoresce blue due to the expression of the CFP-marker. In contrast, approximately 72 hours after treatment at elevated temperature, embryos incubated at 37C begin to develop red fluorescing sectors on the embryo. This observation suggests that the heat inducible gene cassette, CAS1:MS26+, has resulted in double-strand breaks at the MS26 target site between the two overlapping sequences of the RF-FP reporter promoting intramolecular recombination and producing a functional RPF gene which is revealed by the appearance of red fluorescing cells against a background of blue fluoresce. Red fluorescing callus events were selected for plant regeneration and additional molecular and phenotype characterization.

C. Identification of mutations at the TS-MS26 target site in Sorghum tissues

Regenerated plants were screened for mutations at the TS-MS26 target site by amplification of the region by PCR using the primer pair UNIMS26 5'-2 (GACGTGGTGCTCAACTTCGTGAT, SEQ ID NO: 17) and UNIMS26 3'-1 (GCCATGGAGAGGATGGTCATCAT, SEQ ID NO: 18) and digestion of the amplified products with the DNA restriction enzyme, BsiWI, which recognizes the sequence 5'-CGTACG-3'. Products of these reactions were electrophoresed on 1% agarose gels and screened for BsiWI digestion resistant bands indicative of mutations at the TS-MS26 target site.

One hundred twenty nine out of the 389 regenerated plants from PHP42063 heat treated embryos generated contained PCR products resistant to *BsiWI* restriction enzyme digestion indicating mutations at the TS-MS26 target site. Subcloning and DNA sequence analysis of these PCR products revealed a variety of mutations across the TS-MS26 region which consisted of primarily deletions within and across the TS-MS26 target site ranging from 3 to 98

nucleotides. Occasionally, small insertions of single to 11 nucleotides were detected. In total, 16 non-identical mutations were identified in these regenerated sorghum plants (Figure 12, SEQ ID NOs: 62-78). Plants containing mutations at the TS-MS26 target site were used for phenotypic analysis.

D. Phenotypic analysis of Sorghum plants containing *MS26* mutations

Regenerated plants containing mutations at TS-MS26 in Sorghum were grown under greenhouse conditions and allowed to set selfed seed (T1 seed). Male fertility phenotype was screened by planting T1 seed from plants containing the 78 bp deletion (ms26.78Δ across the TS-MS26 target site and allowing these plants to flower. Prior to flowering, seedlings were screened by PCR for mutations at the TS-MS26 target site; MS26/MS26 (wild-type), MS26/ms26.78Δ (heterozygous) and ms26.78Δ/ms26.78Δ (recessive) plants were identified and advanced. Male fertility was determined by examining anthers for the development of starch filling pollen grains coupled with the plant's ability to set self seed. As shown in Figure 13A, panicles of MS26/ ms26.78Δ revealed anther extrusion, pollen shed and seed set. In contrast, ms26.78Δ / ms26.78Δ plants (Figure 13B) extruded small shriveled anthers, did not shed pollen and did not set seed. In contrast to examination of anthers from MS26/ ms26.78Δ plants (Figure 14A), anthers from ms26.78Δ / ms26.78Δ plants were small (Figure 14B). In addition, when anthers from these plants were more closely examined, pollen was easily detected in MS26/ ms26.78Δ anthers (Figure 14C), however pollen was not observed from anthers from ms26.78Δ / ms26.78Δ plants (Figure 14D). Good correlation of the fertility phenotype and the MS26 genotype was observed. In summary, all MS26/MS26 and MS26/ ms26.78Δ plants were male fertile, while all ms26.78Δ / ms26.78Δ plants were male sterile (Table 5). These male sterile plants were female fertile as demonstrated by their ability to set seed when fertilized with wild-type sorghum pollen (data not shown).

TABLE 5.

Fertility scores of Sorghum plants

Genotype	Fertile	Sterile
MS26/MS26	3	0
MS26/ ms26.78Δ	7	0
ms26.78Δ / ms26.78Δ	0	8

All publications and patent applications mentioned in the specification are indicative of the level of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

THAT WHICH IS CLAIMED:

1. A method for making a targeted modification in a male fertility gene in the genome of a plant, said method comprising:

(a) contacting at least one plant cell comprising a target sequence in a male fertility gene with an engineered double-strand-break-inducing agent capable of inducing a double-strand break at the target sequence; and

(b) identifying at least one cell from step (a) comprising an alteration in its genome at the target sequence wherein the alteration is selected from the group consisting of (i) replacement of at least one nucleotide, (ii) a deletion of at least one nucleotide, (iii) an insertion of at least one nucleotide, and (iv) any combination of (i) through (iii);

wherein the alteration of the male fertility gene is a null mutation.

2. The method of claim 1, wherein the male fertility gene is selected from the group consisting of *MS26, MS45, BS92-7, 5126* and *Msca1.*

3. The method of claim 1, further comprising regenerating a fertile plant comprising the alteration.

4. The method of claim 1, wherein the engineered double-strand-break-inducing agent is an endonuclease, a zinc finger nuclease, a TAL effector nuclease, a transposase, or a site-specific recombinase.

5. The method of claim 4, wherein the endonuclease is modified to specifically cut at the target sequence, and wherein the endonuclease no longer cuts at its wild-type endonuclease target sequence.

6. The method of claim 1, wherein a plant that is homozygous for the null mutation is male sterile.

7. The method of claim 3, further comprising selfing the fertile plant and selecting a progeny plant resulting therefrom, wherein said progeny plant is homozygous for the alteration.

8. The method of claim 3, further comprising crossing the fertile plant with a second fertile plant comprising a null mutation in the male fertility gene and selecting a progeny plant resulting therefrom, wherein said progeny plant is male sterile.

9. The method of claim 1, wherein the alteration comprises insertion of a transgene comprising a polynucleotide of interest.

10. The method of claim 9, wherein the transgene further comprises a promoter operably linked to the polynucleotide of interest, and wherein the promoter is capable of driving the expression of the polynucleotide of interest in a plant.

11. The method of claim 9, wherein the polynucleotide of interest encodes a phenotypic marker or an RNA or protein providing an agronomic advantage to the plant.

12. The method of claim 1, wherein the plant is selected from the group consisting of maize, sorghum, rice, wheat, rye, barley, millet and oat.

13. The method of claim 1, wherein the male fertility gene is *MS26*.

14. The method of claim 13, wherein the target sequence comprises the nucleotide sequence set forth in SEQ ID NO: 1.

15. The method of claim 13, wherein the engineered double-strand-break-inducing agent is derived from I-CreI.

16. The method of claim 14, wherein step (a) comprises introducing into the at least one plant cell a nucleic acid construct comprising a nucleotide sequence encoding the engineered double-strand-break-inducing agent.

17. The method of claim 16, wherein the nucleotide sequence is selected from the group consisting of:

(a) the nucleotide sequence set forth in SEQ ID NO: 4, 5, 6, or 7; and

(b) a nucleotide sequence having at least 80% nucleotide sequence identity to at least one nucleotide sequence selected from the group consisting of the nucleotide sequences set forth in SEQ ID NOS: 4, 5, 6, and 7, wherein the nucleotide sequence encodes a polypeptide comprising endonuclease activity.

18. The method of claim 17, wherein the nucleic acid construct further comprises a promoter operably linked to the nucleotide sequence encoding the engineered double-strand-break-inducing agent, wherein the promoter is capable of driving expression of the nucleotide sequence in a plant cell.

19. The method of claim 18, wherein the promoter is a maize ubiquitin promoter.

20. The method of claim 17, wherein the nucleic acid construct further comprises an operably linked coding sequence for a nuclear localization signal.

21. The method of claim 20, wherein the nuclear localization signal comprises an amino acid sequence selected from the group consisting of SEQ ID NOS: 2, 3, and 21.

22. The method of claim 1, wherein the male fertility gene is *MS45*.

23. The method of claim 22, wherein the target sequence comprises the nucleotide sequence set forth in SEQ ID NO: 20.

24. The method of claim 22, wherein the engineered double-strand-break-inducing agent is derived from I-CreI.

25. The method of claim 23, wherein step (a) comprises introducing into the at least one plant cell a nucleic acid construct comprising a nucleotide sequence encoding all or part of the engineered double-strand-break-inducing agent.

26. The method of claim 25, wherein the nucleotide sequence is selected from the group consisting of:

(a) the nucleotide sequence set forth in SEQ ID NO: 22, 23, or 34; and

(b) a nucleotide sequence having at least 80% nucleotide sequence identity to at least one nucleotide sequence selected from the group consisting of the nucleotide sequences set forth in SEQ ID NOS: 22, 23, and 34, wherein the nucleotide sequence encodes a polypeptide comprising endonuclease activity.

27. The method of claim 26, wherein the engineered double-strand-break-inducing agent comprises a first polypeptide and a second polypeptide.

28. The method of claim 26, wherein the nucleic acid construct further comprises a promoter operably linked to the nucleotide sequence encoding the engineered double-strand-break-inducing agent, wherein the promoter is capable of driving expression of the nucleotide sequence in a plant cell.

29. The method of claim 28, wherein the promoter is a maize ubiquitin promoter.

30. The method of claim 26, wherein the nucleic acid construct further comprises an operably linked coding sequence for a nuclear localization signal.

31. The method of claim 30, wherein the coding sequence for a nuclear localization signal is selected from the group consisting of SEQ ID NOS: 2, 3, and 21.

32. A plant produced by the method of claim 1, or a descendant of the plant produced by the method, wherein the descendant comprises the alteration.

33. The plant of claim 32, wherein the plant is a sorghum plant and the male fertility gene is *MS26*.

34. The plant of claim 33, wherein the altered *MS26* gene comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77 and 78.

35. A plant comprising a targeted modification in a male fertility gene in its genome, wherein the targeted modification is a targeted mutation or the insertion of a transgene, and wherein the male fertility gene is selected from the group consisting of *MS26, MS45, BS92-7, 5126 and Msca1.*

36. The plant of claim 35, wherein the targeted modification is a null mutation in the male fertility gene.

37. The plant of claim 36, wherein the plant is homozygous for the alteration.

38. The plant of claim 37, wherein the plant is male sterile.

39. The plant of claim 35, wherein the plant is selected from the group consisting of maize, sorghum, rice, wheat, rye, barley, millet, and oat.

40. The plant of claim 39, wherein the plant is a sorghum plant comprising a targeted mutation in the male fertility gene *MS26*.

41. The plant of claim 40, wherein the *MS26* gene comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77 and 78.

42. An isolated nucleic acid molecule comprising a male fertility gene produced by the method of claim 1, wherein the male fertility gene comprises the alteration.

43. A plant comprising an expression construct, said expression construct comprising a promoter operably linked to a nucleotide sequence encoding an endonuclease, wherein the endonuclease is capable of specifically binding to and creating a double strand break in a target sequence selected from the group consisting of SEQ ID NOS:1 and 20, and wherein the promoter is capable of driving expression of an operably linked nucleotide sequence in a plant cell.

44. The plant of claim 43, wherein the nucleotide sequence comprises a coding sequence of a DNA binding domain of an endonuclease, and wherein the coding sequence is selected from the group consisting of:

(a) nucleotides 100-261 and nucleotides 661-822 of SEQ ID NO: 4;

(b) nucleotides 70-231 and nucleotides 631-792 of SEQ ID NO: 5;

(c) nucleotides 70-231 and nucleotides 820-981 of SEQ ID NO: 6, 7 or 34; and

(d) a degenerate coding sequence of (a), (b), or (c).

45. The plant of claim 43, wherein the nucleotide sequence is selected from the group consisting of SEQ ID NOS: 4, 5, 6, 7, 22, 23, and 34.

46. An expression construct comprising a promoter operably linked to a nucleotide sequence selected from the group consisting of SEQ ID NOS: 4, 5, 6, 7, 22, 23, and 34.

FIG. 1
1/18

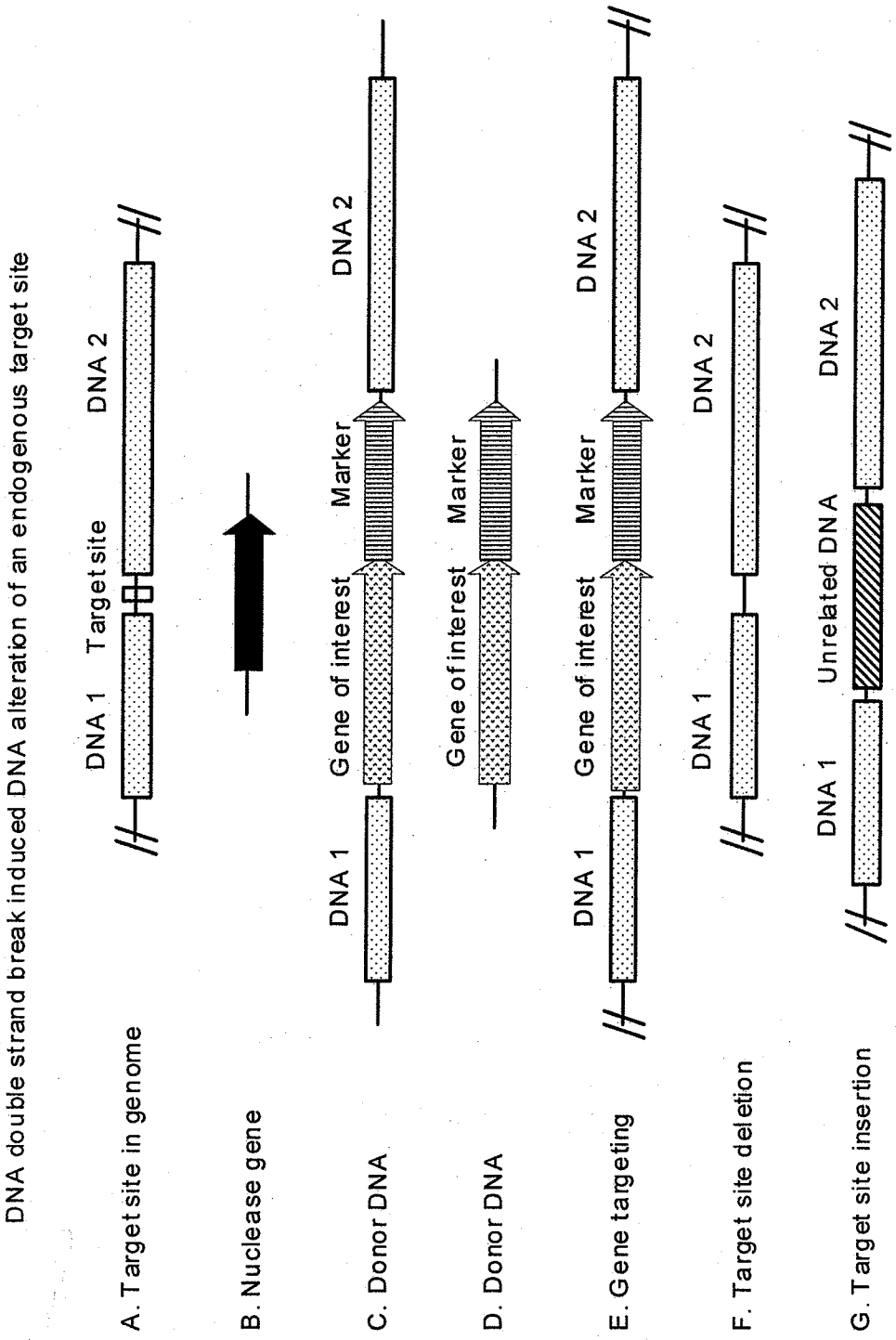

DNA double strand break induced DNA alteration of an endogenous target site
A. Target site in genome
DNA 1
Target site
DNA 2
B. Nuclease gene
C. Donor DNA
DNA 1
Gene of interest
Marker
DNA 2
D. Donor DNA
Gene of interest
Marker
E. Gene targeting
DNA 1
Gene of interest
Marker
DNA 2
F. Target site deletion
DNA 1
DNA 2
G. Target site insertion
DNA 1
Unrelated DNA
DNA 2

FIG. 2
2/18

Wild type	CGGCGG**GATGGTGACGTACGTGCCCTAC**TCGAT
3281	CGGCGG**GATGGTGACG TAC**TCGAT
2963	CGGCGG**GATGGTGACGT CGTGCCCTAC**TCGAT
2980	CGGCGG**GATGGTGACG ACGTGCCCTAC**TCGAT
3861	CGGCGG**GATGGTG ACGTGCCCTAC**TCGAT
3956	CGGCGG**GATGGTGAC TACGTGCCCTAC**TCGAT
3990	CGGCGG**GATGGTGACGT CGTGCCCTAC**TCGAT
6227	CGGCGG**GATGGTGACG ACGTGCCCTAC**TCGAT

FIG. 3
3/18

```
MS26 TS          ----------GATGGTGACGTACGTGCCCTAC----------
maize MS26       AGGTGAGGGCCGGCGGCATGGTGACGTACGTGCCCTACTCGATGGGGCGG
sorghum MS26     AGGTGAGGGCCGGCGGCATGGTGACGTACGTGCCCTACTCGATGGGGCGG
rice MS26        AGGTGCGCGCCGGCGGCATGGTGACGTACGTGCCCTACTCCATGGGGAGG
rye MS26         AGGTGCGCCGCCGGCGGCATGGTGACGTACGTGCCCTACTCCATGGGGCGG
```

FIG. 4
4/18

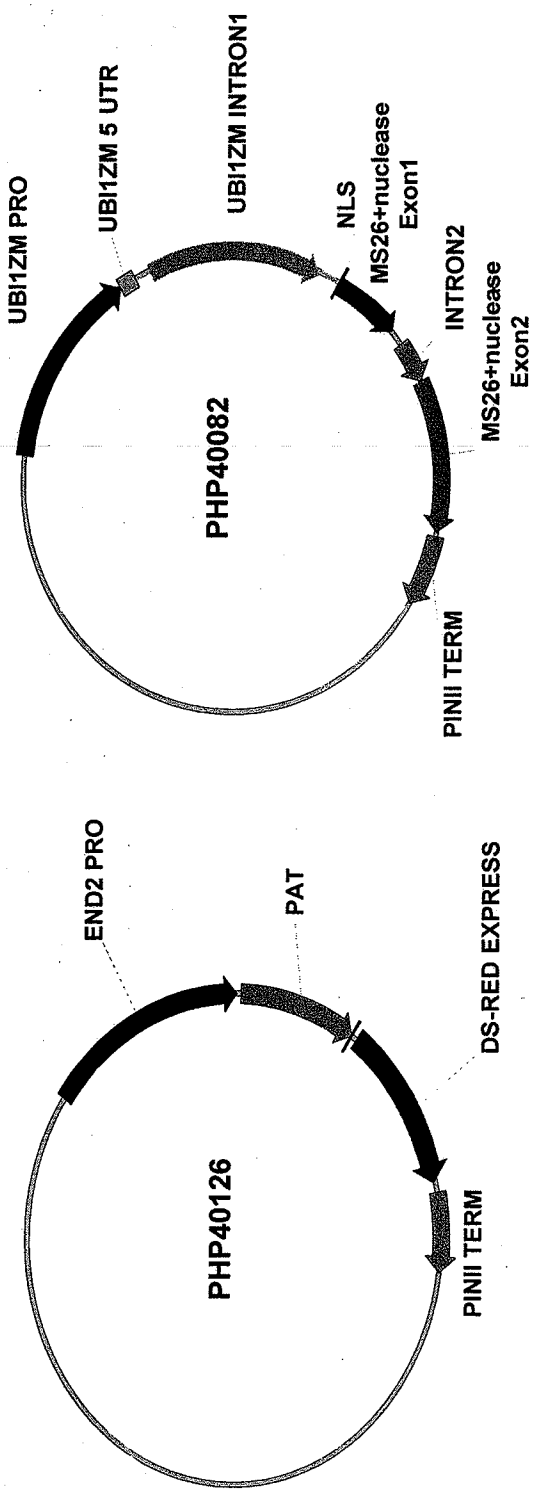

UBI1ZM PRO
UBI1ZM 5 UTR
UBI1ZM INTRON1
NLS
MS26+nuclease Exon1
INTRON2
MS26+nuclease Exon2
PHP40082
PINII TERM
END2 PRO
PAT
DS-RED EXPRESS
PHP40126
PINII TERM

FIG. 5
5/18

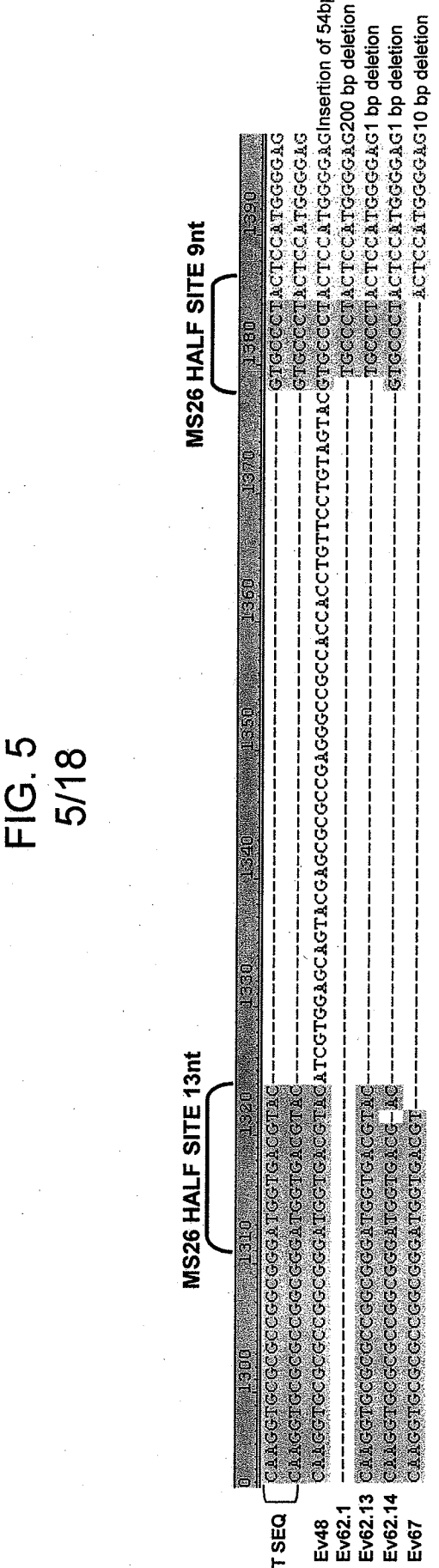

MS26 HALF SITE 13nt
MS26 HALF SITE 9nt
1300
1310
1320
1330
1340
1350
1360
1370
1380
1390
WT SEQ
Ev48
Ev62.1
Ev62.13
Ev62.14
Ev67
Insertion of 54bp RFP
200 bp deletion
1 bp deletion
1 bp deletion
10 bp deletion

FIG. 6
6/18

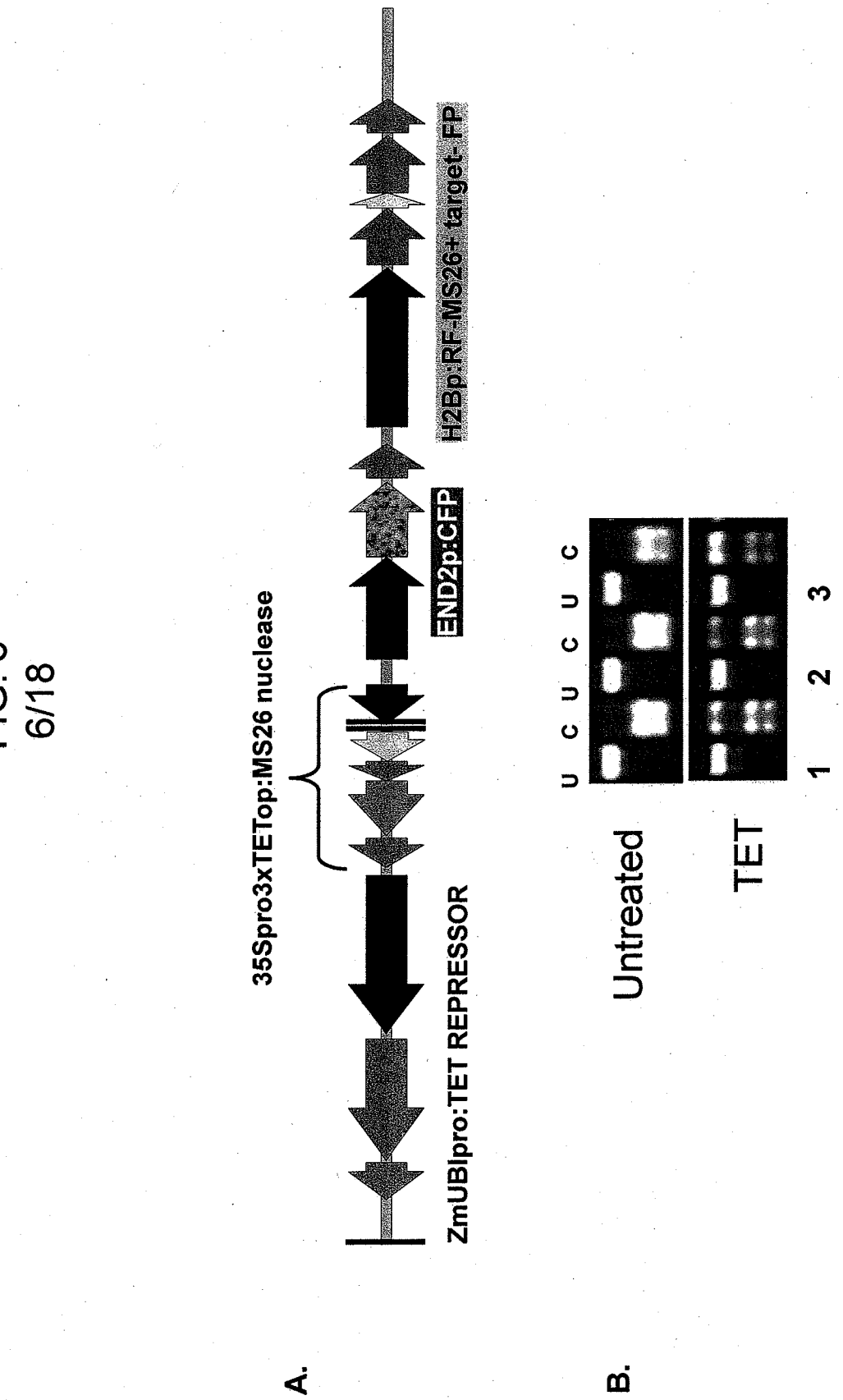

A.
35Spro3xTETop:MS26 nuclease
ZmUBIpro:TET REPRESSOR
END2p:CFP
H2Bp:RF-MS26+ target- FP
B.
Untreated
TET
U C U C U C
1 2 3

FIG. 7
7/18

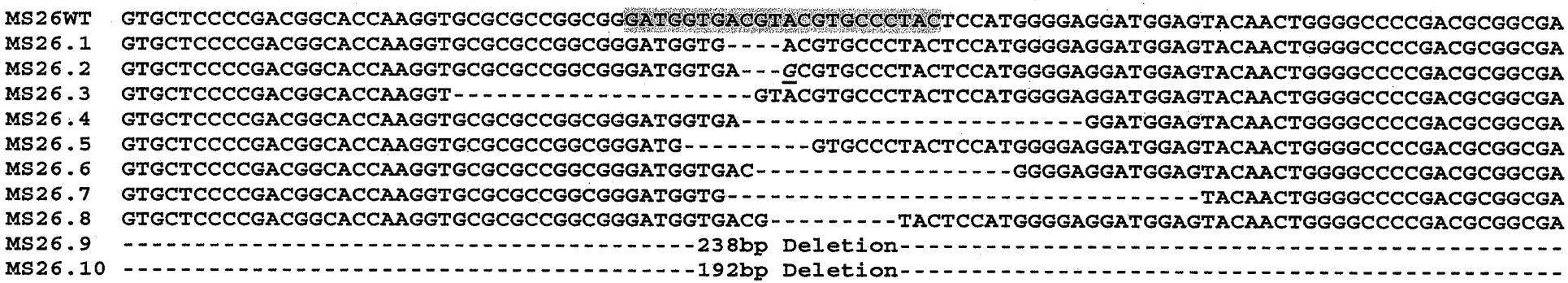

MS26WT GTGCTCCCCGACGGCACCAAGGTGCGCGCCGGCGGGATGGTGACGTACGTGCCCTACTCCATGGGGAGGATGGAGTACAACTGGGGCCCCGACGCGGCGA
MS26.1 GTGCTCCCCGACGGCACCAAGGTGCGCGCCGGCGGGATGGTG----ACGTGCCCTACTCCATGGGGAGGATGGAGTACAACTGGGGCCCCGACGCGGCGA
MS26.2 GTGCTCCCCGACGGCACCAAGGTGCGCGCCGGCGGGATGGTGA---GCGTGCCCTACTCCATGGGGAGGATGGAGTACAACTGGGGCCCCGACGCGGCGA
MS26.3 GTGCTCCCCGACGGCACCAAGGT----------GTACGTGCCCTACTCCATGGGGAGGATGGAGTACAACTGGGGCCCCGACGCGGCGA
MS26.4 GTGCTCCCCGACGGCACCAAGGTGCGCGCCGGCGGGATGGTGA----------GGATGGAGTACAACTGGGGCCCCGACGCGGCGA
MS26.5 GTGCTCCCCGACGGCACCAAGGTGCGCGCCGGCGGGATG---------GTGCCCTACTCCATGGGGAGGATGGAGTACAACTGGGGCCCCGACGCGGCGA
MS26.6 GTGCTCCCCGACGGCACCAAGGTGCGCGCCGGCGGGATGGTGAC----------GGGGAGGATGGAGTACAACTGGGGCCCCGACGCGGCGA
MS26.7 GTGCTCCCCGACGGCACCAAGGTGCGCGCCGGCGGGATGGTG----------TACAACTGGGGCCCCGACGCGGCGA
MS26.8 GTGCTCCCCGACGGCACCAAGGTGCGCGCCGGCGGGATGGTGACG---------TACTCCATGGGGAGGATGGAGTACAACTGGGGCCCCGACGCGGCGA
MS26.9 ----------238bp Deletion----------
MS26.10 ----------192bp Deletion----------

FIG. 8
8/18

FIG. 9
9/18

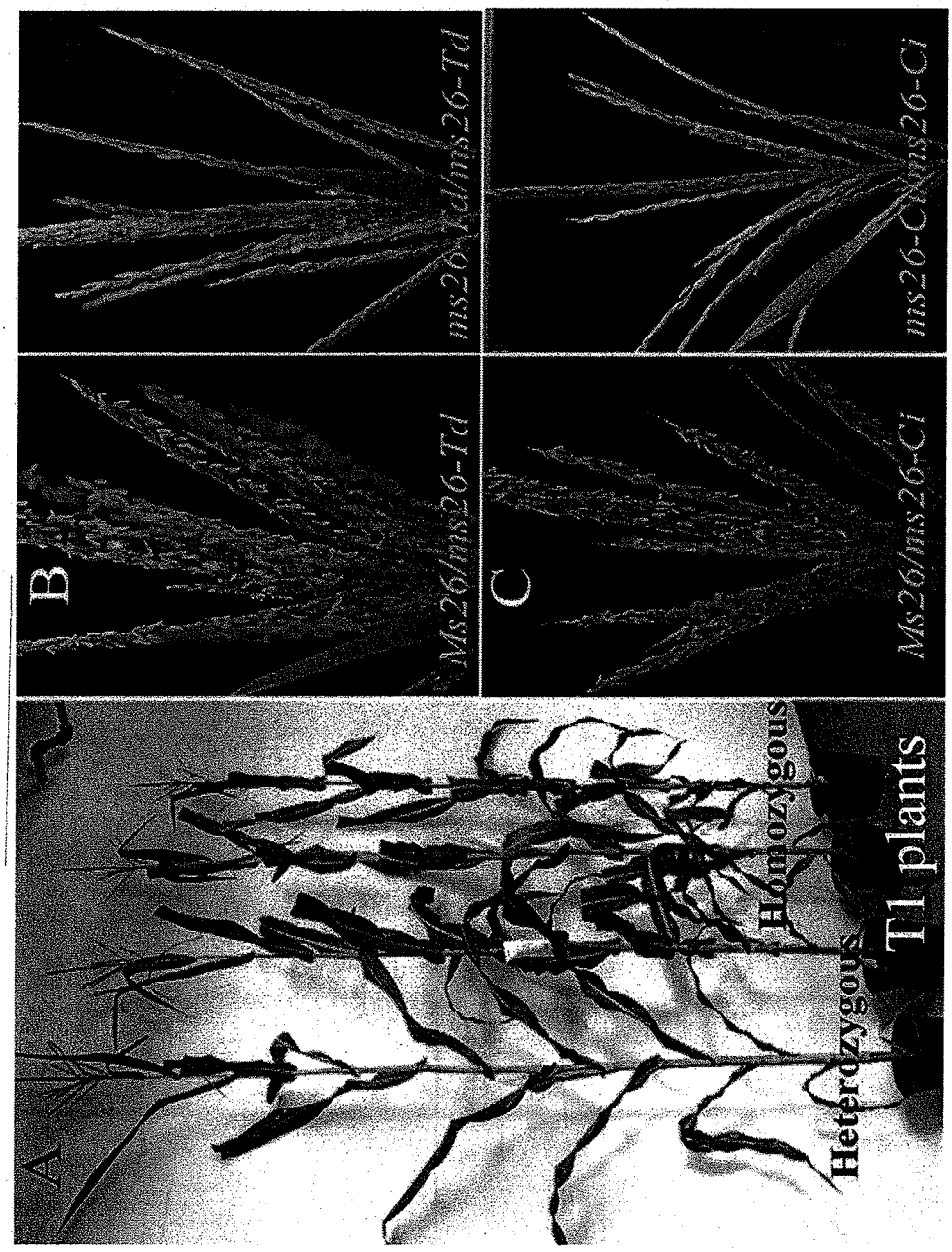

A
Heterozygous
Homozygous
T1 plants
B
Ms26/ms26-Td
ms26-Td/ms26-Td
C
Ms26/ms26-Ci
ms26-Ci/ms26-Ci

FIG. 10
10/18

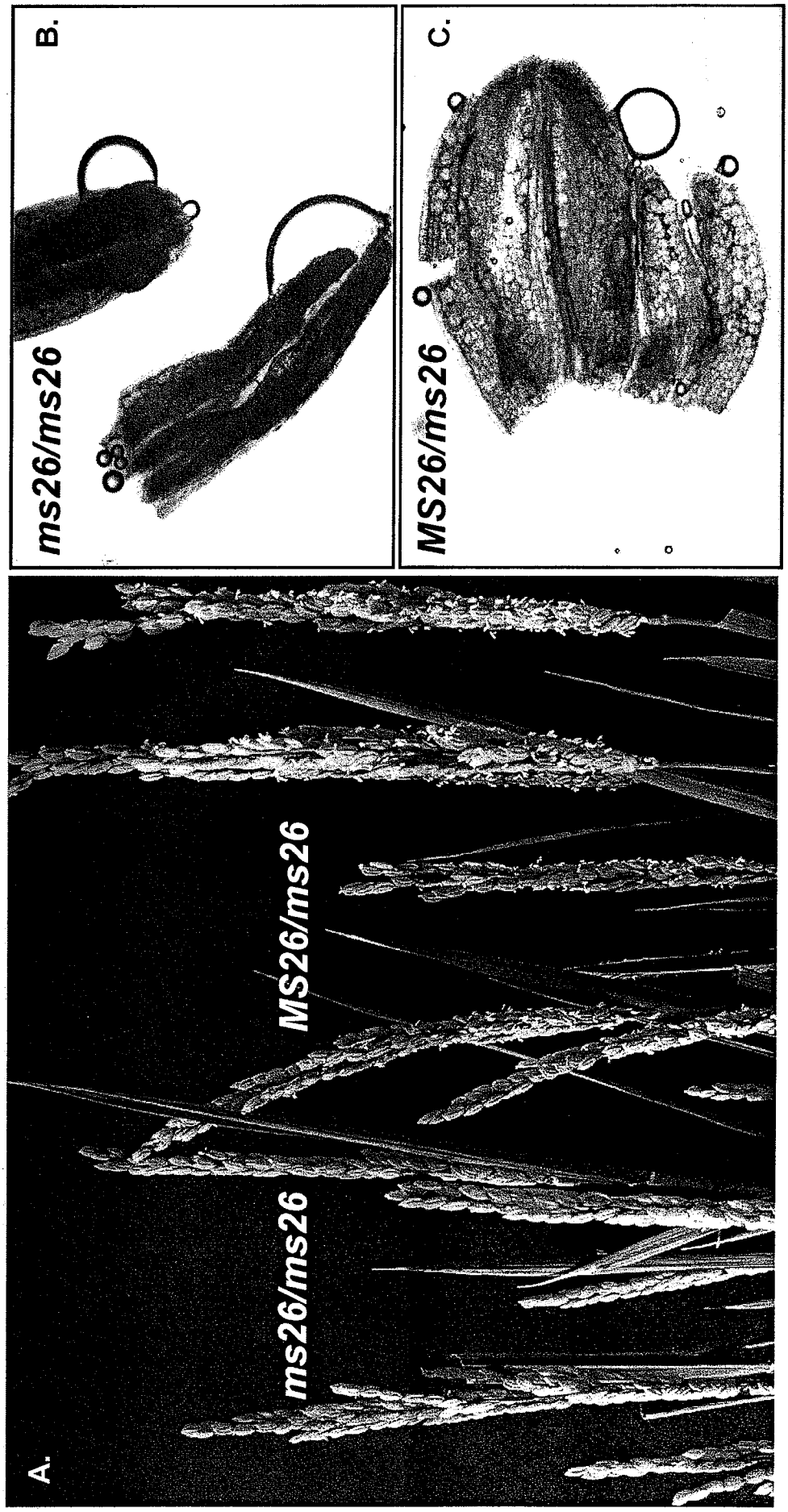


FIG. 11A
11/18

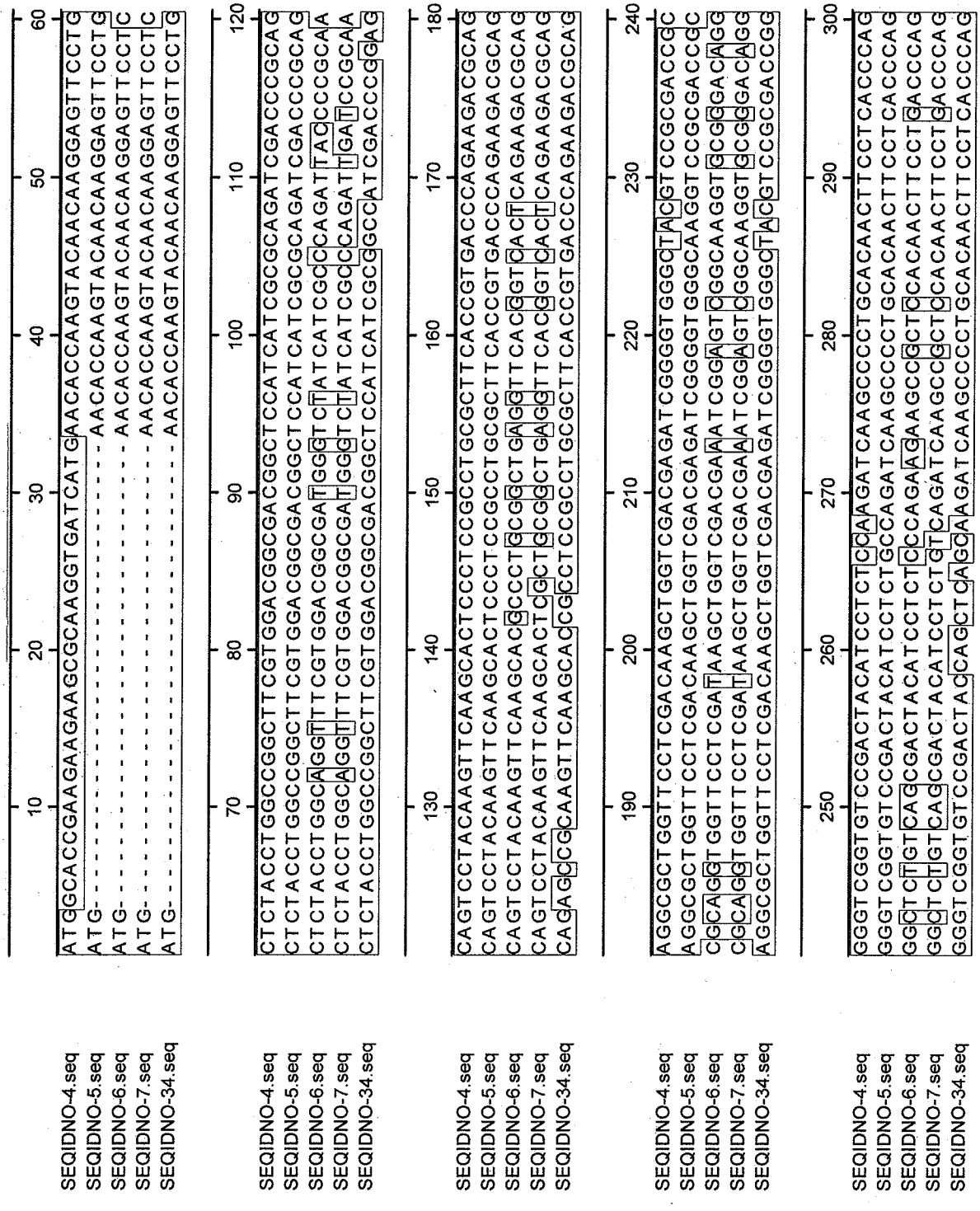

SEQIDNO-4.seq
SEQIDNO-5.seq
SEQIDNO-6.seq
SEQIDNO-7.seq
SEQIDNO-34.seq

FIG. 11B
12/18

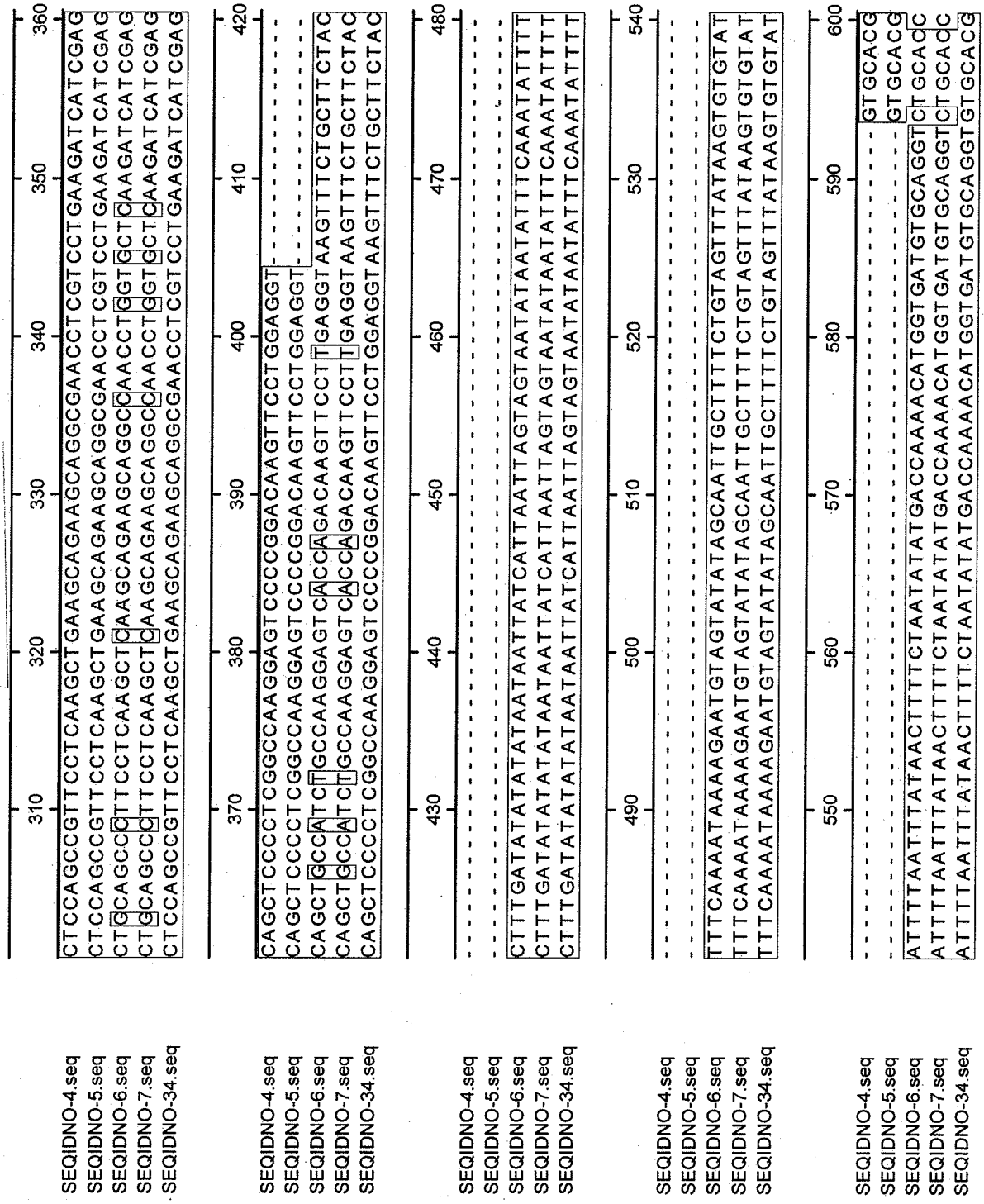

SEQIDNO-4.seq
SEQIDNO-5.seq
SEQIDNO-6.seq
SEQIDNO-7.seq
SEQIDNO-34.seq

FIG. 11C
13/18

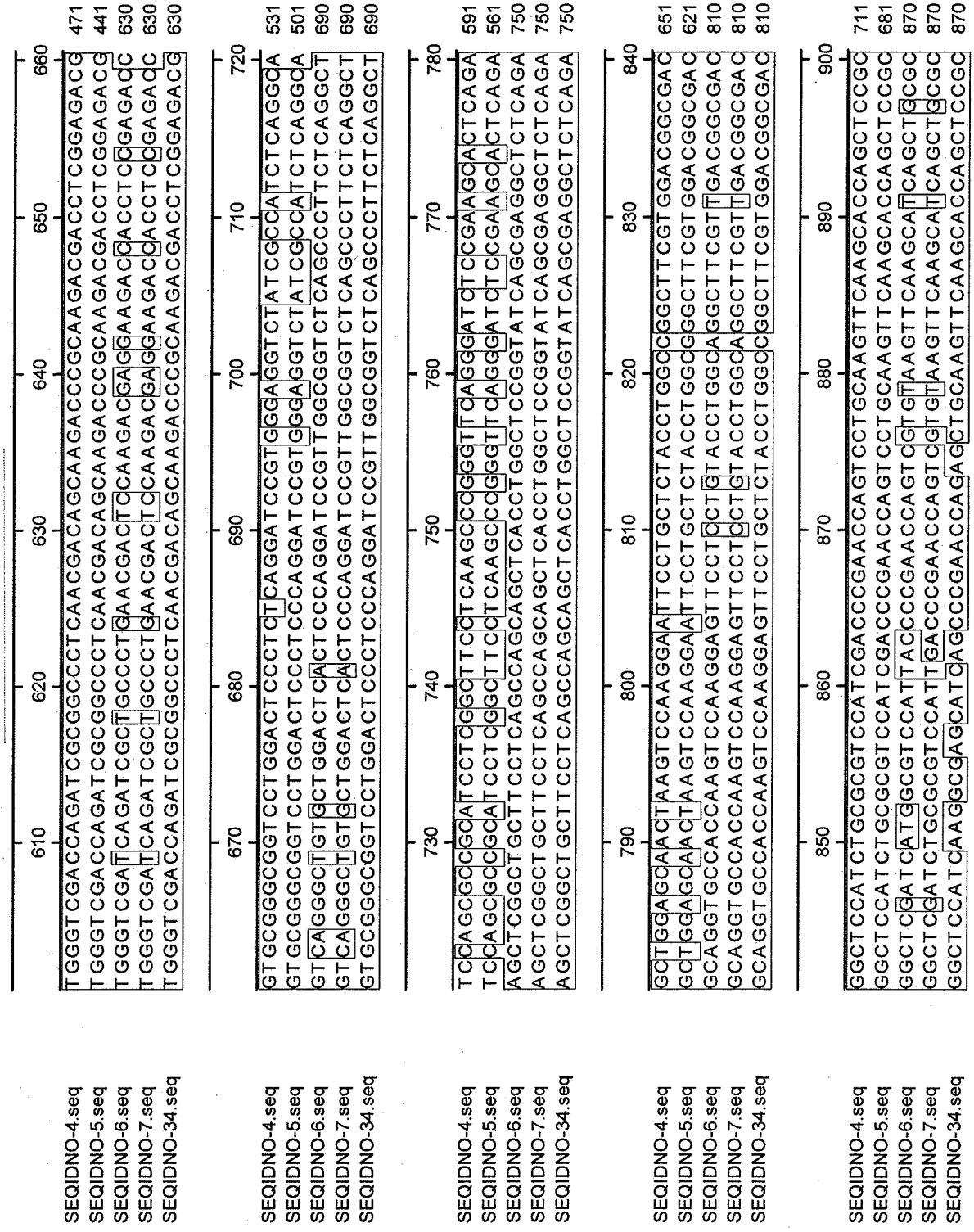

610 620 630 640 650 660
SEQIDNO-4.seq TGGGTCGACCAGATCGCGGCCCTCAACGACAGCAAGACCCGCAAGACCGACCTCGGAGACG 471
SEQIDNO-5.seq TGGGTCGACCAGATCGCGGCCCTCAACGACAGCAAGACCCGCAAGACCGACCTCGGAGACG 441
SEQIDNO-6.seq TGGGTCGATCAGATCGCTGCCCTGAACGACTCCAAGACGAGGAAGACGACCTCGGAGACC 630
SEQIDNO-7.seq TGGGTCGATCAGATCGCTGCCCTGAACGACTCCAAGACGAGGAAGACGACCTCGGAGACC 630
SEQIDNO-34.seq TGGGTCGACCAGATCGCGGCCCTCAACGACAGCAAGACCCGCAAGACCGACCTCGGAGACG 630
670 680 690 700 710 720
SEQIDNO-4.seq GTGCGGGCGGTCCTGGACTCCCTCTCAGGATCCGTGGGAGGTCTATCGCCATCTCAGGCA 531
SEQIDNO-5.seq GTGCGGGCGGTCCTGGACTCCCTCCCAGGATCCGTGGGAGGTCTATCGCCATCTCAGGCA 501
SEQIDNO-6.seq GTCAGGGCTGTGCTGGACTCACTCCCAGGATCCGTTTGGCGGTCTCAGCCCTTCTCAGGCT 690
SEQIDNO-7.seq GTCAGGGCTGTGCTGGACTCACTCCCAGGATCCGTTTGGCGGTCTCAGCCCTTCTCAGGCT 690
SEQIDNO-34.seq GTGCGGGCGGTCCTGGACTCCCTCCCAGGATCCGTTTGGCGGTCTCAGCCCTTCTCAGGCT 690
730 740 750 760 770 780
SEQIDNO-4.seq TCCAGCGCCGCATCCTCGGCTTCCTCAAGCCCGGGTTCAGGGATCTCCGAAGCACTCAGA 591
SEQIDNO-5.seq TCCAGCGCCGCATCCTCGGCTTCCTCAAGCCCGGGTTCAGGGATCTCCGAAGCACTCAGA 561
SEQIDNO-6.seq AGCTCGGCTGCTTCCTCAGCCAGCAGCTCACCTGGCTCCGGTATCAGCGAGGCTCTCAGA 750
SEQIDNO-7.seq AGCTCGGCTGCTTCCTCAGCCAGCAGCTCACCTGGCTCCGGTATCAGCGAGGCTCTCAGA 750
SEQIDNO-34.seq AGCTCGGCTGCTTCCTCAGCCAGCAGCTCACCTGGCTCCGGTATCAGCGAGGCTCTCAGA 750
790 800 810 820 830 840
SEQIDNO-4.seq GCTGGAGCAACTAAGTCCAAGGAATTCCTGCTCTACCTGGCCGGCTTCGTGGACGGCGAC 651
SEQIDNO-5.seq GCTGGAGCAACTAAGTCCAAGGAATTCCTGCTCTACCTGGCGGGCTTCGTGGACGGCGAC 621
SEQIDNO-6.seq GCAGGTGCCACCAAGTCCAAGGAGTTCCTCCTGTACCTGGCAGGCTTCGTTGACGGCGAC 810
SEQIDNO-7.seq GCAGGTGCCACCAAGTCCAAGGAGTTCCTCCTGTACCTGGCAGGCTTCGTTGACGGCGAC 810
SEQIDNO-34.seq GCAGGTGCCACCAAGTCCAAGGAGTTCCTGCTCTACCTGGCCGGCTTCGTGGACGGCGAC 810
850 860 870 880 890 900
SEQIDNO-4.seq GGCTCCATCTGCGCGTCCATCGACCCGAACCAGTCCTGCAAGTTCAAGCACCAGCTCCGC 711
SEQIDNO-5.seq GGCTCCATCTGCGCGTCCATCGACCCGAACCAGTCCTGCAAGTTCAAGCACCAGCTCCGC 681
SEQIDNO-6.seq GGCTCGGATCATGGCGTCCATTACCCCGAACCAGTCGTGTAAGTTCAAGCATCAGCTGCGC 870
SEQIDNO-7.seq GGCTCGGATCTGCGCGTCCATTGACCCGAACCAGTCGTGTAAGTTCAAGCATCAGCTGCGC 870
SEQIDNO-34.seq GGCTCCATCAAGGCGAGCATCAGCCCGAACCAGAGCTGCAAGTTCAAGCACCAGCTCCGC 870

FIG. 11D
14/18

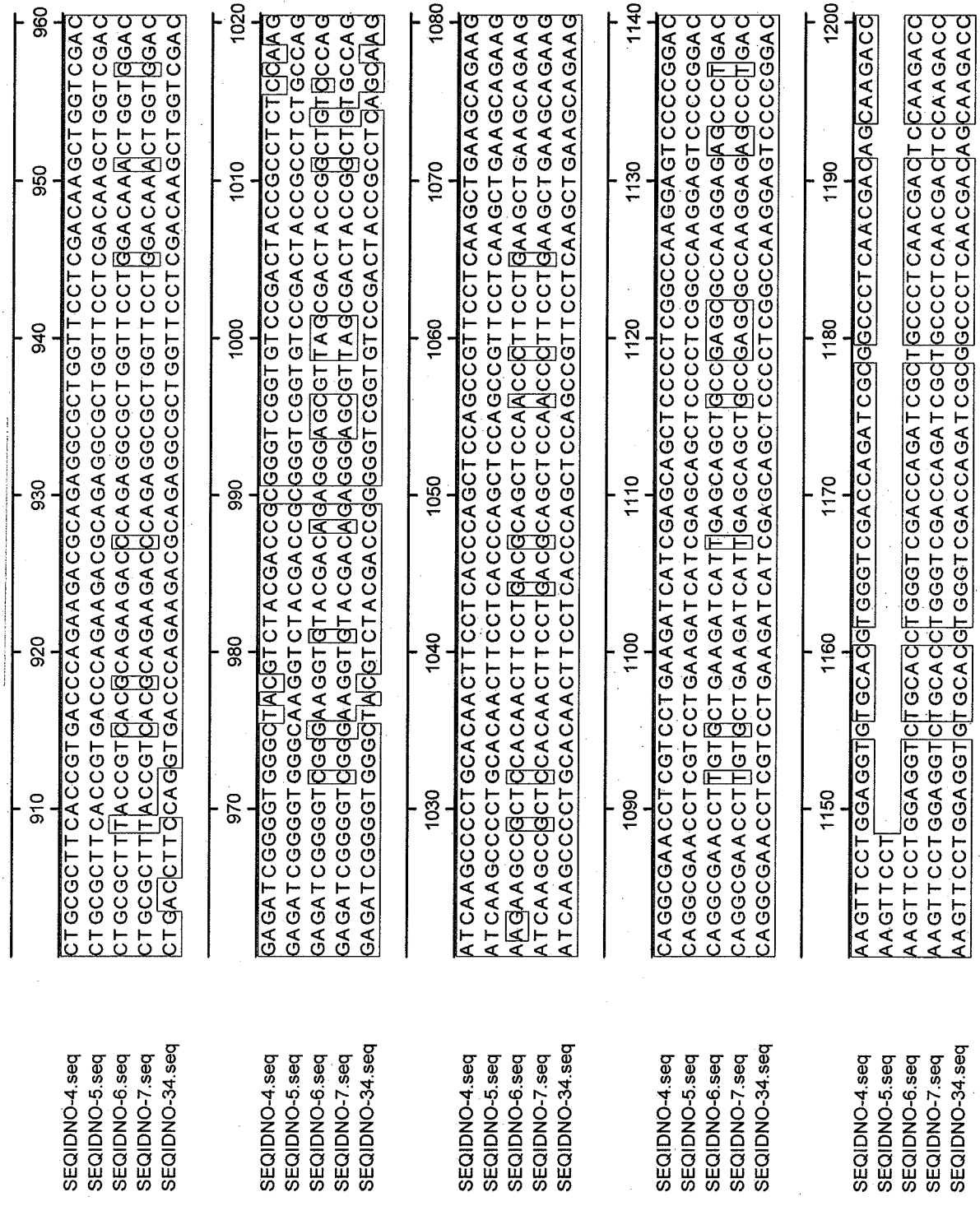

SEQIDNO-4.seq
SEQIDNO-5.seq
SEQIDNO-6.seq
SEQIDNO-7.seq
SEQIDNO-34.seq
910
920
930
940
950
960
771
741
930
930
970
980
990
1000
1010
1020
831
801
990
990
1030
1040
1050
1060
1070
1080
891
861
1050
1050
1090
1100
1110
1120
1130
1140
951
921
1110
1110
1150
1160
1170
1180
1190
1200
1011
929
1170
1170
1170

FIG. 11E
15/18

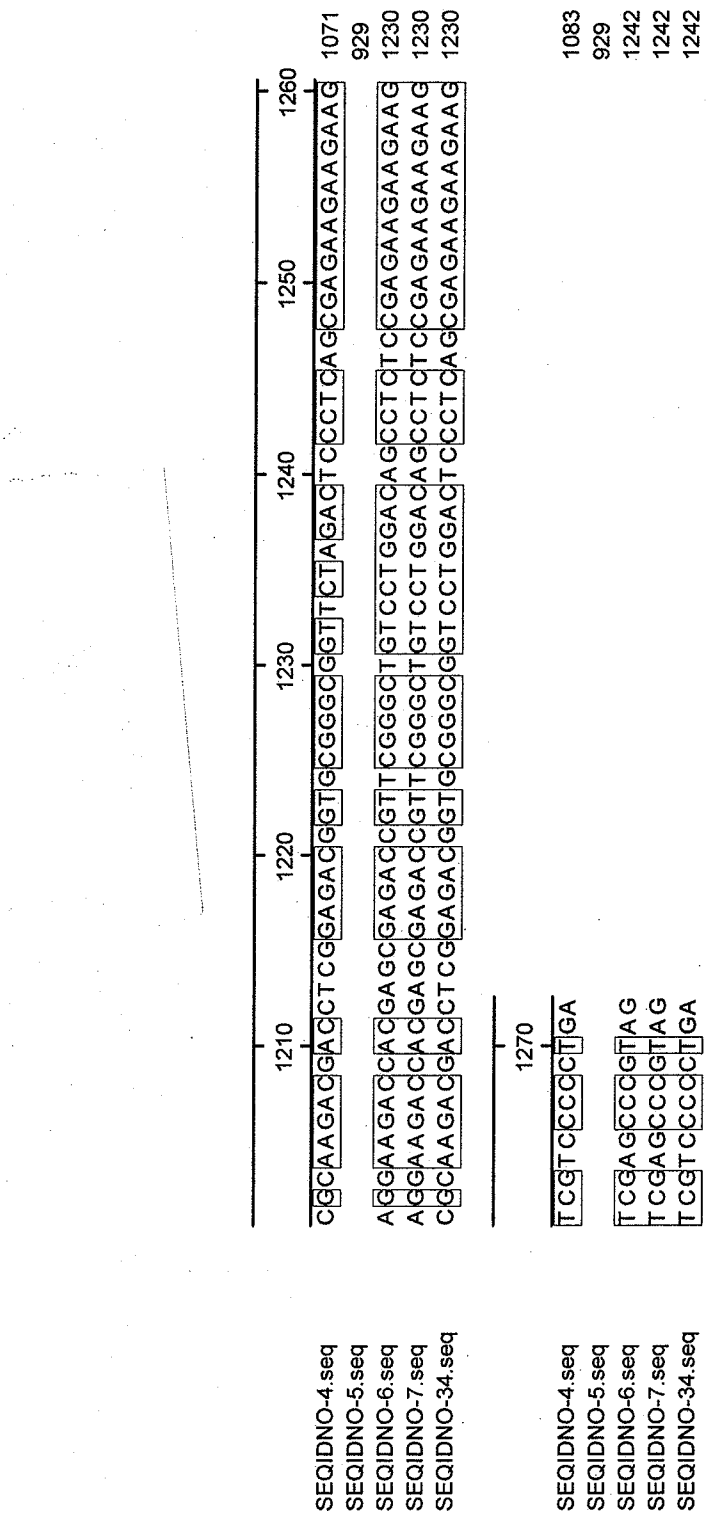
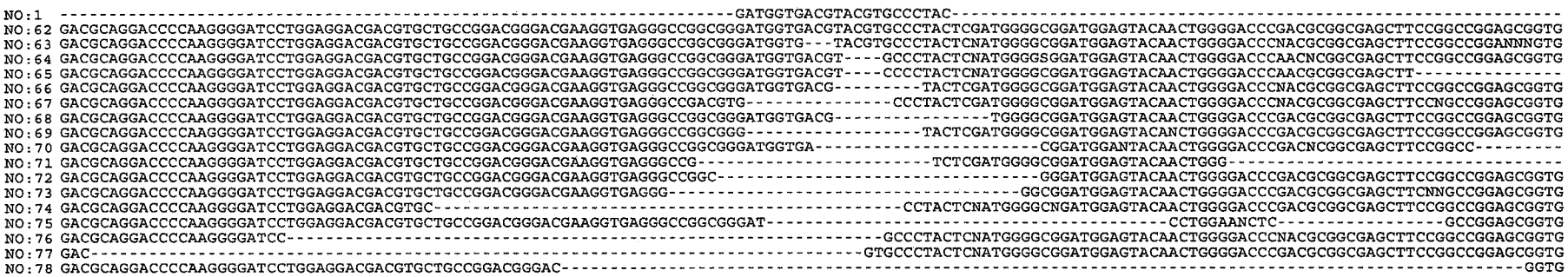

1210
1220
1230
1240
1250
1260
1270
SEQIDNO-4.seq
SEQIDNO-5.seq
SEQIDNO-6.seq
SEQIDNO-7.seq
SEQIDNO-34.seq
1071
929
1230
1230
1230
1083
929
1242
1242
1242

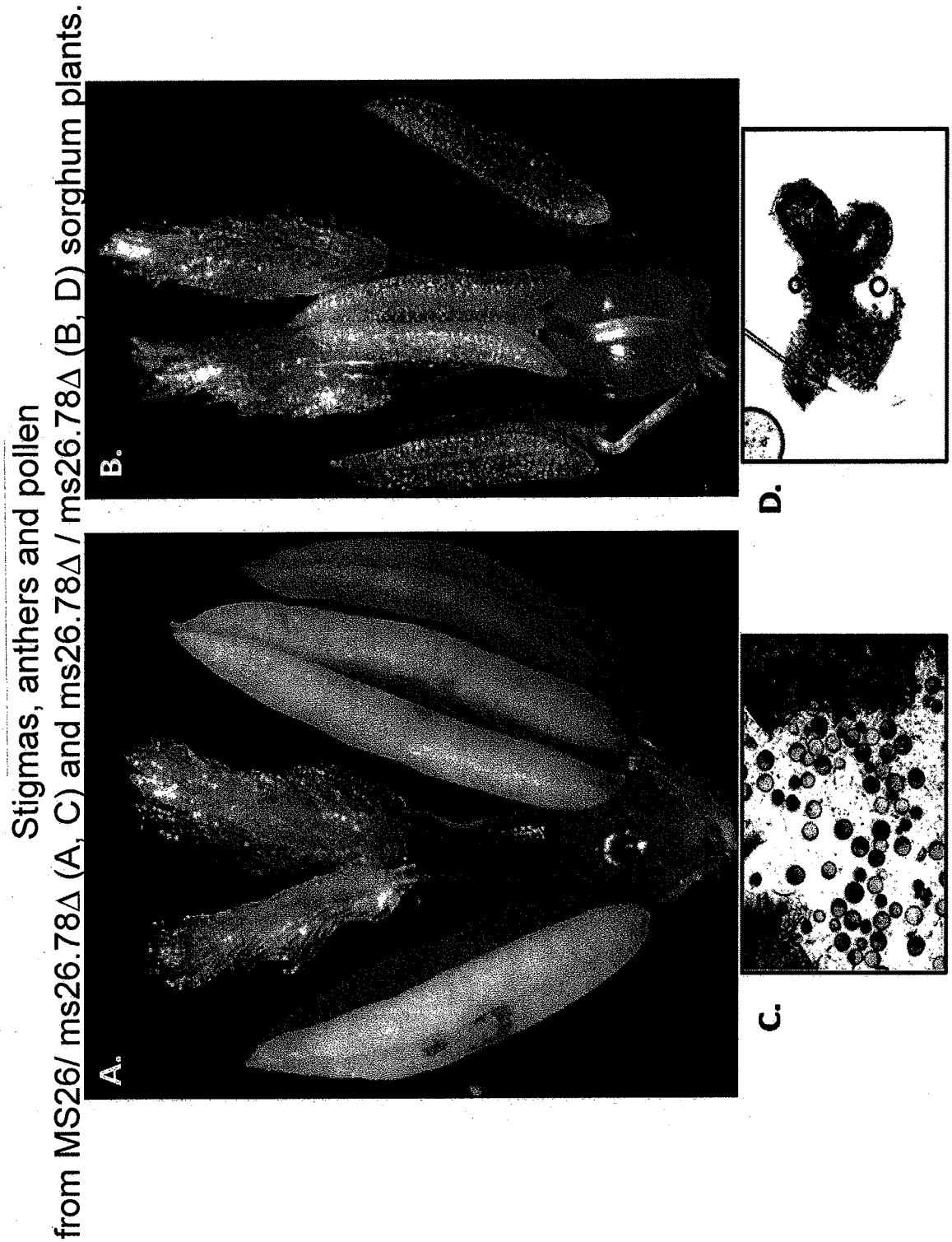
FIG. 12
16/18

Sorghum MS26 mutations and deletions



Flowers from MS26/ ms26.78Δ (A) and ms26.78Δ / ms26.78Δ (B) sorghum plants.

FIG. 14
18/18


Stigmas, anthers and pollen
from MS26/ ms26.78Δ (A, C) and ms26.78Δ / ms26.78Δ (B, D) sorghum plants.
A.
B.
C.
D.