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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 February 2013 (07.02.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/019411 A1

(51) International Patent Classification:
C12N 15/82 (2006.01) *A01H 5/00* (2006.01)

(21) International Application Number:
PCT/US2012/047202

(22) International Filing Date:
18 July 2012 (18.07.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/514,565 3 August 2011 (03.08.2011) US

(71) Applicants *(for all designated States except US)*: **E. I. DU PONT DE NEMOURS AND COMPANY** [US/US]; 1007 Market Street, Wilmington, Delaware 19899 (US). **PIONEER HI-BRED INTERNATIONAL** [US/US]; 7100 N.W. 62nd Avenue, P.O. Box 1014, Johnston, Iowa 50131 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CIGAN, Andrew Mark** [US/US]; 5764 Chatham Circle, Johnston, Iowa 50131 (US). **FALCO, Saverio Carl** [US/US]; 1902 Millers Road, Wilmington, Delaware 19810 (US). **GAO, Huirong** [US/US]; 7068 Ridgedale Court, Johnston, Iowa 50131 (US). **LASSNER, Michael W.** [US/US]; 2935 133rd Way, Urbandale, Iowa 50323 (US). **LI, Zhongsen** [US/US]; 24 Forest Creek Drive, Hockessin, Delaware 19707 (US). **LIU, Zhan-Bin** [US/US]; 1357 Wooded Knoll, West Chester, Pennsylvania 19382 (US). **LYZNIK, Leszek A.** [PL/US]; 5960 Greendale Place, #105, Johnston, Iowa 50131 (US). **PETERSON, David J.** [US/US]; 2126 Stevenson Drive, Ames, Iowa 50010 (US). **SCELONGE, Christopher Jay** [US/US]; 5314 Orvid Avenue, Des Moines, Iowa 50310 (US).

(74) Agent: **STOOP, Johan M.**; E. I. du Pont de Nemours and Company, Legal Patent Records Center, 4417 Lancaster Pike, Wilmington, Delaware 19805 (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:

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

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

WO 2013/019411 A1

(54) Title: METHODS AND COMPOSITIONS FOR TARGETED INTEGRATION IN A PLANT

(57) Abstract: Compositions and methods are provided for the targeted integration of a polynucleotide sequence of interest into the genome of a plant or plant cell. The methods and compositions employ recognition sites for endonucleases and endonucleases in combination with site-specific recombination sites/recombinases to provide an effective system for establishing target sites within the genome of a plant, plant cell or seed. Once such target sites are established, a variety of methods can be employed to further modify the target sites such that they contain a variety of polynucleotides of interest.

METHODS AND COMPOSITIONS FOR TARGETED INTEGRATION IN A PLANT

This application claims the benefit of U.S. Patent Application Serial Number 61/514565, filed August 3, 2011, which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

The invention relates to the field of plant molecular biology. In particular, methods and compositions are provided for altering the genome of a plant.

BACKGROUND OF THE INVENTION

Recombinant DNA technology has made it possible to insert foreign DNA sequences into the genome of an organism, thus, altering the organism's phenotype. The most commonly used plant transformation methods are *Agrobacterium* infection and biolistic particle bombardment in which transgenes integrate into a plant genome in a random fashion and with an unpredictable copy number.

Site-specific integration techniques, which employ site-specific recombination systems, as well as, other types of recombination technologies, have been used to generate targeted insertions of genes of interest in a variety of organism.

Other methods for inserting or modifying a DNA sequence involve homologous DNA recombination by introducing a transgenic DNA sequence flanked by sequences homologous to the genomic target. U.S. Patent No. 5,527,695 describes transforming eukaryotic cells with DNA sequences that are targeted to a predetermined sequence of the eukaryote's DNA. Transformed cells are identified through use of a selectable marker included as a part of the introduced DNA sequences.

While both systems have provided useful techniques for targeted insertion of sequences of interest, there remains a need for methods and compositions which improve these systems and allow for a gene of interest to be targeted to a specific site in the plant genome.

BRIEF SUMMARY OF THE INVENTION

Methods and compositions for targeting a polynucleotide of interest to a specific site in the genome of a plant or plant cell are provided.

Methods for introducing into the genome of a plant cell a target site for site-specific integration are provided. The method comprises providing a plant cell comprising in its genome an endogenous recognition site for an engineered endonuclease, wherein the engineered endonuclease is capable of inducing a double-strand break in the endogenous recognition site, and wherein the endogenous recognition site is located between a first and a second genomic region. A donor DNA comprising a target site for site-specific integration located between a first region of homology to the first genomic region and a second region of homology to the second genomic region is provided, wherein the target site comprises a first and a second recombination site, and the first and the second recombination sites are dissimilar and non-recombinogenic with respect to one another. The plant cell is contacted with the donor DNA and the engineered endonuclease, and least one plant cell comprising in its genome the target site integrated at the endogenous recognition site is identified.

Further provided is a method of integrating a polynucleotide of interest into a specific site in the genome of a plant cell. The method comprises providing at least one plant cell comprising in its genome a target site for site-specific integration, wherein the target site is integrated into an endogenous recognition site for an engineered endonuclease, and wherein the target site is, (i) a target site comprising a first and a second recombination site, or (ii) the target site of (i) further comprising a third recombination site between the first recombination site and the second recombination site, wherein the engineered endonuclease is capable of inducing a double-strand break in the endogenous recognition site, wherein the first, the second, and the third recombination sites are dissimilar and non-recombinogenic with respect to one another. The plant cell is transformed with a transfer cassette comprising, (iii) the first recombination site, a first polynucleotide of interest, and the second recombination site, (iv) the second recombination site, a second polynucleotide of interest, and the third recombination site, or (v) the first recombination site, a third polynucleotide of interest, and the third recombination site. A recombinase is provided that recognizes and implements recombination at the first and the second recombination sites, at the second and the third recombination sites, or at the first and third recombination sites. At least one plant cell comprising integration of the transfer cassette at the target site is selected.

Various compositions include a plant, a seed or a plant cell comprising in its genome a target site for site-specific integration, wherein the target site is integrated into an endogenous recognition site for an engineered endonuclease, wherein the target site comprises in the following order: (a) a first recombination site; and, (b) a second recombination site, and wherein the engineered endonuclease is capable of inducing a double-strand break at the endogenous recognition site, wherein the first and the second recombination sites are dissimilar and non-recombinogenic with respect to one another.

BRIEF DESCRIPTION OF THE DRAWINGS AND 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.

Figure 1. Schematic of the DNA double-strand-break-induced DNA alteration of an endogenous recognition site to integrate a target site followed by FLP recombinase mediated site-specific integration at the target site.

Figure 2. Fragment of plasmid PHP36070 used to create the transgenic insertion at the LIG3-4 recognition site.

Figure 3. Fragment of plasmid PHP44779 used to create the transgenic insertion at the MHP14 recognition site.

Figure 4 . Graphic representation of the transgenic insertion site following homologous recombination mediated by meganuclease induced double stranded breaks. PCR and Southern analysis were used to obtain information about the molecular structure of the transgene insertions.

Figure 5. Transfer Cassettes Plasmids PHP27064 and PHP44951 for FLP recombinase mediated site specific integration and RMCE.

Figure 6. Integration of transfer cassette at the MHP14 SSI target site after FLP recombinase mediated site specific integration.

Table 1 . Sequences 1- 13

SEQ ID NO:	NT/AA	Description
1	NT	TS21 recognition site of soybean ggcactctcg tgtgtgatta aa
2	NT	TS14 recognition site of soybean cagacgtacg caagtagctt tg
3	NT	TS30 recognition site of soybean gagtcccacg caagagcata aa

4	NT	TS5 recognition site of soybean aagacttacg tgtgtactcg tg
5	NT	TS7 recognition site of soybean gacattgtcg tgagaaaaga ga
6	NT	TS4 recognition site of soybean aaatctgtct tgcgaaacgg ca
7	NT	TS22 recognition site of soybean tattctctca taaataaact tt
8	NT	TS24 recognition site of soybean ggaatggaca taagagaact gt
9	NT	FRT1 recombination site gaagttccta ttctctagaa agtataggaa cttc
10	NT	FRT5 recombination site agttcctatt cttcaaaagg tataggaact
11	NT	FRT6 recombination site agttcctatt cttcaaaaag tataggaact
12	NT	FRT12 recombination site agttcctata ctctatgtag aataggaact
13	NT	FRT87 recombination site agttcctata ctttctggag aataggaact

NT= nucleotide seqeunce

SEQ ID NO: 14 PHP36070

SEQ ID NO: 15 is the nucleotide sequence of the LIG3-4 meganuclease

SEQ ID NO: 16 is the nucleotide sequence of the LIG3-4 recognition site.

SEQ ID NO: 17 is the homologous DNA region (HR1) flanking the LIG3-4 recognition site.

SEQ ID NO: 18 is the homologous DNA region (HR2) flanking the LIG3-4 recognition site.

SEQ ID NO: 19 PHP44779

SEQ ID NO: 20 is the nucleotide sequence of the MHP14 recognition site in the maize genome.

SEQ ID NO:21 is the plant optimized nucleotide sequence of the MHP14+ comprising a nuclear localization signal and lacking an intron.

SEQ ID NO: 22 is the HR1 of the MHP14 target site.

SEQ ID NO: 23 is the HR2 of the MHP14 target site.

SEQ ID NO: 24 is the Transfer Cassette plasmid PHP27064 .

SEQ ID NO: 25 is the Transfer Cassette plasmid PHP44951.

SEQ ID NO: 26 is the plant optimized nucleotide sequence of the TS14 meganuclease.

SEQ ID NO: 27 RTW347

SEQ ID NO: 28 RTW365

SEQ ID NO: 29 WOL192 primer sequence

SEQ ID NO: 30 WOL311 primer sequence

SEQ ID NO: 31 PCR product with WOL192 and WOL311

SEQ ID NO: 32 WOL312 primer sequence

SEQ ID NO: 33 WOL193 primer sequence

SEQ ID NO: 34 PCR product with WOL312 and WOL193

SEQ ID NO: 35 is the homologous DNA region (HR1) flanking the TS14 recognition site.

SEQ ID NO: 36 is the homologous DNA region (HR2) flanking theTS14 recognition site.

DETAILED DESCRIPTION OF THE INVENTION

The present inventions now will be described more fully hereinafter with reference to the accompanying drawings, in which some, but not all embodiments of the invention are shown. Indeed, these inventions may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will satisfy applicable legal requirements. Like numbers refer to like elements throughout.

Many modifications and other embodiments of the inventions set forth herein will come to mind to one skilled in the art to which this invention pertains having the benefit of the teachings presented in the foregoing descriptions and the associated drawings. Therefore, it is to be understood that the invention is not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the appended claims. Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes of limitation.

I. Overview

Compositions and methods are provided for the targeted integration of a polynucleotide sequence of interest into the genome of a plant or plant cell. The methods and compositions employ endonucleases, recognition sites for these endonucleases in combination with site-specific recombination sites/recombinases

to provide an effective system for establishing target sites within the genome of a plant, plant cell or seed. Once such target sites are established, a variety of methods can be employed to further modify the target sites such that they contain a variety of polynucleotides of interest.

II. Target Sites Integrated at a Recognition Site

Methods and compositions are provided herein which establish and use plants, plant cells and seeds having stably incorporated into their genome a target site for site-specific integration where the target site is integrated into a recognition site for an endonuclease. As used herein, a target site is "integrated" into a recognition site when an endonuclease induces a double strand break at the recognition site and a homologous recombination event thereby inserts the target site with the boundaries of the original recognition site. It is recognized that the position within a given recognition site in which the target site integrates will vary depending on where the double strand break is induced by the endonuclease. The sequence of the recognition site need not immediately flank the boundaries of the target site. For example, sequences 5' and 3' to the target site found on the donor DNA may also be integrated into the recognition site.

A. Recognition Sites for Endonucleases

As used herein, the term "recognition site for an endonuclease" refers to a DNA sequence at which a double-strand break is induced in the plant cell genome by an endonuclease. The recognition site can be an endogenous site in the plant genome, or alternatively, the recognition site can be heterologous to the plant and thereby not be naturally occurring in the genome, or the recognition site can be found in a heterologous genomic location compared to where it occurs in nature. As used herein, the term "endogenous recognition site" refers to an endonuclease recognition site that is endogenous or native to the genome of a plant and is located at the endogenous or native position of that recognition site in the genome of the plant.

The length of the recognition site can vary, and includes, for example, recognition sites 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 further 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.

In one embodiment, the recognition site of the endonuclease comprises the TS21 (SEQ ID NO: 1), TS14 (SEQ ID NO: 2), TS30 (SEQ ID NO: 3), TS5 (SEQ ID NO: 4), TS7 (SEQ ID NO: 5), TS4 (SEQ ID NO: 6), TS22 (SEQ ID NO: 7), and/or TS24 (SEQ ID NO: 8) recognition sites of soybean which are disclosed in U.S. Provisional Application No. 61/466,602, filed on March 23, 2011, herein incorporated by reference in its entirety. In another embodiment, the recognition site can comprise the LIG3-4 recognition site (SEQ ID NO:16) and the MHP14 recognition site (SEQ ID NO:20)

Any endonuclease that induces a double-strand break into a desired recognition site can be used in the methods and compositions disclosed herein. A naturally-occurring or native endonuclease can be employed so long as the endonuclease induces a double-strand break in a desired recognition site. Alternatively, a modified or engineered endonuclease can be employed. An "engineered endonuclease" refers to an endonuclease that is engineered (modified or derived) from its native form to specifically recognize and induce a double-strand break in the desired recognition site. Thus, an engineered endonuclease can be derived from a native, naturally-occurring endonuclease or it could be artificially created or synthesized. The modification of the endonuclease can be as little as one nucleotide. In some embodiments, the engineered endonuclease induces a double-strand break in a recognition site, wherein the recognition site was not a sequence that would have been recognized by a native (non-engineered or non-modified) endonuclease. Producing a double-strand break in a recognition site or other DNA can be referred to herein as "cutting" or "cleaving" the recognition site or other DNA.

Active variants and fragments of the recognition sites (i.e. SEQ ID NOS: 1-8, 16 and 20) are also provided. Such active variants can comprise at least 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the given recognition site, wherein the active variants retain biological activity and hence are capable of being recognized and cleaved by an endonuclease. Assays to measure the double-strand break of a recognition site by an endonuclease are known in the art and generally measure the ability of an endonuclease to cut the recognition site.

Endonucleases are enzymes that cleave the phosphodiester bond within a polynucleotide chain, and include restriction endonucleases that cleave DNA at specific sites without damaging the bases. Restriction endonucleases include Type I, Type II, Type III, and Type IV endonucleases, which further include subtypes. In the Type I and Type III systems, both the methylase and restriction activities are contained in a single complex.

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 site, 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 open reading frames, 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 a naturally occurring variant, and/or engineered derivative 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.

Any meganuclease can be used herein, 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 active variants or fragments thereof. In a specific embodiment, the engineered endonuclease is derived from I-Cre-I having the sequence set forth in SEQ ID NO: 15, 21 or 26 or an active variant or fragment thereof.

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.

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.

Active variants and fragments of endonucleases (i.e. an engineered endonuclease) are also provided. Such active variants can comprise at least 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the native endonuclease, wherein the active variants retain the ability to cut at a desired recognition site and hence retain double-strand-

break-inducing activity. For example, any of the engineered endonucleases described herein can be modified from a native endonuclease sequence and designed to recognize and induce a double strand break at a recognition site that was not recognized by the native endonuclease. Thus in some embodiments, the engineered endonuclease has a specificity to induce a double-strand break at a recognition site that is different from the corresponding native endonuclease recognition site. Assays for double-strand-break-inducing activity are known and generally measure the overall activity and specificity of the endonuclease on DNA substrates containing the recognition site.

The endonuclease may be introduced by any means known in the art. For example, a plant having the recognition site in its genome is provided. The endonuclease may be transiently expressed or the polypeptide itself can be directly provided to the cell. Alternatively, a nucleotide sequence capable of expressing the endonuclease may be stably integrated into the genome of the plant. In the presence of the corresponding recognition site and the endonuclease, the donor DNA is inserted into the transformed plant's genome. Alternatively, the components of the system may be brought together by sexually crossing transformed plants. Thus a sequence encoding the endonuclease and/or target site can be sexually crossed to one another to allow each component of the system to be present in a single plant. The endonuclease may be under the control of a constitutive or inducible promoter. Such promoters of interest are discussed in further detail elsewhere herein.

B. Integration of a Target Site into the Recognition Site by Homologous Recombination

As outlined above, plants, plant cells and seeds having a target site integrated at a recognition site are provided. Various methods can be used to integrate the target site at the recognition site. Such methods employ homologous recombination to provide integration of the target site at the endonuclease recognition site. In the methods provided, the target site is provided to the plant cell in a donor DNA construct. As used herein, "donor DNA" is a DNA construct that comprises a target site for site-specific integration. The donor DNA construct further comprises a first and a second region of homology that flank the target site

sequence. The first and second regions of homology of the donor DNA share homology to a first and a second genomic region, respectively, present in or flanking the recognition site of the plant genome. By "homology" is meant DNA sequences that are similar. For example, a "region of homology to a genomic region" that is found on the donor DNA is a region of DNA that has a similar sequence to a given "genomic region" in the plant genome. A region of homology can be of any length that is sufficient to promote homologous recombination at the cleaved recognition site. For example, the region of homology can comprise at least 5-10, 5-15, 5-20, 5-25, 5-30, 5-35, 5-40, 5-45, 5- 50, 5-55, 5-60, 5-65, 5- 70, 5-75, 5-80, 5-85, 5-90, 5-95, 5-100, 5-200, 5-300, 5-400, 5-500, 5-600, 5-700, 5-800, 5-900, 5-1000, 5-1100, 5-1200, 5-1300, 5-1400, 5-1500, 5-1600, 5-1700, 5-1800, 5-1900, 5-2000, 5-2100, 5-2200, 5-2300, 5-2400, 5-2500, 5-2600, 5-2700, 5-2800. 5-2900, 5-3000, 5-3100 or more bases in length such that the region of homology has sufficient homology to undergo homologous recombination with the corresponding genomic region. "Sufficient homology" indicates that two polynucleotide sequences have sufficient structural similarity to act as substrates for a homologous recombination reaction.

As used herein, a "genomic region" is a segment of a chromosome in the genome of a plant cell that is present on either side of the recognition site or, alternatively, also comprises a portion of the recognition site. The genomic region can comprise at least 5-10, 5-15, 5-20, 5-25, 5-30, 5-35, 5-40, 5-45, 5- 50, 5-55, 5-60, 5-65, 5- 70, 5-75, 5-80, 5-85, 5-90, 5-95, 5-100, 5-200, 5-300, 5-400, 5-500, 5-600, 5-700, 5-800, 5-900, 5-1000, 5-1100, 5-1200, 5-1300, 5-1400, 5-1500, 5-1600, 5-1700, 5-1800, 5-1900, 5-2000, 5-2100, 5-2200, 5-2300, 5-2400, 5-2500, 5-2600, 5-2700, 5-2800. 5-2900, 5-3000, 5-3100 or more bases such that the genomic region has sufficient homology to undergo homologous recombination with the corresponding region of homology.

The structural similarity between a given genomic region and the corresponding region of homology found on the donor DNA can be any degree of sequence identity that allows for homologous recombination to occur. For example, the amount of homology or sequence identity shared by the "region of homology" of the donor DNA and the "genomic region" of the plant genome can be at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%,

89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity, such that the sequences undergo homologous recombination

The region of homology on the donor DNA can have homology to any sequence flanking the recognition site. While in some embodiments the regions of homology share significant sequence homology to the genomic sequence immediately flanking the recognitions site, it is recognized that the regions of homology can be designed to have sufficient homology to regions that may be further 5' or 3' to the recognition site. In still other embodiments, the regions of homology can also have homology with a fragment of the recognition site along with downstream genomic regions. In one embodiment, the first region of homology further comprises a first fragment of the recognition site and the second region of homology comprises a second fragment of the recognition site, wherein the first and second fragments are dissimilar. In a further embodiment the first region of homology comprises the first 13 bases of the recognition site and the second region of homology comprises the last 9 bases of the recognition site.

As used herein, “homologous recombination” refers to the exchange of DNA fragments between two DNA molecules at the sites of homology. The frequency of homologous recombination is influenced by a number of factors. Different organisms vary with respect to the amount of homologous recombination and the relative proportion of homologous to non-homologous recombination. Generally, the length of the region of homology affects the frequency of homologous recombination events, the longer the region of homology, the greater the frequency. The length of the homology region needed to observe homologous recombination is also species-variable. In many cases, at least 5 kb of homology has been utilized, but homologous recombination has been observed with as little as 25-50 bp of homology. See, for example, Singer *et al.*, (1982) *Cell* 31:25-33; Shen and Huang, (1986) *Genetics* 112:441-57; Watt *et al.*, (1985) *Proc. Natl. Acad. Sci. USA* 82:4768-72, Sugawara and Haber, (1992) *Mol Cell Biol* 12:563-75, Rubnitz and Subramani, (1984) *Mol Cell Biol* 4:2253-8; Ayares *et al.*, (1986) *Proc. Natl. Acad. Sci. USA* 83:5199-203; Liskay *et al.*, (1987) *Genetics* 115:161-7.

Alteration of the genome of a plant cell, for example, through homologous recombination (HR), is a powerful tool for genetic engineering. 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.

Once a double-strand break is induced in the DNA, the cells DNA repair mechanism is activated to repair the break. Error-prone DNA repair mechanisms can produce mutations at double-strand break sites. The most common repair mechanism to bring the broken ends together is the nonhomologous end-joining (NHEJ) pathway (Bleuyard *et al.*, (2006) *DNA Repair* 5:1-12). The structural integrity of chromosomes is typically preserved by the repair, but deletions, insertions, or other rearrangements are possible (Siebert and Puchta, (2002) *Plant Cell* 14:1121-31; Pacher *et al.*, (2007) *Genetics* 175:21-9).

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

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

Provided herein, the methods comprise contacting a plant cell with the donor DNA and the endonuclease. Once a double-strand break is introduced in the recognition site by the endonuclease, the first and second regions of homology of the donor DNA can undergo homologous recombination with their corresponding genomic regions of homology resulting in exchange of DNA between the donor and the genome. As such, the provided method results in the integration of the target site of the donor DNA into the double-strand break in the recognition site in the plant genome.

The donor DNA may be introduced by any means known in the art. For example, a plant having a recognition site is provided. The donor DNA may be provided by any transformation method known in the art including, for example, Agrobacterium-mediated transformation or biolistic particle bombardment. The donor DNA may be present transiently in the cell or it could be introduced via a viral replicon. In the presence of the endonuclease and the recognition site, the donor DNA is inserted into the transformed plant's genome.

Further provided are methods for identifying at least one plant cell comprising in its genome the target site integrated at the recognition site. A variety of methods are available for identifying those plant cells with insertion into the genome at or near to the recognition site 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. See, for example, US Patent Application 12/147,834, herein incorporated by reference in its entirety.

The method also comprises recovering a fertile plant from the plant cell comprising a target site integrated into its genome. As used herein, a "fertile plant" is a plant that is capable of producing a progeny plant. The fertile plant can comprise any of the various target sites as described elsewhere herein integrated in its genome into the recognition site.

III. Methods for Integrating a Polynucleotide of Interest Into the Target Site

As outlined above, various methods and compositions can be employed to obtain a plant having a target site inserted in a recognition site for an endonuclease.

Once such plants and plant cells are generated, a variety of methods can be used to manipulate the sequence within the target site. Such methods employ various components of site-specific recombination systems.

A. The Target Site and Components Thereof

As discussed herein, the various methods and compositions employ a target site. As described in the previous section, the target site is provided in a donor DNA which undergoes homologous recombination with the genomic DNA at the cleaved recognition site resulting in integration of the target site into the genome of the plant cell.

The target site can comprise various components. As used herein, by "target site" is intended a polynucleotide comprising a nucleotide sequence flanked by at least two recombination sites. In some embodiments, the recombination sites of the target site are dissimilar and non-recombinogenic with respect to one another. One or more intervening sequences may be present between the recombination sites of the target site. Intervening sequences of particular interest would include linkers, adapters, selectable markers, polynucleotides of interest, promoters and/or other sites that aid in vector construction or analysis. In addition, the recombination sites of the target site can be located in various positions, including, for example, within intronic sequences, coding sequences, or untranslated regions.

The target site can comprise 1, 2, 3, 4, 5, 6 or more recombination sites. In one embodiment, the target site comprises a first recombination site and a second recombination site wherein the first and the second recombination site are dissimilar and non-recombinogenic to each other. In a further embodiment, the target site comprises a third recombination site between the first recombination site and the second recombination site. In such embodiments, the first, second and third recombination sites may be dissimilar and non-recombinogenic with respect to one another. Such first, second and third recombination sites are able to recombine with their corresponding or identical recombination site when provided with the appropriate recombinase. The various recombination sites and recombinases encompassed by the methods and compositions are described in detail elsewhere herein.

The recombination sites employed in the methods and compositions provided herein can be "corresponding" sites or "dissimilar" sites. By "corresponding recombination sites" or a "set of corresponding recombination sites" is intended that the recombination sites have the same or corresponding nucleotide sequence. A set of corresponding recombination sites, in the presence of the appropriate recombinase, will efficiently recombine with one another (i.e., the corresponding recombination sites are recombinogenic).

In other embodiments, the recombination sites are dissimilar. By "dissimilar recombination sites" or a "set of dissimilar recombination sites" is intended that the recombination sites are distinct (i.e., have at least one nucleotide difference).

The recombination sites within "a set of dissimilar recombination sites" can be either recombinogenic or non-recombinogenic with respect to one other. By "recombinogenic" is intended that the set of recombination sites are capable of recombining with one another. Thus, suitable sets of "recombinogenic" recombination sites for use in the methods and compositions provided herein include those sites where the relative excision efficiency of recombination between the recombinogenic sites is above the detectable limit under standard conditions in an excision assay, typically, greater than 2%, 5%, 10%, 20%, 50%, 100%, or greater.

By "non-recombinogenic" is intended the set of recombination sites, in the presence of the appropriate recombinase, will not recombine with one another or recombination between the sites is minimal. Thus, suitable "non-recombinogenic" recombination sites for use in the methods and compositions provided herein include those sites that recombine (or excise) with one another at a frequency lower than the detectable limit under standard conditions in an excision assay, typically, lower than 2%, 1.5%, 1%, 0.75%, 0.5%, 0.25%, 0.1%, 0.075, 0.005%, 0.001%.

Each recombination site within the "set of non-recombinogenic sites" is biologically active and therefore can recombine with an identical site. Accordingly, it is recognized that any suitable non-recombinogenic recombination sites may be utilized, including a FRT site or an active variant thereof, a LOX site or active variant thereof, any combination thereof, or any other combination of non-recombinogenic recombination sites known in the art. FRT sites that can be employed in the

methods and compositions disclosed herein can be found, for example, in US Publication No. 2011-0047655, herein incorporated by reference.

In a specific embodiment, at least one of the first, the second and the third recombination site comprises FRT1 (SEQ ID NO: 9), FRT5 (SEQ ID NO: 10), FRT6 (SEQ ID NO: 11), FRT12 (SEQ ID NO: 12) or FRT87 (SEQ ID NO: 13). In a specific embodiment, the first recombination site is FRT1, the second recombination site is FRT12 and the third recombination site is FRT87.

B. Transfer Cassettes and Components Thereof

The methods also comprise introducing into the plant cell comprising the integrated target site a transfer cassette. The transfer cassette comprises various components for the incorporation of polynucleotides of interest into the plant genome. As defined herein, the "transfer cassette" comprises at least a first recombination site, a polynucleotide of interest, and a second recombination site, wherein the first and second recombination sites are dissimilar and non-recombinogenic and correspond to the recombination sites in the target site. The transfer cassette is also immediately flanked by the recombination sites. It is recognized that any combination of restriction sites can be employed in the transfer cassettes to provide a polynucleotide of interest.

In one embodiment, the transfer cassette comprises the first recombination site, a first polynucleotide of interest, and the second recombination site. In such methods, the first and second recombination sites of the transfer cassette are recombinogenic (i.e. identical or corresponding) with the first and second recombination sites of the target site, respectively.

In another embodiment of the methods, the transfer cassette comprises the second recombination site, a second polynucleotide of interest, and the third recombination site. In such methods, the second and third recombination sites of the transfer cassette are recombinogenic (i.e. identical or corresponding) with the second and third recombination sites of the target site in the plant genome, respectively.

In yet, another embodiment, the transfer cassette comprises the first recombination site, a third polynucleotide of interest, and the third recombination site. In such cases, the second and third recombination sites of the transfer

cassette are recombinogenic (i.e. identical or corresponding) with the first and third recombination sites of the target site in the plant genome, respectively.

The recombination sites of the transfer cassette may be directly contiguous with the polynucleotide of interest or there may be one or more intervening sequences present between one or both ends of the polynucleotide of interest and the recombination sites. Intervening sequences of particular interest would include linkers, adapters, additional polynucleotides of interest, markers, promoters and/or other sites that aid in vector construction or analysis. It is further recognized that the recombination sites can be contained within the polynucleotide of interest (i.e., such as within introns, coding sequence, or untranslated regions).

In a specific embodiment, the transfer cassette further comprises at least one coding region operably linked to a promoter that drives expression in the plant cell. As discussed elsewhere herein, a recombinase is provided that recognizes and implements recombination at the recombination sites of the target site and the transfer cassette. The recombinase can be provided by any means known in the art and is described in detail elsewhere herein. In a specific embodiment, the coding region of the transfer cassette encodes a recombinase that facilitates recombination between the first and the second recombination sites of the transfer cassette and the target site, the second and the third recombination sites of the transfer cassette and the target site, or the first and the third recombination sites of the transfer cassette and the target site.

Further, the methods provide selecting at least one plant cell comprising integration of the transfer cassette at the target site. Methods for selecting plant cells with integration at the target site, such as selecting for cells expressing a selectable marker, are known in the art and are described elsewhere herein. As such, the methods further comprise recovering a fertile plant from the plant cell comprising in its genome the transfer cassette at the target site.

i. Polynucleotides of Interest

Any polynucleotide of interest (i.e., the "polypeptide of interest") may be provided to the plant cells in the transfer cassettes or target sites of the methods disclosed herein. It is recognized that any polynucleotide of interest can be provided, integrated into the plant genome at the target site by site-specific

integration, and expressed in a plant. The methods disclosed herein, provide for at least 1, 2, 3, 4, 5, 6 or more polynucleotides of interest to be integrated into a specific site in the plant genome.

Various changes in phenotype are of interest, including modifying the fatty acid composition in a plant, altering the amino acid content of a plant, altering a plant's pathogen defense mechanism, and the like. These results can be achieved by providing expression of heterologous products (i.e. polynucleotides of interest) or increased expression of endogenous products in plants. Alternatively, the results can be achieved by providing for a reduction of expression of one or more endogenous products, particularly enzymes or cofactors in the plant. These changes result in a change in phenotype of the transformed plant.

In one embodiment, at least one of the first, the second, and the third polynucleotides of interest comprises a nucleotide sequence for gene silencing, a nucleotide sequence encoding a phenotypic marker, or a nucleotide sequence encoding a protein providing an agronomic advantage.

Polynucleotides of interest are reflective of the commercial markets and interests of those involved in the development of the crop. Crops and markets of interest change, and as developing nations open up world markets, new crops and technologies will emerge also. In addition, as our understanding of agronomic traits and characteristics such as yield and heterosis increase, the choice of genes for transformation will change accordingly. Polynucleotides/polypeptides of interest include, but are not limited to, herbicide-tolerance coding sequences, insecticidal coding sequences, nematicidal coding sequences, antimicrobial coding sequences, antifungal coding sequences, antiviral coding sequences, abiotic and biotic stress tolerance coding sequences, or sequences modifying plant traits such as yield, grain quality, nutrient content, starch quality and quantity, nitrogen fixation and/or utilization, and oil content and/or composition. More specific polynucleotides of interest include, but are not limited to, genes that improve crop yield, polypeptides that improve desirability of crops, genes encoding proteins conferring resistance to abiotic stress, such as drought, nitrogen, temperature, salinity, toxic metals or trace elements, or those conferring resistance to toxins such as pesticides and herbicides, or to biotic stress, such as attacks by fungi, viruses, bacteria, insects, and nematodes, and development of diseases associated with these organisms.

An "herbicide resistance protein" or a protein resulting from expression of an "herbicide resistance-encoding nucleic acid molecule" includes proteins that confer upon a cell the ability to tolerate a higher concentration of an herbicide than cells that do not express the protein, or to tolerate a certain concentration of an herbicide for a longer period of time than cells that do not express the protein. Herbicide resistance traits may be introduced into plants by genes coding for resistance to herbicides that act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides, genes coding for resistance to herbicides that act to inhibit the action of glutamine synthase, such as phosphinothricin or basta (e.g., the bar gene), glyphosate (e.g., the EPSP synthase gene and the GAT gene), HPPD inhibitors (e.g, the HPPD gene) or other such genes known in the art. See, for example, US Patent Nos. 7,626,077, 5,310,667, 5,866,775, 6,225,114, 6,248,876, 7,169,970, 6,867,293, and US Provisional Application No. 61/401,456, each of which is herein incorporated by reference.

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.

Commercial traits can also be encoded on a polynucleotide of interest that could increase for example, starch for ethanol production, or provide expression of proteins. Another important commercial use of transformed plants is the production of polymers and bioplastics such as described in U.S. 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).

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.

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

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

These polynucleotide sequences of interest may encode proteins involved in providing disease or pest resistance. By "disease resistance" or "pest resistance" is intended that the plants avoid the harmful symptoms that are the outcome of the plant-pathogen interactions. Pest resistance genes may encode resistance to pests that have great yield drag such as rootworm, cutworm, European Corn Borer, and the like. Disease resistance and insect resistance genes such as lysozymes or cecropins for antibacterial protection, or proteins such as defensins, glucanases or chitinases for antifungal protection, or *Bacillus thuringiensis* endotoxins, protease inhibitors, collagenases, lectins, or glycosidases for controlling nematodes or insects are all examples of useful gene products. Genes encoding disease resistance traits include detoxification genes, such as against 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.

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

In addition, the polynucleotide of interest may also be used in the sense orientation to suppress the expression of endogenous genes in plants. Methods for suppressing gene expression in plants using polynucleotides in the sense orientation are known in the art. The methods generally involve transforming plants with a DNA construct comprising a promoter that drives expression in a plant operably linked to at least a portion of a nucleotide sequence that corresponds to the transcript of the endogenous gene. Typically, such a nucleotide sequence has substantial sequence identity to the sequence of the transcript of the endogenous

gene, generally greater than about 65% sequence identity, about 85% sequence identity, or greater than about 95% sequence identity. See, U.S. Patent Nos. 5,283,184 and 5,034,323; herein incorporated by reference.

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

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

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.

Active variants or fragments of polynucleotides/polypeptides of interest are also provided. Such active variants can comprise at least 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the native polynucleotide/polypeptide of interest, wherein the active variants retain the biological activity of the native polynucleotide/polypeptide.

C. Site-Specific Recombination System

A site-specific recombination system can be employed in a variety of ways to manipulate the target site that has been integrated at the recognition site. The site-specific recombination system employs various components which are described in detail below and in U.S. Patent Nos. 6187994, 6262341, 6331661 and 6300545, each of which is herein incorporated by reference.

Various recombination sites can be employed in the methods and compositions provided herein (i.e. in the various target sites or transfer cassettes disclosed herein). By "recombination site" is intended a naturally occurring recombination site and active variants thereof. Many recombination systems are known in the art and one of skill will recognize the appropriate recombination site to be used with the recombination system of interest. As discussed in greater detail elsewhere herein, various combinations of recombination sites can be employed including, sets of dissimilar sites and corresponding recombination sites and/or

dissimilar and non-recombinogenic sites can be used in the various methods provided herein. Accordingly, any suitable recombination site or set of recombination sites may be utilized herein, including a FRT site, a biologically active variant of a FRT site (i.e. a mutant FRT site), a LOX site, a biologically active variant of a LOX site (i.e. a mutant LOX site), any combination thereof, or any other combination of recombination sites known in the art. Examples of FRT sites include, for example, the wild type FRT site (FRT1, SEQ ID NO: 9), and various mutant FRT sites, including but not limited to, FRT5 (SEQ ID NO: 10), FRT6 (SEQ ID NO: 11), FRT12 (SEQ ID NO: 12) and FRT87 (SEQ ID NO: 13). See, for example, U.S. Patent No. 6,187,994.

Recombination sites from the Cre/Lox site-specific recombination system can also be used. Such recombination sites include, for example, wild type LOX sites and mutant LOX sites. An analysis of the recombination activity of mutant LOX sites is presented in Lee *et al.* (1998) *Gene* 216:55-65, herein incorporated by reference. Also, see for example, Schlake and Bode (1994) *Biochemistry* 33:12746-12751; Huang *et al.* (1991) *Nucleic Acids Research* 19:443-448; Sadowski (1995) In *Progress in Nucleic Acid Research and Molecular Biology* Vol. 51, pp. 53-91; Cox (1989) In *Mobile DNA*, Berg and Howe (eds) American Society of Microbiology, Washington D.C., pp. 116-670; Dixon *et al.* (1995) *Mol. Microbiol.* 18:449-458; Umlauf and Cox (1988) *EMBO* 7:1845-1852; Buchholz *et al.* (1996) *Nucleic Acids Research* 24:3118-3119; Kilby *et al.* (1993) *Trends Genet.* 9:413-421; Rossant and Geagy (1995) *Nat. Med.* 1: 592-594; Albert *et al.* (1995) *The Plant J.* 7:649-659; Bayley *et al.* (1992) *Plant Mol. Biol.* 18:353-361; Odell *et al.* (1990) *Mol. Gen. Genet.* 223:369-378; Dale and Ow (1991) *Proc. Natl. Acad. Sci. USA* 88:10558-10562; Qui *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:1706-1710; Stuurman *et al.* (1996) *Plant Mol. Biol.* 32:901-913; Dale *et al.* (1990) *Gene* 91:79-85; Albert *et al.* (1995) *The Plant J.* 7:649-659 and WO 01/00158; all of which are herein incorporated by reference.

Active variants and fragments of recombination sites (i.e SEQ ID NOS: 9-13) are also encompassed by the compositions and methods provided herein. Fragments of a recombination site retain the biological activity of the recombination site and hence facilitate a recombination event in the presence of the appropriate recombinase. Thus, fragments of a recombination site may range from at least

about 5, 10, 15, 20, 25, 30, 35, 40 nucleotides, and up to the full-length of a recombination site. Active variants can comprise at least 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the native recombination site, wherein the active variants retain biological activity and hence facilitate a recombination event in the presence of the appropriate recombinase. Assays to measure the biological activity of recombination sites are known in the art. See, for example, Senecoll *et al.* (1988) *J. Mol. Biol.* 201:406-421; Voziyanov *et al.* (2002) *Nucleic Acid Research* 30:7, U.S. Patent No. 6,187,994, WO/01/00158, and Albert *et al.* (1995) *The Plant Journal* 7:649-659.

Recombinases are also employed in the methods and compositions provided herein. By "recombinase" is intended a native polypeptide that catalyzes site-specific recombination between compatible recombination sites. For reviews of site-specific recombinases, see Sauer (1994) *Current Opinion in Biotechnology* 5:521-527; and Sadowski (1993) *FASEB* 7:760-767; the contents of which are incorporated herein by reference. The recombinase used in the methods can be a naturally occurring recombinase or a biologically active fragment or variant of the recombinase. Recombinases useful in the methods and compositions include recombinases from the Integrase and Resolvase families, biologically active variants and fragments thereof, and any other naturally occurring or recombinantly produced enzyme or variant thereof that catalyzes conservative site-specific recombination between specified DNA recombination sites.

The Integrase family of recombinases has over one hundred members and includes, for example, FLP, Cre, Int, and R. For other members of the Integrase family, see for example, Esposito *et al.* (1997) *Nucleic Acid Research* 25:3605-3614 and Abremski *et al.* (1992) *Protein Engineering 5*:87-91, both of which are herein incorporated by reference. Other recombination systems include, for example, the streptomycete bacteriophage phi C31 (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.* 237:334-342); and a retroviral integrase-based integration system (Tanaka *et al.* (1998) *Gene* 17:67-76). In other embodiments, the recombinase is one that does not require cofactors or a supercoiled substrate. Such recombinases include Cre , FLP , or active variants or fragments thereof (SEQ ID NOS: 15, 21, 26).

The FLP recombinase is a protein that catalyzes a site-specific reaction that is involved in amplifying the copy number of the two-micron plasmid of *S. cerevisiae* during DNA replication. As used herein, FLP recombinase refers to a recombinase that catalyzes site-specific recombination between two FRT sites. The FLP protein has been cloned and expressed. See, for example, Cox (1993) *Proc. Natl. Acad. Sci. U.S.A. 80*:4223-4227. The FLP recombinase for use in the methods and with the compositions may be derived from the genus *Saccharomyces*. One can also synthesize a polynucleotide comprising the recombinase using plant-preferred codons for optimal expression in a plant of interest. A recombinant FLP enzyme encoded by a nucleotide sequence comprising maize preferred codons (FLPm) that catalyzes site-specific recombination events is known. See, for example, U.S. Patent 5,929,301, herein incorporated by reference. Additional functional variants and fragments of FLP are known. See, for example, Buchholz *et al.* (1998) *Nat. Biotechnol. 16*:617-618, Hartung *et al.* (1998) *J. Biol. Chem. 273*:22884-22891, Saxena *et al.* (1997) *Biochim Biophys Acta 1340*(2):187-204, and Hartley *et al.* (1980) *Nature 286*:860-864, all of which are herein incorporated by reference.

The bacteriophage recombinase Cre catalyzes site-specific recombination between two *lox* sites. The Cre recombinase is known in the art. See, for example, Guo *et al.* (1997) *Nature* 389:40-46; Abremski *et al.* (1984) *J. Biol. Chem.* 259:1509-1514; Chen *et al.* (1996) *Somat. Cell Mol. Genet.* 22:477-488; Shaikh *et al.* (1977) *J. Biol. Chem.* 272:5695-5702; and, Buchholz *et al.* (1998) *Nat. Biotechnol. 16*:617-618, all of which are herein incorporated by reference. The Cre polynucleotide sequences may also be synthesized using plant-preferred codons. Such sequences (moCre) are described in WO 99/25840, herein incorporated by reference and set forth in SEQ ID NO: 21.

It is further recognized that a chimeric recombinase can be used in the methods. By "chimeric recombinase" is intended a recombinant fusion protein which is capable of catalyzing site-specific recombination between recombination sites that originate from different recombination systems. That is, if a set of functional recombination sites, characterized as being dissimilar and non-recombinogenic with respect to one another, is utilized in the methods and compositions and comprises a FRT site and a LoxP site, a chimeric FLP/Cre recombinase or active variant or fragment thereof will be needed or, alternatively,

both recombinases may be separately provided. Methods for the production and use of such chimeric recombinases or active variants or fragments thereof are described in WO 99/25840, herein incorporated by reference.

By utilizing various combinations of recombination sites in the target sites and the transfer cassettes provided herein, the methods provide a mechanism for the site-specific integration of polynucleotides of interest into a specific site in the plant genome. The methods also allow for the subsequent insertion of additional polynucleotides of interest into the specific genomic site.

In one embodiment, providing the recombinase comprises integrating into the genome of the plant cell a nucleotide sequence encoding the recombinase. In a specific embodiment, the recombinase is FLP. In yet another embodiment, the FLP recombinase is synthesized using maize-preferred codons.

As used herein, by "providing" is intended any method that allows for an amino acid sequence and/or a polynucleotide to be brought together with the recited components. A variety of methods are known in the art for the introduction of nucleotide sequence into a plant. Any means can be used to bring together the various components of the recombination system (i.e., the target site, transfer cassette, and the appropriate recombinase), including, for example, transformation and sexual crossing. See, also, WO99/25884 herein incorporated by reference. In addition, as discussed elsewhere herein, the recombinase may also be provided by the introduction of the polypeptide or mRNA into the cell.

Active variants and fragments of recombinases (i.e FLP or Cre) are also encompassed by the compositions and methods provided herein. Such active variants can comprise at least 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the native recombinase, wherein the active variants retain biological activity and hence implement a recombination event. Assays for recombinase activity are known and generally measure the overall activity of the enzyme on DNA substrates containing recombination sites. For example, to assay for FLP activity, inversion of a DNA sequence in a circular plasmid containing two inverted FRT sites can be detected as a change in position of restriction enzyme sites. This assay is described in Vetter *et al.* (1983) *PNAS* 80:7284. Alternatively, excision of DNA from a linear molecule or intermolecular recombination frequency induced by the enzyme may be assayed, as

described, for example, in Babineau *et al.* (1985) *Journal of Biological Chemistry* 260:12313; Meyer-Leon *et al.* (1987) *Nucleic Acid Res* 15:6469; and Gronostajski *et al.* (1985) *Journal of Biological Chemistry* 260:12328. Alternatively, recombinase activity may also be assayed by excision of a sequence flanked by recombinogenic FRT sites that upon removal will activate an assayable marker gene.

D. Methods of Manipulating the Target Site Integrated Into the Genome

As discussed above, various methods can be used to insert polynucleotides of interest into the target site in a plant or plant cell. Non-limiting examples of various DNA constructs, target sites, and transfer cassettes that can be used to insert a polynucleotide of interest into a plant or plant cell are set forth in Table 2. In each of the examples presented in Table 2, once the target site has integrated into the recognition site or once the transfer cassette has integrated into the target site, the appropriate selective agent can be employed to identify the plant cell having the desired DNA construct.

Once a target site has been established within the genome, additional recombination sites may be introduced by incorporating such sites within the nucleotide sequence of the transfer cassette. Thus, once a target site has been established, it is possible to subsequently add or alter sites through recombination. Such methods are described in detail in WO 99/25821, herein incorporated by reference.

In one embodiment, multiple genes or polynucleotides of interest can be stacked at the target site in the genome of the plant. For example, as illustrated in Table 1, scheme D, the target site integrated at the recognition site can comprise the following components: RSF1::P1::R1::S1::T1-P2::NT1::T2-P3::R2-R3::RSF2, where RSF is a fragment of the recognition site, P is a promoter active in a plant, R is a recombination site, S is the selection marker, T is a termination region, and NT is a polynucleotide of interest. The following transfer cassette comprising the following components could be introduced: R2::S2::T3-P4::NT2::T4-R3. The plant with this transfer cassette integrated at the target site, can then be selected for based on the second selection marker. In this manner, multiple sequences can be stacked at predetermined locations in the target site. Various alterations can be

made to the stacking method described above and still achieve the desired outcome of having the polynucleotides of interest stacked in the genome of the plant.

Table 2. Non-Limiting Examples of Various Integrated Target Sites, Transfer Cassettes, and Integrated Transfer Cassettes

	Target site integrated at the recognition site		Transfer cassette		Transfer cassette integrated at the target site
A	RSF1-P1::R1::NT1::T1-R2-RSF2	X	R1::S1::T2-R2	→	RSF1-P1::R1::S1::T2-R2-RSF2
B	RSF1-P1::R1::NT1::T1-R2-RSF2	X	R1::S1::T2-P2::NT2::T3-R2	→	RSF1-P1::R1::S1-T2-P2::NT2::T3::R2-RSF2
C	RSF1-P1::R1::NT1::T1-R2-RSF2	X	R1::S1::T2-P2::NT2::T3-P3::R2-R3	→	RSF1-P1::R1::S1::T2-P2::NT2::T3-P3::R2::R3-RSF2
D	RSF1-P1::R1::S1::T1-P2::NT1::T2-P3::R2-R3-RSF2	X	R2::S2::T3-P4::NT2::T4-R3	→	RSF1-P1::R1::S1::T1-P2::NT1::T2-P3::R2::S2::T3-P4::NT2::T4-R3-RSF2

RSF = recognition site fragment; P = promoter active in a plant; R = recombination site; S = selection marker; T = terminator region; NT = polynucleotide of interest; the symbol :: implies a fusion between adjacent elements and implies that the sequences are put together to generate an inframe fusion that results in a properly expressed and functional gene product.

IV. Methods of Introducing Sequences

As outlined above, methods and compositions provided herein combine an endonuclease integration system with a site-specific recombinase system which allow for improved methods and compositions for the targeted insertion of a sequence of interest in the genome of a plant. Such systems employ a variety of components and for ease of reference, herein the term “site-specific integration system” generically refers to all the components of the endonuclease integration system (i.e. the various endonucleases, recognition sites, target sites, donor DNA or any active variants or fragments thereof provided herein) and the site-specific recombination system (i.e. the various transfer cassettes, site-specific recombination sites, site-specific recombinases, polynucleotides of interest or any active variants or fragments thereof provided herein).

The methods provided herein comprise introducing into a plant cell, plant or seed a polynucleotide or polypeptide construct comprising the various components of the site-specific integration system provided herein.

The methods provided herein do not depend on a particular method for introducing any component of the site-specific integration system into the host cell, only that the polynucleotide gains access to the interior of a least one cell of the host. Methods for introducing polynucleotides into host cells (i.e. plants) are known in the art and include, but are not limited to, stable transformation methods, transient transformation methods, and virus-mediated methods.

By "introducing" is intended presenting to the plant the sequence (polypeptide or polynucleotide) in such a manner that the sequence gains access to the interior of a cell of the plant. The methods provided herein do not depend on a particular method for introducing a sequence into a plant, only that the polynucleotide or polypeptide gains access to the interior of at least one cell of the plant. Methods for introducing sequences into plants are known in the art and include, but are not limited to, stable transformation methods, transient transformation methods, virus-mediated methods, and sexual breeding. Thus, “introduced” in the context of inserting a nucleic acid fragment (e.g., various components of the site-specific integration system provided herein) 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).

In some embodiments, the plant cells, plants and seeds employed in the methods and compositions have a DNA construct stably incorporated into their genome. By "stably incorporated" or "stably introduced" is intended the introduction of a polynucleotide into the plant such that the nucleotide sequence integrates into the genome of the plant and is capable of being inherited by progeny thereof. Any protocol may be used for the stable incorporation of the DNA constructs or the various components of the site-specific integration system employed herein.

Transformation protocols as well as protocols for introducing polypeptides or polynucleotide sequences into plants may vary depending on the type of plant or

plant cell, i.e., monocot or dicot, targeted for transformation. Suitable methods of introducing polypeptides and polynucleotides into plant cells include microinjection (Crossway *et al.* (1986) *Biotechniques* 4:320-334), electroporation (Riggs *et al.* (1986) *Proc. Natl. Acad. Sci. USA* 83:5602-5606, *Agrobacterium*-mediated transformation (U.S. Patent No. 5,563,055 and U.S. Patent No. 5,981,840), direct gene transfer (Paszkowski *et al.* (1984) *EMBO J.* 3:2717-2722), and ballistic particle acceleration (see, for example, U.S. Patent Nos. 4,945,050; U.S. Patent No. 5,879,918; U.S. Patent No. 5,886,244; and, 5,932,782; Tomes *et al.* (1995) in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, ed. Gamborg and Phillips (Springer-Verlag, Berlin); McCabe *et al.* (1988) *Biotechnology* 6:923-926); and Lec1 transformation (WO 00/28058). Also see Weissinger *et al.* (1988) *Ann. Rev. Genet.* 22:421-477; Sanford *et al.* (1987) *Particulate Science and Technology* 5:27-37 (onion); Christou *et al.* (1988) *Plant Physiol.* 87:671-674 (soybean); McCabe *et al.* (1988) *Bio/Technology* 6:923-926 (soybean); Finer and McMullen (1991) *In Vitro Cell Dev. Biol.* 27P:175-182 (soybean); Singh *et al.* (1998) *Theor. Appl. Genet.* 96:319-324 (soybean); Datta *et al.* (1990) *Biotechnology* 8:736-740 (rice); Klein *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:4305-4309 (maize); Klein *et al.* (1988) *Biotechnology* 6:559-563 (maize); U.S. Patent Nos. 5,240,855; 5,322,783; and, 5,324,646; Klein *et al.* (1988) *Plant Physiol.* 91:440-444 (maize); Fromm *et al.* (1990) *Biotechnology* 8:833-839 (maize); Hooykaas-Van Slogteren *et al.* (1984) *Nature (London)* 311:763-764; U.S. Patent No. 5,736,369 (cereals); Bytebier *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:5345-5349 (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 Reports* 9:415-418 and Kaeppler *et al.* (1992) *Theor. Appl. Genet.* 84:560-566 (whisker-mediated transformation); D'Halluin *et al.* (1992) *Plant Cell* 4:1495-1505 (electroporation); Li *et al.* (1993) *Plant Cell Reports* 12:250-255 and Christou and Ford (1995) *Annals of Botany* 75:407-413 (rice); Osjoda *et al.* (1996) *Nature Biotechnology* 14:745-750 (maize via *Agrobacterium tumefaciens*); all of which are herein incorporated by reference.

In other embodiments, any of the polynucleotides employed herein may be introduced into plants by contacting plants with a virus or viral nucleic acids. Generally, such methods involve incorporating a desired polynucleotide within a

viral DNA or RNA molecule. It is recognized that a sequence employed in the methods or compositions provided herein may be initially synthesized as part of a viral polyprotein, which later may be processed by proteolysis *in vivo* or *in vitro* to produce the desired recombinant protein. Further, it is recognized that promoters employed herein also encompass promoters utilized for transcription by viral RNA polymerases. Methods for introducing polynucleotides into plants and expressing a protein encoded therein, involving viral DNA or RNA molecules, are known in the art. See, for example, U.S. Patent Nos. 5,889,191, 5,889,190, 5,866,785, 5,589,367, 5,316,931, and Porta *et al.* (1996) *Molecular Biotechnology 5*:209-221; herein incorporated by reference.

In other embodiments, various components of the site-specific integration system can be provided to a plant using a variety of transient transformation methods. "Transient transformation" is intended to mean that a polynucleotide is introduced into the host (i.e., a plant) and expressed temporally. Such transient transformation methods include, but are not limited to, the introduction of any of the components of the site-specific integration system or active fragments or variants thereof directly into the plant or the introduction of the transcript into the plant. Such methods include, for example, microinjection or particle bombardment. See, for example, Crossway *et al.* (1986) *Mol Gen. Genet. 202*:179-185; Nomura *et al.* (1986) *Plant Sci. 44:53-58*; Hepler *et al.* (1994) *Proc. Natl. Acad. Sci. 91*: 2176-2180 and Hush *et al.* (1994) *The Journal of Cell Science 107*:775-784, all of which are herein incorporated by reference. Alternatively, the polynucleotide can be transiently transformed into the plant using techniques known in the art. Such techniques include viral vector system and the precipitation of the polynucleotide in a manner that precludes subsequent release of the DNA. Thus, the transcription from the particle-bound DNA can occur, but the frequency with which it is released to become integrated into the genome is greatly reduced. Such methods include the use particles coated with polyethylimine (PEI; Sigma #P3143).

The cells that have been transformed may be grown into plants in accordance with conventional ways. See, for example, McCormick *et al.* (1986) *Plant Cell Reports* 5:81-84. These plants may then be grown, and either pollinated with the same transformed strain or different strains, and the resulting progeny having constitutive expression of the desired phenotypic characteristic identified.

Two or more generations may be grown to ensure that expression of the desired phenotypic characteristic is stably maintained and inherited and then seeds harvested to ensure expression of the desired phenotypic characteristic has been achieved. In this manner, transformed seed (also referred to as "transgenic seed") having the recited DNA construct stably incorporated into their genome is provided.

V. Plants

Compositions provided herein encompass a plant cell, a plant, a plant part, and a seed comprising any of the components, or combination thereof, of the site-specific integration system disclosed herein (i.e. an endonuclease, a recognition site, a target site, a donor DNA, a transfer cassette, various site-specific recombination sites, site-specific recombinases, polynucleotides of interest, or any active variants or fragments thereof).

In one embodiment, a plant cell, a plant, a plant part and/or a seed is provided comprising a target site for site-specific integration integrated at the recognition site.

The compositions further provide a plant cell, a plant, a plant part and a seed comprising a transfer cassette integrated at the target site. In one embodiment, the plant cell, a plant, a plant part and a seed having the target site integrated at the recognition site comprises a target site comprising in the following order, a first recombination site, a second recombination site and wherein the first and the second recombination sites are dissimilar and non-recombinogenic with respect to one another. The target site can further comprise a polynucleotide of interest between the first and the second recombination sites. As described elsewhere herein, the recombination sites can be any combination of recombination sites known in the art. For example, the recombination sites can be a FRT site, a mutant FRT site, a LOX site or a mutant LOX site.

In specific embodiments, the target site of the plant cell, plant, plant part and seed further comprises a third recombination site between the first and the second recombination site, wherein the third recombination site is dissimilar and non-recombinogenic to the first and the second recombination sites. The first, second, and third recombination sites can comprise, for example, FRT1, FRT5, FRT6, FRT12, or FRT87. Also, provided is a plant cell, plant, or seed wherein the first

recombination site is FRT1, the second recombination site is FRT12 and the third recombination site is FRT87.

The plant cell, a plant, a plant part and a seed can comprise any of the recognition sites provided herein. For example, the recognition site can be selected from the group consisting of SEQ ID NO:1-8, 16 and 20 or an active variant thereof.

As used herein, the term plant includes plant cells, plant protoplasts, plant cell tissue cultures from which a plant can be regenerated, plant calli, plant clumps, and plant cells that are intact in plants or parts of plants such as embryos, pollen, ovules, seeds, leaves, flowers, branches, fruit, kernels, ears, cobs, husks, stalks, roots, root tips, anthers, and the like. Grain is intended to mean the mature seed produced by commercial growers for purposes other than growing or reproducing the species. Progeny, variants, and mutants of the regenerated plants are also included herein, provided that these parts comprise the recited DNA construct.

A transformed plant or transformed plant cell provided herein is one in which genetic alteration, such as transformation, has been affected as to a gene of interest, or is a plant or plant cell which is descended from a plant or cell so altered and which comprises the alteration. A “transgene” is a gene that has been introduced into the genome by a transformation procedure. Accordingly, a “transgenic plant” is a plant that contains a transgene, whether the transgene was introduced into that particular plant by transformation or by breeding; thus, descendants of an originally-transformed plant are encompassed by the definition. A “control” or “control plant” or “control plant cell” provides a reference point for measuring changes in phenotype of the subject plant or plant cell. A control plant or plant cell may comprise, for example: (a) a wild-type plant or cell, i.e., of the same genotype as the starting material for the genetic alteration which resulted in the subject plant or cell; (b) a plant or plant cell of the same genotype as the starting material but which has been transformed with a null construct (i.e., with a construct which does not express the transgene, such as a construct comprising a marker gene); (c) a plant or plant cell which is a non-transformed segregant among progeny of a subject plant or plant cell; (d) a plant or plant cell genetically identical to the subject plant or plant cell but which is not exposed to conditions or stimuli that would induce expression of the transgene; or (e) the subject plant or plant cell itself, under conditions in which the construct is not expressed.

Plant cells that have been transformed to have a component(s) of the site-specific integration system provided herein can be grown into whole plants. The regeneration, development, and cultivation of plants from single plant protoplast transformants or from various transformed explants is well known in the art. See, for example, McCormick *et al.* (1986) *Plant Cell Reports* 5:81-84; Weissbach and Weissbach, In: Methods for Plant Molecular Biology, (Eds.), Academic Press, Inc. San Diego, Calif., (1988). This regeneration and growth process typically includes the steps of selection of transformed cells, culturing those individualized cells through the usual stages of embryonic development through the rooted plantlet stage. Transgenic embryos and seeds are similarly regenerated. The resulting transgenic rooted shoots are thereafter planted in an appropriate plant growth medium such as soil. Preferably, the regenerated plants are self-pollinated to provide homozygous transgenic plants. Otherwise, pollen obtained from the regenerated plants is crossed to seed-grown plants of agronomically important lines. Conversely, pollen from plants of these important lines is used to pollinate regenerated plants. Two or more generations may be grown to ensure that expression of the desired phenotypic characteristic is stably maintained and inherited and then seeds harvested to ensure expression of the desired phenotypic characteristic has been achieved. In this manner, the compositions presented herein provide transformed seed (also referred to as "transgenic seed") having a polynucleotide provided herein, for example, a target site, stably incorporated into their genome.

The components of the site-specific integration system provided herein may be used for transformation of any plant species, including, but not limited to, monocots and dicots. Examples of plant species of interest include, but are not limited to, corn (maize) (*Zea mays*), *Brassica* sp. (e.g., *B. napus, B. rapa, B. juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), 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*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*), sweet

potato (*Ipomoea batatus*), cassava (*Manihot esculenta*), coffee (*Coffea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, barley, vegetables, ornamentals, and conifers.

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

Conifers that may be employed include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*); Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*). In specific embodiments, plants are crop plants (for example, corn, alfalfa, sunflower, *Brassica*, soybean, cotton, safflower, peanut, sorghum, wheat, millet, tobacco, etc.). In other embodiments, corn and soybean plants are optimal, and in yet other embodiments corn plants are optimal.

Other plants of interest include grain plants that provide seeds of interest, oil-seed plants, and leguminous plants. Seeds of interest include grain seeds, such as corn, wheat, barley, rice, sorghum, rye, etc. Oil-seed plants include cotton, soybean, safflower, sunflower, *Brassica*, maize, alfalfa, palm, coconut, etc. Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, etc.

It is recognized that the plant having stably incorporated the DNA construct can be further characterized for site-specific integration potential, agronomic potential, and copy number. See, U.S. Patent No. 6,187,994.

Depending on the polynucleotide(s) of interest incorporated into the target site, the transgenic plants, plant cells, or seeds comprising a target site with a polynucleotide(s) of interest provided herein may have a change in phenotype, including, but not limited to, an altered pathogen or insect defense mechanism, an increased resistance to one or more herbicides, an increased ability to withstand stressful environmental conditions, a modified ability to produce starch, a modified level of starch production, a modified oil content and/or composition, a modified carbohydrate content and/or composition, a modified fatty acid content and/or composition, a modified ability to utilize, partition and/or store nitrogen, and the like.

VI. Polynucleotides

Provided herein are polynucleotides or nucleic acid molecules comprising the various components of the site-specific integration system (i.e. an endonuclease, a recognition site, a target site, a donor DNA, a transfer cassette, various site-specific recombination sites, site-specific recombinases, polynucleotides of interest or any active variants or fragments thereof). Also provided are nucleic acid molecules comprising any of the various target sites provided herein integrated at the recognition site in the plant genome.

The terms “polynucleotide,” “polynucleotide sequence,” “nucleic acid sequence,” and “nucleic acid fragment” are used interchangeably herein. These terms encompass nucleotide sequences and the like. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof. The use of the term “polynucleotide” is not intended to limit the present invention to polynucleotides comprising DNA. Those of ordinary skill in the art will recognize that polynucleotides, can comprise ribonucleotides and combinations of ribonucleotides and deoxyribonucleotides. Such deoxyribonucleotides and ribonucleotides include both naturally occurring molecules and synthetic analogues. The polynucleotides provided herein also

encompass all forms of sequences including, but not limited to, single-stranded forms, double-stranded forms, hairpins, stem-and-loop structures, and the like.

The compositions provided herein can comprise an isolated or substantially purified polynucleotide. An "isolated" or "purified" polynucleotide is substantially or essentially free from components that normally accompany or interact with the polynucleotide as found in its naturally occurring environment. Thus, an isolated or purified polynucleotide 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.

Further provided are recombinant polynucleotides comprising the various components of the site-specific integration system. The terms "recombinant polynucleotide" and "recombinant DNA construct" are used interchangeably herein. A recombinant construct comprises an artificial or heterologous combination of nucleic acid sequences, e.g., regulatory and coding sequences that are not found together in nature. For example, a transfer cassette can comprise restriction sites and a heterologous polynucleotide of interest. In other embodiments, a recombinant 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 comprising any of the isolated nucleic acid fragments provided herein. The skilled artisan will also recognize that different

independent transformation events will result in different levels and patterns of expression (Jones *et al.*, *EMBO J.* 4:2411-2418 (1985); De Almeida *et al.*, *Mol. Gen. Genetics* 218:78-86 (1989)), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, immunoblotting analysis of protein expression, or phenotypic analysis, among others.

In specific embodiments, one or more of the components of the site-specific integration system described herein can be provided in an expression cassette for expression in a plant or other organism or cell type of interest. The cassette can include 5' and 3' regulatory sequences operably linked to a polynucleotide provided herein. "Operably linked" is intended to mean a functional linkage between two or more elements. For example, an operable linkage between a polynucleotide of interest and a regulatory sequence (i.e., a promoter) is a functional link that allows for expression of the polynucleotide of interest. Operably linked elements may be contiguous or non-contiguous. When used to refer to the joining of two protein coding regions, by operably linked is intended that the coding regions are in the same reading frame. The cassette may additionally contain at least one additional gene to be cotransformed into the organism. Alternatively, the additional gene(s) can be provided on multiple expression cassettes. Such an expression cassette is provided with a plurality of restriction sites and/or recombination sites for insertion of a recombinant polynucleotide to be under the transcriptional regulation of the regulatory regions. The expression cassette may additionally contain selectable marker genes.

The expression cassette can include in the 5'-3' direction of transcription, a transcriptional and translational initiation region (i.e., a promoter), a recombinant polynucleotide provided herein, and a transcriptional and translational termination region (i.e., termination region) functional in plants. The regulatory regions (i.e., promoters, transcriptional regulatory regions, and translational termination regions) and/or a polynucleotide provided herein may be native/analogous to the host cell or to each other. Alternatively, the regulatory regions and/or a polynucleotide provided herein may be heterologous to the host cell or to each other. As used herein, "heterologous" in reference to a sequence is a sequence that originates from a

foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, a promoter operably linked to a heterologous polynucleotide is from a species different from the species from which the polynucleotide was derived, or, if from the same/analogous species, one or both are substantially modified from their original form and/or genomic locus, or the promoter is not the native promoter for the operably linked polynucleotide. Alternatively, the regulatory regions and/or a recombinant polynucleotide provided herein may be entirely synthetic.

The termination region may be native with the transcriptional initiation region, may be native with the operably linked recombinant polynucleotide, may be native with the plant host, or may be derived from another source (i.e., foreign or heterologous) to the promoter, the recombinant polynucleotide, the plant host, or any combination thereof. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also Guerineau *et al.* (1991) *Mol. Gen. Genet.* 262:141-144; Proudfoot (1991) *Cell* 64:671-674; Sanfacon *et al.* (1991) *Genes Dev.* 5:141-149; Mogen *et al.* (1990) *Plant Cell* 2:1261-1272; Munroe *et al.* (1990) *Gene* 91:151-158; Ballas *et al.* (1989) *Nucleic Acids Res.* 17:7891-7903; and Joshi *et al.* (1987) *Nucleic Acids Res.* 15:9627-9639.

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

A number of promoters can be used in the expression cassettes provided herein. The promoters can be selected based on the desired outcome. It is recognized that different applications can be enhanced by the use of different promoters in the expression cassettes to modulate the timing, location and/or level of expression of the polynucleotide of interest. Such expression constructs may also contain, if desired, a promoter regulatory region (e.g., one conferring inducible, constitutive, environmentally- or developmentally-regulated, or cell- or tissue-

specific/selective expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

In some embodiments, an expression cassette provided herein can be combined with constitutive, tissue-preferred, or other promoters for expression in plants. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'-promoter derived from T-DNA of Agrobacterium tumefaciens, the ubiquitin 1 promoter, the Smas promoter, the cinnamyl alcohol dehydrogenase promoter (U.S. Pat. No. 5,683,439), the Nos promoter, the pEmu promoter, the rubisco promoter, the GRP1-8 promoter and other transcription initiation regions from various plant genes known to those of skill. If low level expression is desired, weak promoter(s) may be used. Weak constitutive promoters include, for example, the core promoter of the Rsyn7 promoter (WO 99/43838 and U.S. Pat. No. 6,072,050), the core 35S CaMV promoter, and the like. Other constitutive promoters include, for example, U.S. Pat. Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; and 5,608,142. See also, U.S. Pat. No. 6,177,611, herein incorporated by reference.

Examples of inducible promoters are the Adh1 promoter which is inducible by hypoxia or cold stress, the Hsp70 promoter which is inducible by heat stress, the PPDK promoter and the pepcarboxylase promoter which are both inducible by light. Also useful are promoters which are chemically inducible, such as the In2-2 promoter which is safener induced (U.S. Pat. No. 5,364,780), the ERE promoter which is estrogen induced, and the Axig1 promoter which is auxin induced and tapetum specific but also active in callus (PCT US01/22169).

Examples of promoters under developmental control include promoters that initiate transcription preferentially in certain tissues, such as leaves, roots, fruit, seeds, or flowers. An exemplary promoter is the anther specific promoter 5126 (U.S. Pat. Nos. 5,689,049 and 5,689,051). Examples of seed-preferred promoters include, but are not limited to, 27 kD gamma zein promoter and waxy promoter, Boronat, A. *et al.* (1986) *Plant Sci.* 47:95-102; Reina, M. *et al. Nucl. Acids Res.* 18(21):6426; and Kloesgen, R. B. *et al.* (1986) *Mol. Gen. Genet.* 203:237-244. Promoters that express in the embryo, pericarp, and endosperm are disclosed in

U.S. Pat. No. 6,225,529 and PCT publication WO 00/12733. The disclosures for each of these are incorporated herein by reference in their entirety.

Chemical-regulated promoters can be used to modulate the expression of a gene in a plant through the application of an exogenous chemical regulator. Depending upon the objective, 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 are known in the art and include, but are not limited to, the maize In2-2 promoter, which is activated by benzenesulfonamide herbicide safeners, the maize GST promoter, which is activated by hydrophobic electrophilic compounds that are used as pre-emergent herbicides, and the tobacco PR-1a promoter, which is activated by salicylic acid. Other chemical-regulated promoters of interest include steroid-responsive promoters (see, for example, the glucocorticoid-inducible promoter in Schena *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:10421-10425 and McNellis *et al.* (1998) *Plant J.* 14(2):247-257) and tetracycline-inducible and tetracycline-repressible promoters (see, for example, Gatz *et al.* (1991) *Mol. Gen. Genet.* 227:229-237, and U.S. Pat. Nos. 5,814,618 and 5,789,156), herein incorporated by reference.

Tissue-preferred promoters can be utilized to target enhanced expression of a polynucleotide of interest within a particular plant tissue. Tissue-preferred promoters are known in the art. See, for example, Yamamoto *et al.* (1997) *Plant J.* 12(2):255-265; Kawamata *et al.* (1997) *Plant Cell Physiol.* 38(7):792-803; Hansen *et al.* (1997) *Mol. Gen Genet.* 254(3):337-343; Russell *et al.* (1997) *Transgenic Res.* 6(2):157-168; Rinehart *et al.* (1996) *Plant Physiol.* 112(3):1331-1341; Van Camp *et al.* (1996) *Plant Physiol.* 112(2):525-535; Canevascini *et al.* (1996) *Plant Physiol.* 112(2):513-524; Yamamoto *et al.* (1994) *Plant Cell Physiol.* 35(5):773-778; Lam (1994) *Results Probl. Cell Differ.* 20:181-196; Orozco *et al.* (1993) *Plant Mol Biol.* 23(6):1129-1138; Matsuoka *et al.* (1993) *Proc Natl. Acad. Sci. USA* 90(20):9586-9590; and Guevara-Garcia *et al.* (1993) *Plant J.* 4(3):495-505. Such promoters can be modified, if necessary, for weak expression.

Leaf-preferred promoters are known in the art. See, for example, Yamamoto *et al.* (1997) *Plant J.* 12(2):255-265; Kwon *et al.* (1994) *Plant Physiol.* 105:357-67; Yamamoto *et al.* (1994) *Plant Cell Physiol.* 35(5):773-778; Gotor *et al.* (1993) *Plant*

J. 3:509-18; Orozco *et al.* (1993) *Plant Mol. Biol.* 23(6):1129-1138; and Matsuoka *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90(20):9586-9590. In addition, the promoters of cab and rubisco can also be used. See, for example, Simpson *et al.* (1958) *EMBO J* 4:2723-2729 and Timko *et al.* (1988) *Nature* 318:57-58.

Root-preferred promoters are known and can be selected from the many available from the literature or isolated *de novo* from various compatible species. See, for example, Hire *et al.* (1992) *Plant Mol. Biol.* 20(2):207-218 (soybean root-specific glutamine synthetase gene); Keller and Baumgartner (1991) *Plant Cell* 3(10):1051-1061 (root-specific control element in the GRP 1.8 gene of French bean); Sanger *et al.* (1990) *Plant Mol. Biol.* 14(3):433-443 (root-specific promoter of the mannopine synthase (MAS) gene of *Agrobacterium tumefaciens*); and Miao *et al.* (1991) *Plant Cell* 3(1):11-22 (full-length cDNA clone encoding cytosolic glutamine synthetase (GS), which is expressed in roots and root nodules of soybean). See also Bogusz *et al.* (1990) *Plant Cell* 2(7):633-641, where two root-specific promoters isolated from hemoglobin genes from the nitrogen-fixing nonlegume *Parasponia andersonii* and the related non-nitrogen-fixing nonlegume *Trema tomentosa* are described. The promoters of these genes were linked to a β-glucuronidase reporter gene and introduced into both the nonlegume *Nicotiana tabacum* and the legume *Lotus corniculatus*, and in both instances root-specific promoter activity was preserved. Leach and Aoyagi (1991) describe their analysis of the promoters of the highly expressed roIC and roID root-inducing genes of *Agrobacterium rhizogenes* (see *Plant Science* (Limerick) 79(1):69-76). They concluded that enhancer and tissue-preferred DNA determinants are dissociated in those promoters. Teeri *et al.* (1989) used gene fusion to lacZ to show that the Agrobacterium T-DNA gene encoding octopine synthase is especially active in the epidermis of the root tip and that the TR2' gene is root specific in the intact plant and stimulated by wounding in leaf tissue, an especially desirable combination of characteristics for use with an insecticidal or larvicidal gene (see *EMBO J.* 8(2):343-350). The TR1' gene, fused to nptII (neomycin phosphotransferase II) showed similar characteristics. Additional root-preferred promoters include the VfENOD-GRP3 gene promoter (Kuster *et al.* (1995) P*lant Mol. Biol.* 29(4):759-772); and roIB promoter (Capana *et al.* (1994) *Plant Mol. Biol.* 25(4):681-691. See also U.S. Pat. Nos. 5,837,876; 5,750,386; 5,633,363; 5,459,252; 5,401,836; 5,110,732; and 5,023,179. The phaseolin gene

(Murai *et al.* (1983) *Science* 23:476-482 and Sengopta-Gopalen *et al.* (1988) *PNAS* 82:3320-3324.

The expression cassette containing the polynucleotides provided herein can also comprise a selectable marker gene for the selection of transformed cells. Selectable marker genes are utilized for the selection of transformed cells or tissues. Marker genes include genes encoding antibiotic resistance, such as those encoding neomycin phosphotransferase II (NEO) and hygromycin phosphotransferase (HPT), as well as genes conferring resistance to herbicidal compounds, such as glufosinate ammonium, bromoxynil, imidazolinones, and 2,4-dichlorophenoxyacetate (2,4-D) and sulfonylureas. Additional selectable markers include phenotypic markers such as beta-galactosidase and fluorescent proteins such as green fluorescent protein (GFP) (Su *et al.* (2004) *Biotechnol. Bioeng.* 85:610-9 and Fetter *et al.* (2004) *Plant Cell* 16:215-28), cyan fluorescent protein (CYP) (Bolte et al. (2004) *J. Cell Science* 117:943-54 and Kato *et al.* (2002) *Plant Physiol.* 129:913-42), and yellow fluorescent protein (PhiYFP.TM. from Evrogen; see, Bolte *et al.* (2004) *J. Cell Science* 117:943-54). For additional selectable markers, see generally, Yarranton (1992) *Curr. Opin. Biotech.* 3:506-511; Christopherson *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6314-6318; Yao *et al.* (1992) Cell 71:63-72; Reznikoff (1992) *Mol. Microbiol.* 6:2419-2422; Barkley *et al.* (1980) in The Operon, pp. 177-220; Hu *et al.* (1987) *Cell* 48:555-566; Brown *et al.* (1987) *Cell* 49:603-612; Figge *et al.* (1988) *Cell* 52:713-722; Deuschle *et al.* (1989) *Proc. Natl. Acad. Aci. USA* 86:5400-5404; Fuerst *et al.* (1989) *Proc. Nail. Acad. Sci.* USA 86:2549-2553; Deuschle *et al.* (1990) *Science* 248:480-483; Gossen (1993) Ph.D. Thesis, University of Heidelberg; Reines *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:1917-1921; Labow *et al.* (1990) *Mol. Cell. Biol.* 10:3343-3356; Zambretti *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3952-3956; Baim *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:5072-5076; Wyborski *et al.* (1991) *Nucleic Acids Res.* 19:4647-4653; Hillenand-Wissman (1989) *Topics Mol. Struc. Biol.* 10:143-162; Degenkolb *et al.* (1991) *Antimicrob. Agents Chemother.* 35:1591-1595; Kleinschnidt *et al.* (1988) *Biochemistry* 27:1094-1104; Bonin (1993) Ph.D. Thesis, University of Heidelberg; Gossen *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Oliva *et al.* (1992) *Antimicrob. Agents Chemother.* 36:913-919; Hlavka *et al.* (1985) Handbook of Experimental Pharmacology, Vol. 78 (Springer-Verlag, Berlin); Gill *et al.* (1988)

Nature 334:721-724. Such disclosures are herein incorporated by reference. The above list of selectable marker genes is not meant to be limiting. Any selectable marker gene can be used in the compositions presented herein.

Where appropriate, the sequences employed in the methods and compositions (i.e., the polynucleotide of interest, the recombinase, the endonuclease, etc.) may be optimized for increased expression in the transformed plant. That is, the genes can be synthesized using plant-preferred codons for improved expression. See, for example, Campbell and Gowri (1990) *Plant Physiol.* 92:1-11 for a discussion of host-preferred codon usage. Methods are available in the art for synthesizing plant-preferred genes. See, for example, U.S. Patent Nos. 5,380,831, and 5,436,391, and Murray *et al.* (1989) *Nucleic Acids Res.* 17:477-498, herein incorporated by reference.

VII. Fragments, Variants and Sequence Comparisons

The methods and compositions provided herein employ a variety of different components of the site-specific integration system (i.e. an endonuclease, a recognition site, a target site, a donor DNA, a transfer cassette, various site-specific recombination sites, site-specific recombinases, polynucleotides of interest or any active variants or fragments thereof). It is recognized throughout the description that some components of the site-specific integration system can have active variants and fragments. Such components include, for example, endonucleases (i.e. engineered endonucleases), endonuclease recognition sites (i.e. SEQ ID NOS: 1-8, 16, 120), recombinases (i.e. SEQ ID NOS: 15, 21, 26), recombination sites (i.e. SEQ ID NO: 9-13), and polynucleotides of interest. Biological activity for each of these components is described elsewhere herein.

Fragments and variants of the endonucleases, endonuclease recognition sites, recombinases, recombination sites, and polynucleotides of interest are also encompassed herein. By "fragment" is intended a portion of the polynucleotide or a portion of the amino acid sequence and hence protein encoded thereby. Fragments of a polynucleotide may encode protein fragments that retain the biological activity of the native protein (i.e., a fragment of a recombinase implements a recombination event). As used herein, a "native" polynucleotide or polypeptide comprises a naturally occurring nucleotide sequence or amino acid sequence, respectively.

Thus, fragments of a polynucleotide may range from at least about 20 nucleotides, about 50 nucleotides, about 100 nucleotides, and up to the full-length polynucleotide. A fragment of a polynucleotide that encodes a biologically active portion of a protein employed in the methods or compositions will encode at least 15, 25, 30, 50, 100, 150, 200, or 250 contiguous amino acids, or up to the total number of amino acids present in a full-length protein. Alternatively, fragments of a polynucleotide that are useful as a hybridization probe generally do not encode fragment proteins retaining biological activity. Thus, fragments of a nucleotide sequence may range from at least about 10, 20, 30, 40, 50, 60, 70, 80 nucleotides or up to the full length sequence.

A biologically active portion of a polypeptide can be prepared by isolating a portion of one of the polynucleotides encoding the portion of the polypeptide of interest and expressing the encoded portion of the protein (e.g., by recombinant expression *in vitro*), and assessing the activity of the portion of the polypeptide. For example, polynucleotides that encode fragments of a recombinase polypeptide can comprise nucleotide sequence comprising at least 16, 20, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 800, 900, 1,000, 1,100, 1,200, 1,300, or 1,400 nucleotides, or up to the number of nucleotides present in a nucleotide sequence employed in the methods and compositions provided herein.

"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 the native polynucleotide. 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 polypeptides employed in the compositions and methods provided herein. Naturally occurring allelic variants such as these, or naturally occurring allelic variants of polynucleotides can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques as outlined below. Variant polynucleotides also include synthetically derived polynucleotides, such as those generated, for example, by using site-directed mutagenesis. Generally, variants of

a particular polynucleotide employed in the methods and compositions provided herein will have 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 to that particular polynucleotide as determined by sequence alignment programs and parameters as described elsewhere herein.

Variants of a particular polynucleotide employed in the methods and compositions provided herein (i.e., endonucleases, endonuclease recognition sites, recombinases, recombination sites, and polynucleotides of interest) 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. Thus, for example, an isolated polynucleotide that encodes a polypeptide with a given percent sequence identity to the polypeptide are disclosed. 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 provided herein 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 employed in the methods and compositions provided herein are biologically active, that is they continue to possess the desired biological activity of the native protein. Such variants may result from, for example, genetic polymorphism or from human manipulation. Biologically active variants of a native protein provided herein will have 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 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 provided herein may differ from that protein by as few as 1-15 amino

acid residues, as few as 1-10, such as 6-10, as few as 5, as few as 4, 3, 2, or even 1 amino acid residue.

Proteins may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of the recombinase proteins can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. See, for example, Kunkel (1985) *Proc. Natl. Acad. Sci. USA* 82:488-492; Kunkel *et al.* (1987) *Methods in Enzymol.* 154:367-382; U.S. 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 preferable.

Thus, the polynucleotides used herein can include the naturally occurring sequences, the "native" sequences, as well as mutant forms. Likewise, the proteins used in the methods provided herein encompass both naturally occurring proteins as well as variations and modified forms thereof. Obviously, the mutations that will be made in the polynucleotide encoding the variant polypeptide must not place the sequence out of reading frame and preferably 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.

Variant polynucleotides and proteins also encompass sequences and proteins derived from a mutagenic and recombinogenic procedure such as DNA shuffling. With such a procedure, for example, one or more different recombinase coding sequences can be manipulated to create a new recombinase protein

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 following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides. As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence. As used herein, "comparison window" makes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Sequence relationships can be analyzed and described using computer-implemented algorithms. The sequence relationship between two or more polynucleotides, or two or more polypeptides can be determined by determining the best alignment of the sequences, and scoring the matches and the gaps in the alignment, which yields the percent sequence identity, and the percent sequence similarity. Polynucleotide relationships can also be described based on a comparison of the polypeptides each encodes. Many programs and algorithms for the comparison and analysis of sequences are well-known in the art.

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 GAP Weight of 50 and Length Weight of 3, and the nwsgapdna.cmp scoring matrix; % identity and % similarity for an amino acid sequence using GAP Weight of 8 and Length Weight of 2, and the BLOSUM62 scoring matrix (Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci USA 89*:10915); or any equivalent program thereof. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by GAP Version 10

GAP uses the algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443-453, to find the 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. It allows for the provision of a gap creation penalty and a gap extension penalty in units of matched bases. GAP must make a profit of gap creation penalty number of matches for each gap it inserts. If a gap extension penalty greater than zero is chosen, GAP must, in addition, make a profit for each gap inserted of the length of the gap times the gap extension penalty. Default gap creation penalty values and gap extension penalty values in Version 10 of the GCG Wisconsin Genetics Software Package for protein sequences are 8 and 2, respectively. For nucleotide sequences the default gap creation penalty is 50 while the default gap extension penalty is 3. The gap creation and gap extension penalties can be expressed as an integer selected from the group of integers consisting of from 0 to 200. Thus, for example, the gap creation and gap extension penalties can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 or greater.

GAP presents one member of the family of best alignments. There may be many members of this family, but no other member has a better quality. GAP displays four figures of merit for alignments: Quality, Ratio, Identity, and Similarity. The Quality is the metric maximized in order to align the sequences. Ratio is the quality divided by the number of bases in the shorter segment. Percent Identity is the percent of the symbols that actually match. Percent Similarity is the percent of

the symbols that are similar. Symbols that are across from gaps are ignored. A similarity is scored when the scoring matrix value for a pair of symbols is greater than or equal to 0.50, the similarity threshold.

As used herein, "sequence identity" or "identity" in the context of two polynucleotides or polypeptide sequences makes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have "sequence similarity" or "similarity". Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California).

As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide 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 result by 100 to yield the percentage of sequence identity.

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

1. A method for introducing into the genome of a plant cell a target site for site-specific integration, the method comprising:

(a) providing a plant cell comprising in its genome an endogenous recognition site for an engineered endonuclease, wherein the engineered endonuclease is capable of inducing a double-strand break in said endogenous recognition site, and wherein the endogenous recognition site is located between a first and a second genomic region ;

(b) providing a donor DNA comprising the target site for site-specific integration located between a first region of homology to said first genomic region and a second region of homology to said second genomic region, wherein the target site comprises a first and a second recombination site, wherein the first and the second recombination sites are dissimilar and non-recombinogenic with respect to one another;

(c) contacting the plant cell with the donor DNA and the engineered endonuclease, and

(d) identifying at least one plant cell from (c) comprising in its genome the target site integrated at the endogenous recognition site.

2. The method of embodiment 1, wherein the first region of homology further comprises a first fragment of said endogenous recognition site of (a), and wherein the second region of homology comprises a second fragment of said endogenous recognition site of (a), wherein the first and second fragments are dissimilar.

3. The method of embodiment 1, wherein the first region of homology further comprises the first 13 bases of said endogenous recognition site of (a), and wherein the second region of homology comprises the last 9 bases of said endogenous recognition site of (a).

4. The method of any one of embodiments 1-3, further comprising recovering a fertile plant from the cell of (d), the fertile plant comprising in its genome the target site integrated into the endogenous recognition site.

5. The method of any one of embodiments 1-4, wherein the endogenous recognition site is selected from the group consisting of SEQ ID NO:1-8, 16 and 20 or a sequence having at least 90% sequence identity to SEQ ID NO:1-8, 16 and 20.

6. The method of any one of embodiments 1-5, wherein the target site further comprises a polynucleotide of interest between the first recombination site and the second recombination site.

7. The method of any one of embodiments 1-6, wherein at least one of the first and the second recombination sites comprises an FRT site, a mutant FRT site, a LOX site, and a mutant LOX site.

8. The method of any one of embodiments 1-7, wherein the target site further comprises a third recombination site between the first and the second recombination site, wherein the third recombination site is dissimilar and non-recombinogenic to the first and the second recombination sites.

9. The method of embodiment 8, wherein at least one of the first, the second, and the third recombination sites comprises FRT1 (SEQ ID NO: 9), FRT 5 (SEQ ID NO: 10), FRT6 (SEQ ID NO: 11), FRT12 (SEQ ID NO: 12) and FRT87 (SEQ ID NO: 13).

10. The method of embodiment 8, wherein the first recombination site is FRT1 (SEQ ID NO: 9), the second recombination site is FRT12 (SEQ ID NO: 12), and the third recombination site is FRT87 (SEQ ID NO: 13).

11. The method of any one of embodiments 1-10, wherein the engineered endonuclease is derived from I-CreI.

12. The method of any one of embodiments 1-11, wherein said plant cell is from a monocot.

13. The method of embodiment 12, wherein said monocot is maize, wheat, rice, barley, sorghum, or rye.

14. The method of any one of embodiments 1-11, wherein said plant cell is from a dicot.

15. The method of embodiment 14, wherein said dicot is soybean, Brassica, sunflower, cotton, or alfalfa.

16. A plant cell, plant part, plant, or seed comprising the target site integrated at the endogenous recognition site according to any one of embodiments 1-15.

17. A nucleic acid molecule comprising the target site integrated at the endogenous recognition site according to any one of embodiments 1-15.

18. A plant cell, plant part, plant, or seed comprising the nucleic acid molecule of embodiment 17.

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

(a) providing at least one plant cell comprising in its genome a target site for site-specific integration, wherein the target site is integrated into an endogenous recognition site for an engineered endonuclease, and wherein the target site is,

(i) a target site comprising a first and a second recombination site; or,

(ii) the target site of (i) further comprising a third recombination site between the first recombination site and the second recombination site,

wherein the engineered endonuclease is capable of inducing a double-strand break in the endogenous recognition site, wherein the first, the second, and the third recombination sites are dissimilar and non-recombinogenic with respect to one another,

(b) introducing into the plant cell of (a) a transfer cassette comprising,

(iii) the first recombination site, a first polynucleotide of interest, and the second recombination site,

(iv) the second recombination site, a second polynucleotide of interest, and the third recombination sites, or

(v) the first recombination site, a third polynucleotide of interest, and the third recombination sites;

(c) providing a recombinase that recognizes and implements recombination at the first and the second recombination sites, at the second and the third recombination sites, or at the first and third recombination sites; and

(d) selecting at least one plant cell comprising integration of the transfer cassette at the target site.

20. The method of embodiment 19, further comprising recovering a fertile plant from the plant cell of (d), the fertile plant comprising in its genome the transfer cassette at the target site.

21. The method of any one of any one of embodiments 19-20, wherein at least one of the first, the second, and the third polynucleotides of interest comprises a nucleotide sequence for gene silencing, a nucleotide sequence encoding a

phenotypic marker, or a nucleotide sequence encoding a protein providing an agronomic advantage.

22. The method of any one of embodiments 19-21, wherein providing the recombinase comprises integrating into the genome of the plant cell a nucleotide sequence encoding the recombinase.

23. The method of any one of embodiment 19-22, wherein the transfer cassette further comprises at least one coding region operably linked to a promoter that drives expression in the plant cell.

24. The method of any one of embodiments 19-23, wherein the transfer cassette further comprises a coding region operably linked to a promoter that drives expression in the plant cell, wherein the coding region encodes a recombinase that facilitates recombination between, the first and the second recombination sites of the transfer cassette and the target site, the second and the third recombination sites of the transfer cassette and the target site, or the first and the third recombination sites of the transfer cassette and the target site.

25. The method of any one of embodiment 19-24, wherein at least one of the first, the second, and the third recombination sites comprises an FRT site, a mutant FRT site, a LOX site, or a mutant LOX site.

26. The method of any one of embodiments 19-24, wherein the first recombination site is FRT1 (SEQ ID NO: 9), the second recombination site is FRT12 (SEQ ID NO: 12), and the third recombination site is FRT87 (SEQ ID NO: 13).

27. The method of any one of embodiments 19-26, wherein the recombinase is FLP.

28. The method of embodiment 27, wherein the FLP has been synthesized using maize-preferred codons.

29. The method of any one of embodiments 19-28, wherein said plant cell is from a monocot.

30. The method of embodiment 29, wherein said monocot is maize, wheat, rice, barley, sorghum, or rye.

31. The method of any one of embodiments 19-28 wherein said plant cell is from a dicot.

32. The method of embodiment 31, wherein said dicot is soybean, Brassica, sunflower, cotton, or alfalfa.

33. A plant cell, plant part, plant, or seed comprising the transfer cassette integrated at the target site according to any one of embodiments 19-32.

34. A plant, seed or plant cell comprising in its genome a target site for site-specific integration, wherein the target site is integrated into an endogenous recognition site for an engineered endonuclease, wherein the target site comprises in the following order:

(a) a first recombination site;

(b) a second recombination site, and

wherein the engineered endonuclease is capable of inducing a double-strand break at the endogenous recognition site, wherein the first and the second recombination sites are dissimilar and non-recombinogenic with respect to one another.

35. The plant, seed or plant cell of embodiment 34, wherein the endogenous recognition site is selected from the group consisting of SEQ ID NO:1-8, 16 and 20 or a sequence having at least 90% sequence identity to SEQ ID NO:1-8,16 and 20.

36. The plant, seed or plant cell of any of embodiments 35-36, wherein the target site further comprises a polynucleotide of interest between the first recombination site and the second recombination site.

37. The plant, seed or plant cell of any one of embodiments 35-36, wherein at least one of the first and the second recombination sites comprises an FRT site, a mutant FRT site, a LOX site, or a mutant LOX site.

38. The plant, seed or plant cell of any one of embodiments 35-37, wherein the target site further comprises a third recombination site between the first and the second recombination site, wherein the third recombination site is dissimilar and non-recombinogenic to the first and the second recombination sites.

39. The plant, seed or plant cell of embodiment 38, wherein at least one of the first, the second, and the third recombination sites comprises FRT1 (SEQ ID NO: 9), FRT 5 (SEQ ID NO: 10), FRT6 (SEQ ID NO: 11), FRT12 (SEQ ID NO: 12) and FRT87 (SEQ ID NO: 13).

40. The plant, seed or plant cell of embodiment 38, wherein the first recombination site is FRT1 (SEQ ID NO: 9), the second recombination site is FRT12 (SEQ ID NO: 12), and the third recombination site is FRT87 (SEQ ID NO: 13).

41. The plant, seed, or plant cell of any one of embodiments 35-40, wherein said plant, seed or plant cell is from a monocot.

42. The plant, seed, or plant cell of embodiment 41, wherein said monocot is maize, wheat, rice, barley, sorghum, or rye.

43. The plant, seed, or plant cell of any one of embodiments 35-40, wherein said plant, seed or plant cell is from a dicot.

44. The plant, seed, or plant cell of embodiment 43, wherein said dicot is soybean, Brassica, sunflower, cotton, or alfalfa.

Recombinase Medicated DNA Casette exchange RMCE using different recombinase systems have been achieved successfully in several plants (Nanto K, Yamada-Watanabet K, Ebinuma H(2005) Agrobacterium-mediated RMCE approach for gene replacement. Plant Biotechnol J, 3: 203–214; Louwerse JD et al. 2007. Stable recombinase-mediated cassette exchange in Arabidopsis using Agrobacterium tumefaciens. Plant Physiol 145: 1282–1293; Li Z. et al. 2009, Site-specific integration of transgenes in soybean via recombinase-mediated DNA cassette exchange. Plant Physiol 151: 1087–1095). Groups of transgenes can be stacked to the same site through multiple rounds of RMCE (Li et al 2010, Published online before print August 2010, doi:10.1104/pp.110.160093; Plant Physiology October 2010 vol. 154 no. 2 622-631). Taking advantage of reversible DNA cassette exchange in RMCE, an RMCE product can be used as a new target for subsequent SSI transformation. RMCE is a complex process especially when there are two targets, one on each homologous chromosome, and the two recombinase recognition sites involved are only partially incompatible (Li et al., 2009). The process is further complicated in gene stacking by using three recognition sites and large donor DNA containing multiple genes with some repeated sequences (Li et al 2010).

The DNA repair mechanisms of cells are the basis of transformation to introduce extraneous DNA or induce mutations on endogenous genes. DNA homologous recombination is a specialized way of DNA repair that the cells repair DNA damages using a homologous sequence. In plants, DNA homologous recombination happens at frequencies too low to be 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).

Recent developments in plant gene targeting demonstrate that endogenous genomic sites can be specifically targeted for modification through DNA double-strand break-induced homologous recombination (US patent application 12/147834, filed on June 27, 2008 and US provisional application 61/466602 filed on March 23, 2011, and herein incorporated by reference in their entirety). DNA double-strand breaks can be created with either designed zinc finger nucleases or modified homing endonucleases. Customized zinc finger nucleases have been employed to introduce successfully an herbicide resistance gene, PAT, to a tobacco (Nicotiana tabacum) endochitinase gene locus, a maize (Zea mays) inositol-1,3,4,5,6-petakisphosphate 2-kinase gene locus, or to introduce specific mutations to a tobacco acetolactate synthase gene to gain resistance to sulfonyl urea (Li. et al 2010). Similarly, an engineered I-CreI endonuclease derivative designed to recognize a selected sequence adjacent to the maize LIGULELESS1 gene has been used to produce mutations with small deletions or insertions specifically at expected cleavage sites (US patent application 12/147834, filed on June 27, 2008).

The following examples are offered by way of illustration and not by way of limitation.

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 followed by Site specific integration

Figure 1 provides a non-limiting example of targeted integration of a target site for site-specific recombination at a double strand break induced by an endonuclease at an endogenous recognition site and subsequent modification of the integrated target site. In Figure 1A, a plant having in its genome an endogenous recognition site for an endonuclease flanked by a first genomic region (DNA1) and a second genomic region (DNA2) is provided. A donor DNA is introduced into the plant cell comprising a nuclease gene for inducing a double strand break, a promoter, and a target site comprising two dissimilar and non-recombinogenic site-specific recombination sites (i.e. FRT1 and FRT87) and a first marker gene (Marker 1) whereby the target site is flanked by a first region of homology to DNA1 (HR1) and a second region of homology to DNA2 (HR2). The endonuclease induces a double strand break in the genomic DNA at the endogenous recognition site and the donor DNA and genomic DNA undergoes homologous recombination at the corresponding DNA1 and DNA2 regions. The resulting genomic structure with the integrated target site is depicted in Figure 1D. The integrated target site can be altered by site-specific recombination by providing to the cell a transfer cassette comprising the same dissimilar and non-recombinogenic recombination sites as the integrated target site (i.e. FRT1 and FRT87), a second marker gene (Marker 2), and a gene of interest. In the presence of the appropriate recombinase (FLP recombinase), the integrated target site is altered by site-specific recombination and the resulting genomic target site comprises Marker 2 and the gene of interest (Figure 1G).

EXAMPLE 2

Transformation of Maize Immature Embryos

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

a. Particle-mediated delivery

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

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

A plasmid comprising the Zm-BBM (also referred to as Zm-ODP2) coding sequence (set forth in SEQ ID NO: 9) operably linked to a promoter is constructed. This could be a weak promoter such as nos, a tissue-specific promoter, such as globulin-1 or oleosin, an inducible promoter such as In2, or a strong promoter such as ubiquitin plus a plasmid containing the selectable marker gene phosphinothricin N-acetyltransferase (PAT; Wohlleben *et al*. (1988) *Gene* 70:25 37) that confers resistance to the herbicide bialaphos. Furthermore, plasmids containing the double strand brake inducing agent and donor DNA such as PHP44285 or PHP44779 are constructed as described above and co-bombareded with the plasmids containing the developmental genes ODP2 (AP2 domain transcription factor ODP2 (Ovule development protein 2); US20090328252 A1) and Wushel.

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

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

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

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

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

Bombardment medium (560Y) comprises 4.0 g/l N6 basal salts (SIGMA C-1416), 1.0 ml/l Eriksson's Vitamin Mix (1000X SIGMA-1511), 0.5 mg/l thiamine HCl, 120.0 g/l sucrose, 1.0 mg/l 2,4-D, and 2.88 g/l L-proline (brought to volume with D-I

H2O following adjustment to pH 5.8 with KOH); 2.0 g/l Gelrite (added after bringing to volume with D-I H2O); and 8.5 mg/l silver nitrate (added after sterilizing the medium and cooling to room temperature).

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

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

Hormone-free medium (272V) comprises 4.3 g/l MS salts (GIBCO 11117-074), 5.0 ml/l MS vitamins stock solution (0.100 g/l nicotinic acid, 0.02 g/l thiamine HCL, 0.10 g/l pyridoxine HCL, and 0.40 g/l glycine brought to volume with polished D-I H2O), 0.1 g/l myo-inositol, and 40.0 g/l sucrose (brought to volume with polished D-I H2O after adjusting pH to 5.6); and 6 g/l bacto-agar (added after bringing to volume with polished D-I H2O), sterilized and cooled to 60ºC.

b. *Agrobacterium*-mediated transformation

Agrobacterium-mediated transformation was performed essentially as described in Djukanovic *et al*. (2006) *Plant Biotech J* 4:345-57. Briefly, 10-12 day old immature embryos (0.8 -2.5 mm in size) were dissected from sterilized kernels and placed into liquid medium (4.0 g/L N6 Basal Salts (Sigma C-1416), 1.0 ml/L Eriksson's Vitamin Mix (Sigma E-1511), 1.0 mg/L thiamine HCl, 1.5 mg/L 2, 4-D, 0.690 g/L L-proline, 68.5 g/L sucrose, 36.0 g/L glucose, pH 5.2). After embryo collection, the medium was replaced with 1 ml *Agrobacterium* at a concentration of 0.35-0.45 OD550. Maize embryos were incubated with *Agrobacterium* for 5 min at room temperature, then the mixture was poured onto a media plate containing 4.0

g/L N6 Basal Salts (Sigma C-1416), 1.0 ml/L Eriksson's Vitamin Mix (Sigma E-1511), 1.0 mg/L thiamine HCl, 1.5 mg/L 2, 4-D, 0.690 g/L L-proline, 30.0 g/L sucrose, 0.85 mg/L silver nitrate, 0.1 nM acetosyringone, and 3.0 g/L Gelrite, pH 5.8. Embryos were incubated axis down, in the dark for 3 days at 20°C, then incubated 4 days in the dark at 28°C, then transferred onto new media plates containing 4.0 g/L N6 Basal Salts (Sigma C-1416), 1.0 ml/L Eriksson's Vitamin Mix (Sigma E-1511), 1.0 mg/L thiamine HCl, 1.5 mg/L 2, 4-D, 0.69 g/L L-proline, 30.0 g/L sucrose, 0.5 g/L MES buffer, 0.85 mg/L silver nitrate, 3.0 mg/L Bialaphos, 100 mg/L carbenicillin, and 6.0 g/L agar, pH 5.8. Embryos were subcultured every three weeks until transgenic events were identified. Somatic embryogenesis was induced by transferring a small amount of tissue onto regeneration medium (4.3 g/L MS salts (Gibco 11117), 5.0 ml/L MS Vitamins Stock Solution, 100 mg/L myo-inositol, 0.1 µM ABA, 1 mg/L IAA, 0.5 mg/L zeatin, 60.0 g/L sucrose, 1.5 mg/L Bialaphos, 100 mg/L carbenicillin, 3.0 g/L Gelrite, pH 5.6) and incubation in the dark for two weeks at 28°C. All material with visible shoots and roots were transferred onto media containing 4.3 g/L MS salts (Gibco 11117), 5.0 ml/L MS Vitamins Stock Solution, 100 mg/L myo-inositol, 40.0 g/L sucrose, 1.5 g/L Gelrite, pH 5.6, and incubated under artificial light at 28°C. One week later, plantlets were moved into glass tubes containing the same medium and grown until they were sampled and/or transplanted into soil.

EXAMPLE 3

Transient Expression of BBM Enhances Transformation

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

Briefly, gold particles were coated with PEI as follows. First, the gold particles were washed. Thirty-five mg of gold particles, 1.0 in average diameter (A.S.I. #162-0010), were weighed out in a microcentrifuge tube, and 1.2 ml absolute

EtOH was added and vortexed for one minute. The tube was incubated for 15 minutes at room temperature and then centrifuged at high speed using a microfuge for 15 minutes at 4oC. The supernatant was discarded and a fresh 1.2 ml aliquot of ethanol (EtOH) was added, vortexed for one minute, centrifuged for one minute, and the supernatant again discarded (this is repeated twice). A fresh 1.2 ml aliquot of EtOH was added, and this suspension (gold particles in EtOH) was stored at –20oC for weeks. To coat particles with polyethylimine (PEI; Sigma #P3143), 250 µl of the washed gold particle/EtOH mix was centrifuged and the EtOH discarded. The particles were washed once in 100 µl ddH2O to remove residual ethanol, 250 µl of 0.25 mM PEI was added, followed by a pulse-sonication to suspend the particles and then the tube was plunged into a dry ice/EtOH bath to flash-freeze the suspension, which was then lyophilized overnight. At this point, dry, coated particles could be stored at -80oC for at least 3 weeks. Before use, the particles were rinsed 3 times with 250 µl aliquots of 2.5 mM HEPES buffer, pH 7.1, with 1x pulse-sonication, and then a quick vortex before each centrifugation. The particles were then suspended in a final volume of 250 µl HEPES buffer. A 25 µl aliquot of the particles was added to fresh tubes before attaching DNA. To attach uncoated DNA, the particles were pulse-sonicated, then 1 µg of DNA (in 5 µl water) was added, followed by mixing by pipetting up and down a few times with a Pipetteman and incubated for 10 minutes. The particles were spun briefly (i.e. 10 seconds), the supernatant removed, and 60 µl EtOH added. The particles with PEI-precipitated DNA-1 were washed twice in 60 µl of EtOH. The particles were centrifuged, the supernatant discarded, and the particles were resuspended in 45 µl water. To attach the second DNA (DNA-2), precipitation using TFX-50 was used. The 45 µl of particles/DNA-1 suspension was briefly sonicated, and then 5 µl of 100 ng/µl of DNA-2 and 2.5 µl of TFX-50 were added. The solution was placed on a rotary shaker for 10 minutes, centrifuged at 10,000g for 1 minute. The supernatant was removed, and the particles resuspended in 60 µl of EtOH. The solution was spotted onto macrocarriers and the gold particles onto which DNA-1 and DNA-2 had been sequentially attached were delivered into scutellar cells of 10 DAP Hi-II immature embryos using a standard protocol for the PDS-1000. For this experiment, the DNA-1 plasmid contained a UBI::RFP::pinII expression cassette, and DNA-2 contained a UBI::CFP::pinII expression cassette. Two days after bombardment,

transient expression of both the CFP and RFP fluorescent markers was observed as numerous red & blue cells on the surface of the immature embryo. The embryos were then placed on non-selective culture medium and allowed to grow for 3 weeks before scoring for stable colonies. After this 3-week period, 10 multicellular, stably-expressing blue colonies were observed, in comparison to only one red colony. This demonstrated that PEI-precipitation could be used to effectively introduce DNA for transient expression while dramatically reducing integration of the PEI-introduced DNA and thus reducing the recovery of RFP-expressing transgenic events. In this manner, PEI-precipitation can be used to deliver transient expression of BBM and/or WUS2.

For example, the particles are first coated with UBI::BBM::pinII using PEI, then coated with UBI::moPAT~YFP using TFX-50, and then bombarded into scutellar cells on the surface of immature embryos. PEI-mediated precipitation results in a high frequency of transiently expressing cells on the surface of the immature embryo and extremely low frequencies of recovery of stable transformants (relative to the TFX-50 method). Thus, it is expected that the PEI-precipitated BBM cassette expresses transiently and stimulates a burst of embryogenic growth on the bombarded surface of the tissue (i.e. the scutellar surface), but this plasmid will not integrate. The PAT~GFP plasmid released from the Ca++/gold particles is expected to integrate and express the selectable marker at a frequency that results in substantially improved recovery of transgenic events. As a control treatment, PEI-precipitated particles containing a UBI::GUS::pinII (instead of BBM) are mixed with the PAT~GFP/Ca++ particles. Immature embryos from both treatments are moved onto culture medium containing 3mg/l bialaphos. After 6-8 weeks, it is expected that GFP+, bialaphos-resistant calli will be observed in the PEI/BBM treatment at a much higher frequency relative to the control treatment (PEI/GUS).

As an alternative method, the BBM plasmid is precipitated onto gold particles with PEI, and then introduced into scutellar cells on the surface of immature embryos, and subsequent transient expression of the BBM gene elicits a rapid proliferation of embryogenic growth. During this period of induced growth, the explants are treated with Agrobacterium using standard methods for maize (see Example 1), with T-DNA delivery into the cell introducing a transgenic expression cassette such as UBI::moPAT~GFPm::pinII. After co-cultivation, explants are allowed to recover on

normal culture medium, and then are moved onto culture medium containing 3 mg/l bialaphos. After 6-8 weeks, it is expected that GFP+, bialaphos-resistant calli will be observed in the PEI/BBM treatment at a much higher frequency relative to the control treatment (PEI/GUS).

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

EXAMPLE 4

Meganuclease generated recombinant target loci (RTL) for FLP/FRT site specific integration (SSI) in maize.

A. Introducing FRT sites at a specific maize genomic locus.

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

Plasmid PHP36070 (SEQ ID NO: 14) was used to create a transgenic insertion at the LIG3-4 locus. The LIG3-4 locus is described in U.S. patent application 12/147,834 filed on June 27, 2008 which is herein incorporated by

reference in its entirety. In short, an endogenous maize genomic sequence was selected for design of a custom double-strand break inducing agent derived from I-CreI meganuclease.

The LIG3-4 recognition site is a 22 bp polynucleotide having the following sequence:

(SEQ ID NO: 16) ATATACCTCACAC▼GTACGCGTA

The double strand break site and overhang region is shown in bold, the enzyme cuts after C13, as indicated by the solid triangle. The I-CreI meganuclease was modified to produce the LIG3-4 meganuclease designed to recognize the LIG3-4 recognition sequence as described in US patent application 12/147,834 filed on June 27, 2008. PHP36070 contains the LIG3-4 meganuclease plant optimized DNA sequence transcribed by the maize ubiquitin promoter (Figure 2). Homologous region 1 (HR1, SEQ ID NO: 17) and homologous region 2 (HR2, SEQ ID NO: 18) are the regions of maize homology that flank LIG3-4 recognition site. A marker cassette consisting of the maize ubiquitin promoter (UBI promoter) driving the herbicide resistance gene mopat (MO-PAT, Figure 2) and a potato protease inhibitor II (PINII) terminator was located between the two HR regions. MoPAT encodes a phosphinothricin acetyltransferase. The marker cassette also included two non-identical FRT sites, FRT1 and FRT87, to allow for site specific integration mediated by FLP recombinase. The positioning of the FRT sites between the promoter and mopat gene and following the PINII has been described in US patent 7,820,880 (filed on November 16, 2008 and issued on October 26, 2010) and incorporated by reference herein. Agrobacterium borders are shown as RB and LB in Figure 2, however this construct was not introduced via Agrobacterium. Instead, PHP36070 was introduced by particle bombardment and co-bombarded with additional plasmids, PHP21875 (UBI PRO-bbm) and PHP21139 (IN2-2 PRO-wus), expressing the maize bbm gene and maize wuschel gene, respectively to enhance the frequency of recovery of transgenic insertions (as described in Example 2 and 3, PCT/US2010/062531).

Plasmid PHP44779 (SEQ ID NO: 19) was used to create a transgenic insertion at the MHP locus on maize chromosome 1.

The MHP locus is described in U.S. patent application 61/499,443 filed on June 21, 2011 which is herein incorporated by reference in its entirety. In short, an endogenous maize genomic target recognition sequence was selected for design of

a custom double-strand break inducing agent derived from I-CreI meganuclease. The MHP14 recognition site is a 22 bp polynucleotide having the following sequence: (SEQ ID NO: 20): caaacagattcacgtcagattt.

PHP44779 (Figure 2) contains the MHP14 meganuclease plant optimized DNA sequence (SEQ ID NO 21) transcribed by the maize ubiquitin promoter Homologous region 1 (also referred to as ZM-Seq8 MHP14-1 in Figure 3, SEQ ID NO: 22) and homologous region 2 (also referred to as ZM-Seq9 MHP14-2 in Figure 3, SEQ ID NO: 23) are the regions of maize homology that flank the MHP14 recognition site. A marker cassette consisting of the maize ubiquitin promoter (UBI promoter) driving the herbicide resistance gene mopat and a potato protease inhibitor II (PINII) terminator was located between the two HR regions (Figure 3). PHP44779 was introduced by particle bombardment and co-bombarded with the additional plasmids, PHP44779 and PHP31729 (OLE PRO-bbm) and PHP21139 (IN2-2 PRO-wus), expressing the maize bbm gene and maize wuschel gene, respectively, to enhance the frequency of recovery of transgenic insertions (Example 2 and 3, PCT/US2010/062531).

Transgenic callus bombarded with either PHP36070 or PHP44779 and showing resistance to bialaphos was regenerated into plants. Leaf tissue of transgenic plants was used for molecular analysis to confirm that the insertion or recombination occurred at the specific target sites for each meganuclease. The desired molecular configuration of the transgenic insert of these experiments is detailed in Figure 4. Figure 4 shows a graphic representation of the transgenic insertion site following homologous recombination mediated by meganuclease induced double stranded breaks. The meganuclease can cut the maize genome at a specific target sequence leaving a double stranded break which subsequently promotes the cell's DNA recombination and repair mechanisms. A copy of the plasmid with homologous regions is in the vicinity and gets recombined into the genome creating a transgenic insertion site via homologous recombination. The recombination exchange places a fragment of the transformation plasmid in the insertion site which contains only the marker cassette with associated FLP recombinase sites. Ideally, the HR regions of the construct have inserted in a way that they are seamless with the existing genome sequence of the target

chromosome. The meganuclease gene is left behind on the construct and degraded.

PCR and Southern analysis were used to obtain information about the molecular structure of the transgene insertions. The desired molecular configurations containing the marker gene cassette with associated FLP/FRT SSI features (Figure 4) were obtained as a small percentage of the total number of insertion events.

B. Characterization of recombinant target loci (RTL) for FLP/FRT site specific integration at a specific LIG3-4 or MHP maize genomic site.

Transgenic plants regenerated form callus bombarded with either PHP36070 or PHP44779 were grown in the greenhouse to maturity and seed was harvested. The next generation seed from self pollinations was used to obtain a homozygous seed supply, by using quantitative PCR (QPCR) to screen populations from a self pollination of either the first or second generation. Homozygous individuals identified by QPCR were self-pollinated to increase homozygous seed supply and were carried on to non-transgenic plants to provide a supply of immature embryos for FLP/FRT SSI transformation experiments. Large numbers of immature embryos heterozygous for the transgenic insertion at the LIG3-4 or MHP site were obtained for these experiments by carrying pollen from populations of homozygous LIG3-4 or MHP plants to populations of non-transgenic plants of the same corn genotype supplied for embryo source.

Transgenic plants containing the desired insertions (Figure 4) from meganuclease facilitated homologous recombination were identified and became recombinant target loci (RTL) for FLP/FRT site specific integration.

Methods for site specific integration using FLP recombinase include the combination of a 'target' transgenic locus, also known as an Recombinant Trait Locus (RTL), with two non-identical FRT sites and a 'donor' plasmid or insert with the same two non-identical FRT sites (US patent 7,462,766 filed May 4, 2006 and issued December 09, 2008). FLP recombinase binds to the FRT sites of both target and donor, bring the FRT sites together in the cell, and then recombine FRT sites of identical sequence.

A LIG-3-4 event (E8815.112.3.28) and MHP14 events containing the desired configuration and non-identical FRT sites (Figure 4) were identified.

C. Recombinase Medicated DNA Casette exchange in SSI sites created at specific genomic loci.

Maize transformation of events containing the recombinant target loci (RTL) (described in Example 4 B) for FLP/FRT SSI at a specific LIG3-4 or MHP locus was accomplished by particle bombardment (also referred herein as SSI transformation). For each experiment, several plasmids were co-bombarded including the Transfer Cassette plasmids PHP27064 (SEQ ID NO:24) or PHP44951(SEQ ID NO:25)(Figure 5), a plasmid to transiently express FLP recombinase (PHP5096, UBI PRO-flp), and plasmids to transiently express bbm and wus (PHP31729, PHP21139).

Two independent Transfer Cassette plasmids (PHP27064 (SEQ ID NO:24) or PHP44951(SEQ ID NO:25) were used to test the effect of molecule size on transformation frequency at the LIG3-4 and MHP14 RTL, as well as to introduce multiple genes that allow the measurement of expression levels in the case of PHP44951 **(Figure 5).** Both transfer cassettes includes an FRT1 upstream of the first gene and FRT87 at the 3' end of the terminator for the last gene. This allows functional recombination with the target sites that have FRT sites in similar positions. The upstream gene in the donor does not have its own promoter because it is activated by the ubiquitin promoter in the target site following recombination. The integration of the transfer cassettes at the MHP14 site after FLP recombinase mediated site specific integration is shown in Figure 6.

The SSI process involves recombination between the FRT1 and FRT87 sites of the target and transfer cassette in a process called double reciprocal crossover. The end result is that the mopat gene (Figure 4) is replaced at the SSI target by gat4621 and zs-yellow1 N1 when donor PHP27064 is bombarded or those genes plus gus and mopat if PHP44951 is bombarded.

Table 3 shows QPCR results from a number of independent FLP/FRT SSI transformation experiments involving a relatively large number of treated immature embryos. RMCE events were obtained from the LIG3-4 RTL. Table 3 also illustrates that the RMCE frequency may be affected by the size of the transfer cassette. A

higher RMCE frequency was observed for the smaller transfer cassette PHP27064 when compared to the larger PHP44951.

Table 3: RMCE frequencies at the LIG3-4 target site. The LIG3/4 RTL was obtained from meganuclease assisted homologous recombination.

RTL	Transfer Cassette	Embryo TRT	# RMCE (T0 plants)	RMCE FREQ	YFP expression in callus	Regeneration
LIG3/4 E8815.112.3.28	PHP27064	2696	17	1.10	Bright	Good
LIG3/4 E8815.112.3.28	PHP44951	2693	10	0.69	Bright	OK

Embryo TRT = the number of immature embryos treated for SSI transformation; RMCE = recombinase mediated cassette exchange is the term given to the desired molecular result from FLP/FRT SSI transformation; RMCE FREQ = the transformation frequency on a per embryo basis of obtaining RMCE events.

Similar FLP/FRT SSI transformation data was generated the MHP14 target site as shown in Table 4.

Table 4: RMCE frequencies at the MHP14 target site. The MHP14 RTL was obtained from meganuclease assisted homologous recombination.

RTL	Transfer Cassette	Embryo TRT	# RMCE (T0 plants)	RMCE FREQ
MHP14	PHP27064	3753	19	0.51
MHP14	PHP44951	3696	2	0.05

Table 4 indicates that RMCE events were obtained from the MHP14 SSI target site RTL. As with the Lig3-4 site, a higher RMCE frequency was observed at the MHP14 site with the smaller transfer cassette PHP27064 when compared to PHP44951.

Table 3 and Table 4 clearly demonstrate that we can obtain targeted integration of a target site for site-specific recombination at a double strand break

induced by a meganuclease at an endogenous recognition site and subsequent modify the integrated target site by RMCE.

EXAMPLE 5

Creation of FRT1/FRT87 sites in the TS14 target site in soybean genome by soybean TS14 meganuclease

Soybean lines comprising an endogenous recognition sequence in their genome were contacted with an engineered meganuclease derived from I-CreI designed to specifically recognize and create a double-strand break in the endogenous recognition sequence. Immature embryos comprising an endogenous recognition site were contacted with the components described below, events selected and characterized.

In order to introduce the FRT1/FRT87 sites into the TS14 recognition site (SEQ ID NO:2) in the soybean genome, two expression cassettes (RTW347 and RTW365) were used. RTW347 (SEQ ID NO: 27) contains the plant optimized DNA sequence encoding the TS14 meganuclease (SEQ ID NO:26) driven by the soybean UBQ promoter and PinII terminator. RTW365 (SEQ ID NO:28) is the Transfer Cassette construct in which the FRT1/FRT87 sites are flanking the Gm-HRA::Gm-ALS terminator. The Gm-HRA gene was driven by the soybean SAMS promoter. The SAMS promoter::FRT1::Gm-HRA::Gm-ALS Terminator::FRT87 cassette is flanked by the 1000 bp homologous region1 (SEQ ID NO: 35) and 928bp region2 (SEQ ID NO:36) flanking the TS14 recognition site. The TS14 recognition site sequence, the plant-optimized nucleotide sequence of the TS14 meganuclease homologous region1 and region2 have been disclosed in US provisional application 61/466602 filed on March 23, 2011 which is herein incorporated by reference in their entirety.

RTW347 and RTW365 were co-bombarded into soybean cells using standard soybean transformation methods. qPCR and genomic PCR were used to identify the transgene integration event containing the SAMS promoter::FRT1::GM-HRA::GM-ALS Terminator::FRT87. The qPCR assay specific to the TS14 target sequence was developed to identify sequence changes in the region. The primers (Mega14-13F, Mega14-128R and probe Mega14-85T), were used to identify the transgenic events with the TS14 target sites reduced to 1 or 0 copy as compared to the 2 copies in the wild type soybean genome. The border specific genomic PCR

assays were used to further identify the transgene integration event. For example, the primer set WOL192 (SEQ ID NO:29) and WOL311 (SEQ ID NO:30) were designed and used to amplify the left border integration. The WOL192 is a sequence specific primer located in soybean genome 5' beyond the TS14 HR1 region and the WOL311 is a sequence specific primer to the 5' SAMS promoter in the reverse orientation. A 1334 bp PCR product (SEQ ID NO:31) can only be obtained when the RTW365 repair DNA get integrated by homologous integration enable by TS14 meganuclease. Another set of primer WOL312 (SEQ ID NO:32) and WOL193 (SEQ ID NO:33) were also designed and used to amplify the right border integration. The WOL312 is the sense primer from the GM-ALS terminator and the WOL193 is a sequence specific primer located in soybean genome 3' beyond the TS14 HR2 region. A 1620 bp PCR product (SEQ ID NO:34) can only be obtained when the RTW365 repair DNA get integrated by homologous integration enable by TS14 meganuclease. For the TS14 target site, 18 qPCR positive events were identified from total 68 events by qPCR analyses. Out of the 18 qPCR positive events, three events were confirmed to be perfect TS14 meganuclease mediated SAMS promoter::FRT1::GM-HRA::GM-ALS Terminator::FRT87 transgene integration events by homologous recombination. The Introduction of the FRT1 and FRT87 sites in the soybean TS14 target provided the ability to use the FLP/FRT technology to perform gene stacking by the SSI technology.

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 introducing into the genome of a plant cell a target site for site-specific integration, the method comprising:

(a) providing a plant cell comprising in its genome an endogenous recognition site for an engineered endonuclease, wherein the engineered endonuclease is capable of inducing a double-strand break in said endogenous recognition site, and wherein the endogenous recognition site is located between a first and a second genomic region ;

(b) providing a donor DNA comprising the target site for site-specific integration located between a first region of homology to said first genomic region and a second region of homology to said second genomic region, wherein the target site comprises a first and a second recombination site, wherein the first and the second recombination sites are dissimilar and non-recombinogenic with respect to one another;

(c) contacting the plant cell with the donor DNA and the engineered endonuclease, and

(d) identifying at least one plant cell from (c) comprising in its genome the target site integrated at the endogenous recognition site.

2. The method of claim 1, wherein the first region of homology further comprises a first fragment of said endogenous recognition site of (a), and wherein the second region of homology comprises a second fragment of said endogenous recognition site of (a), wherein the first and second fragments are dissimilar.

3. The method of claim 1, wherein the first region of homology further comprises the first 13 bases of said endogenous recognition site of (a), and wherein the second region of homology comprises the last 9 bases of said endogenous recognition site of (a).

4. The method of any one of claims 1-3, further comprising recovering a fertile plant from the cell of (d), the fertile plant comprising in its genome the target site integrated into the endogenous recognition site.

5. The method of any one of claims 1-4, wherein the endogenous recognition site is selected from the group consisting of SEQ ID NO:1-8, 16 and 20 or a sequence having at least 90% sequence identity to SEQ ID NO:1-8, 16 and 20.

6. The method of any one of claims 1-5, wherein the target site further comprises a polynucleotide of interest between the first recombination site and the second recombination site.

7. The method of any one of claims 1-6, wherein at least one of the first and the second recombination sites comprises an FRT site, a mutant FRT site, a LOX site, and a mutant LOX site.

8. The method of any one of claims 1-7, wherein the target site further comprises a third recombination site between the first and the second recombination site, wherein the third recombination site is dissimilar and non-recombinogenic to the first and the second recombination sites.

9. The method of claim 8, wherein at least one of the first, the second, and the third recombination sites comprises FRT1 (SEQ ID NO: 9), FRT 5 (SEQ ID NO: 10), FRT6 (SEQ ID NO: 11), FRT12 (SEQ ID NO: 12) and FRT87 (SEQ ID NO: 13).

10. The method of claim 8, wherein the first recombination site is FRT1 (SEQ ID NO: 9), the second recombination site is FRT12 (SEQ ID NO: 12), and the third recombination site is FRT87 (SEQ ID NO: 13).

11. The method of any one of claims 1-10, wherein the engineered endonuclease is derived from I-CreI.

12. The method of any one of claims 1-11, wherein said plant cell is from a monocot.

13. The method of claim 12, wherein said monocot is maize, wheat, rice, barley, sorghum, or rye.

14. The method of any one of claims 1-11, wherein said plant cell is from a dicot.

15. The method of claim 14, wherein said dicot is soybean, Brassica, sunflower, cotton, or alfalfa.

16. A plant cell, plant part, plant, or seed comprising the target site integrated at the endogenous recognition site according to any one of claims 1-15.

17. A nucleic acid molecule comprising the target site integrated at the endogenous recognition site according to any one of claims 1-15.

18. A plant cell, plant part, plant, or seed comprising the nucleic acid molecule of claim 17.

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

(a) providing at least one plant cell comprising in its genome a target site for site-specific integration, wherein the target site is integrated into an endogenous recognition site for an engineered endonuclease, and wherein the target site is,

(i) a target site comprising a first and a second recombination site; or

(ii) the target site of (i) further comprising a third recombination site between the first recombination site and the second recombination site,

wherein the engineered endonuclease is capable of inducing a double-strand break in the endogenous recognition site, wherein the first, the second, and the third recombination sites are dissimilar and non-recombinogenic with respect to one another,

(b) introducing into the plant cell of (a) a transfer cassette comprising,

(iii) the first recombination site, a first polynucleotide of interest, and the second recombination site,

(iv) the second recombination site, a second polynucleotide of interest, and the third recombination sites, or

(v) the first recombination site, a third polynucleotide of interest, and the third recombination sites;

(c) providing a recombinase that recognizes and implements recombination at the first and the second recombination sites, at the second and the third recombination sites, or at the first and third recombination sites; and

(d) selecting at least one plant cell comprising integration of the transfer cassette at the target site.

20. The method of claim 19, further comprising recovering a fertile plant from the plant cell of (d), the fertile plant comprising in its genome the transfer cassette at the target site.

21. The method of any one of any one of claims 19-20, wherein at least one of the first, the second, and the third polynucleotides of interest comprises a nucleotide sequence for gene silencing, a nucleotide sequence encoding a phenotypic marker, or a nucleotide sequence encoding a protein providing an agronomic advantage.

22. The method of any one of claims 19-21, wherein providing the recombinase comprises integrating into the genome of the plant cell a nucleotide sequence encoding the recombinase.

23. The method of any one of claim 19-22, wherein the transfer cassette further comprises at least one coding region operably linked to a promoter that drives expression in the plant cell.

24. The method of any one of claims 19-23, wherein the transfer cassette further comprises a coding region operably linked to a promoter that drives expression in the plant cell, wherein the coding region encodes a recombinase that

facilitates recombination between, the first and the second recombination sites of the transfer cassette and the target site, the second and the third recombination sites of the transfer cassette and the target site, or the first and the third recombination sites of the transfer cassette and the target site.

25. The method of any one of claim 19-24, wherein at least one of the first, the second, and the third recombination sites comprises an FRT site, a mutant FRT site, a LOX site, or a mutant LOX site.

26. The method of any one of claims 19-24, wherein the first recombination site is FRT1 (SEQ ID NO: 9), the second recombination site is FRT12 (SEQ ID NO: 12), and the third recombination site is FRT87 (SEQ ID NO: 13)

27. The method of any one of claims 19-26, wherein the recombinase is FLP.

28. The method of claim 27, wherein the FLP has been synthesized using maize-preferred codons.

29. The method of any one of claims 19-28, wherein said plant cell is from a monocot.

30. The method of claim 29, wherein said monocot is maize, wheat, rice, barley, sorghum, or rye.

31. The method of any one of claims 19-28 wherein said plant cell is from a dicot.

32. The method of claim 31, wherein said dicot is soybean, Brassica, sunflower, cotton, or alfalfa.

33. A plant cell, plant part, plant, or seed comprising the transfer cassette integrated at the target site according to any one of claims 19-32.

34. A plant, seed or plant cell comprising in its genome a target site for site-specific integration, wherein the target site is integrated into an endogenous recognition site for an engineered endonuclease, wherein the target site comprises in the following order:

(a) a first recombination site;

(b) a second recombination site, and

wherein the engineered endonuclease is capable of inducing a double-strand break at the endogenous recognition site, wherein the first and the second recombination sites are dissimilar and non-recombinogenic with respect to one another.

35. The plant, seed or plant cell of claim 34, wherein the endogenous recognition site is selected from the group consisting of SEQ ID NO:1-8, 16 and 20 or a sequence having at least 90% sequence identity to SEQ ID NO:1-8, 16 and 20.

36. The plant, seed or plant cell of any of claims 35-36, wherein the target site further comprises a polynucleotide of interest between the first recombination site and the second recombination site.

37. The plant, seed or plant cell of any one of claims 35-36, wherein at least one of the first and the second recombination sites comprises an FRT site, a mutant FRT site, a LOX site, or a mutant LOX site.

38. The plant, seed or plant cell of any one of claims 35-37, wherein the target site further comprises a third recombination site between the first and the second recombination site, wherein the third recombination site is dissimilar and non-recombinogenic to the first and the second recombination sites.

39. The plant, seed or plant cell of claim 38, wherein at least one of the first, the second, and the third recombination sites comprises FRT1 (SEQ ID NO: 9), FRT 5 (SEQ ID NO: 10), FRT6 (SEQ ID NO: 11), FRT12 (SEQ ID NO: 12) and FRT87 (SEQ ID NO: 13).

40. The plant, seed or plant cell of claim 38, wherein the first recombination site is FRT1 (SEQ ID NO: 9), the second recombination site is FRT12 (SEQ ID NO: 12), and the third recombination site is FRT87 (SEQ ID NO: 13).

41. The plant, seed, or plant cell of any one of claims 35-40, wherein said plant, seed or plant cell is from a monocot.

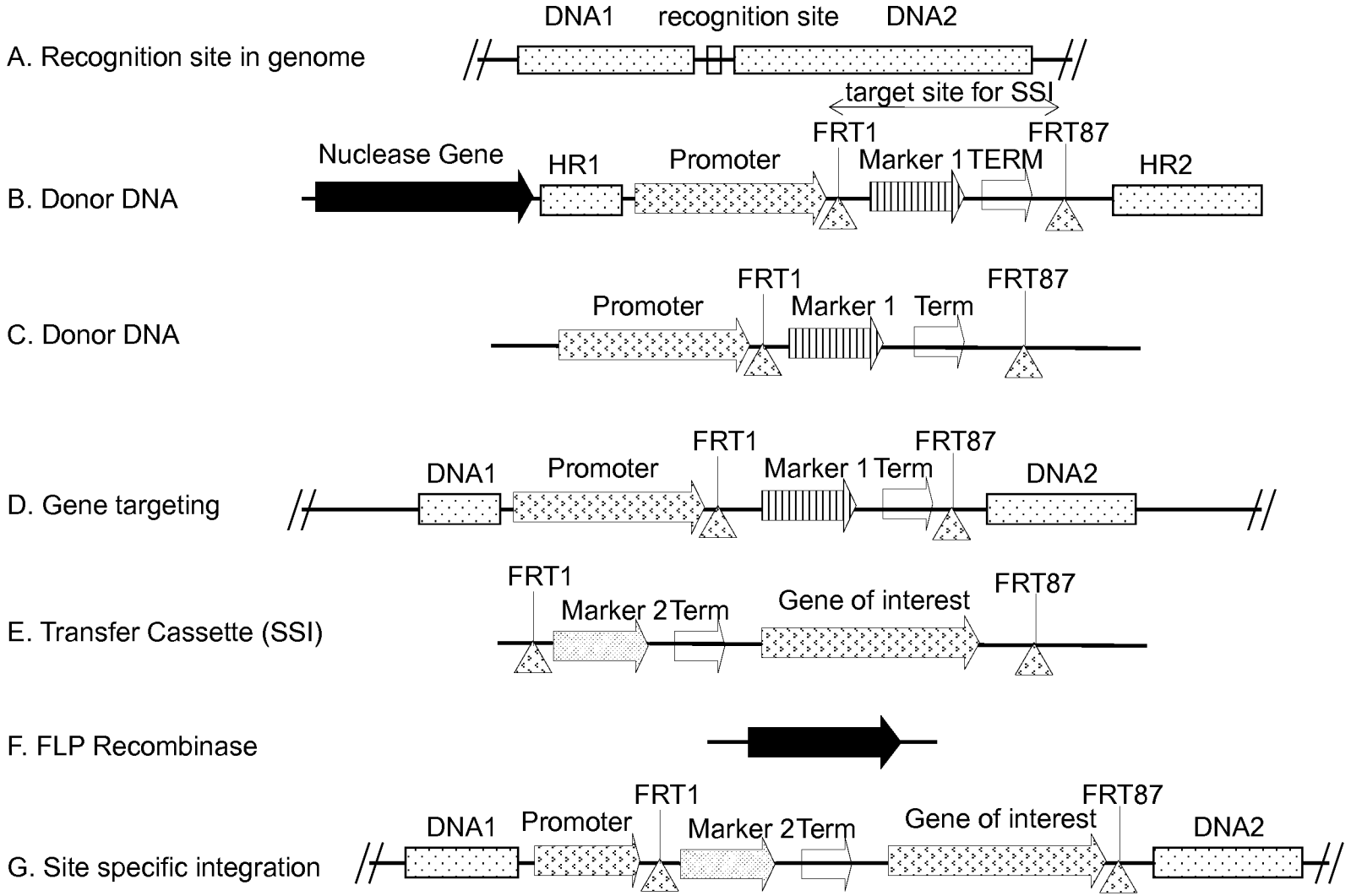
42. The plant, seed, or plant cell of claim 41, wherein said monocot is maize, wheat, rice, barley, sorghum, or rye.

43. The plant, seed, or plant cell of any one of claims 35-40, wherein said plant, seed or plant cell is from a dicot.

44. The plant, seed, or plant cell of claim 43, wherein said dicot is soybean, Brassica, sunflower, cotton, or alfalfa.

Figure 1

DNA Double-Strand-Break-Induced DNA Alteration of an Endogenous Target Site Followed by FLP Recombinase Mediated Site Specific Integration



WO 2013/019411 PCT/US2012/047202

Figure 2

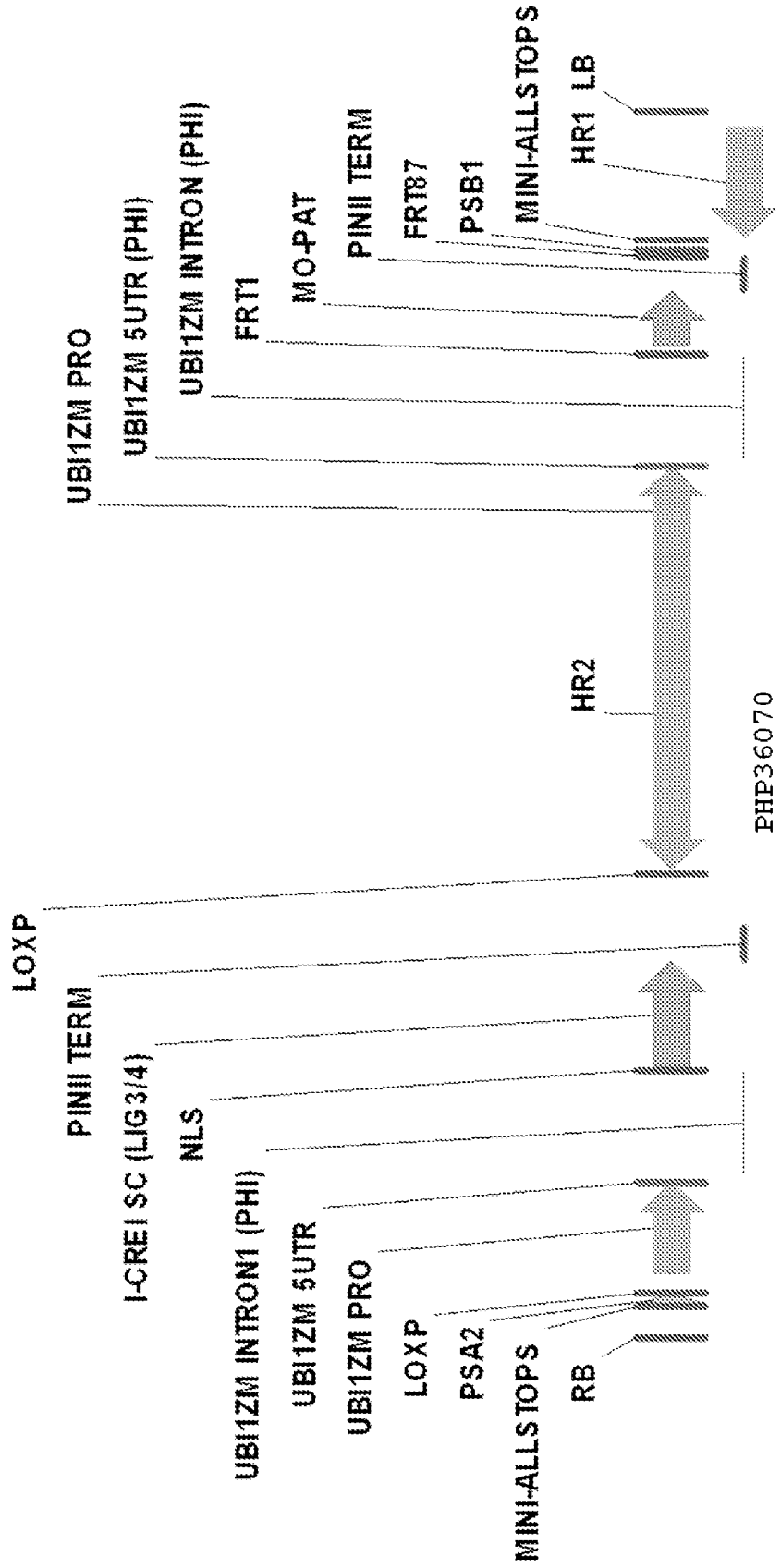

LOXP
PINII TERM
I-CREI SC (LIG3/4)
NLS
UBI1ZM INTRON1 (PHI)
UBI1ZM 5UTR
UBI1ZM PRO
LOXP
PSA2
MINI-ALLSTOPS
RB
HR2
UBI1ZM PRO
UBI1ZM 5UTR (PHI)
UBI1ZM INTRON (PHI)
FRT1
MO-PAT
PINII TERM
FRT87
PSB1
MINI-ALLSTOPS
HR1 LB
PHP36070

Figure 3

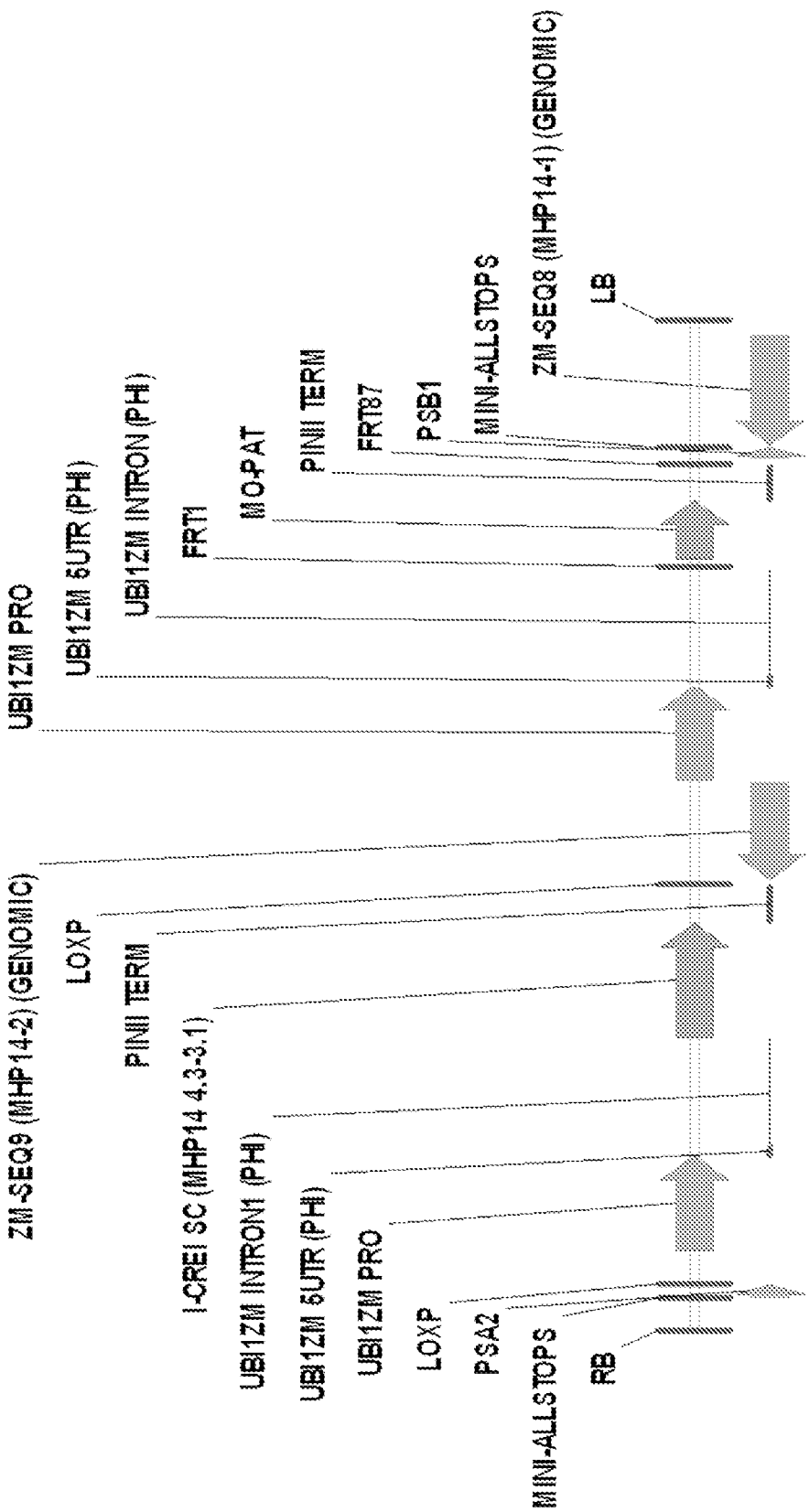

ZM-SEQ9 (MHP14-2) (GENOMIC)
LOXP
PINII TERM
I-CREI SC (MHP14 4.3-3.1)
UBI1ZM INTRON1 (PHI)
UBI1ZM 5UTR (PHI)
UBI1ZM PRO
LOXP
PSA2
MINI-ALLSTOPS
RB
UBI1ZM PRO
UBI1ZM 5UTR (PHI)
UBI1ZM INTRON (PHI)
FRT1
MO-PAT
PINII TERM
FRT87
PSB1
MINI-ALLSTOPS
ZM-SEQ8 (MHP14-1) (GENOMIC)
LB
PHP44779

Figure 4

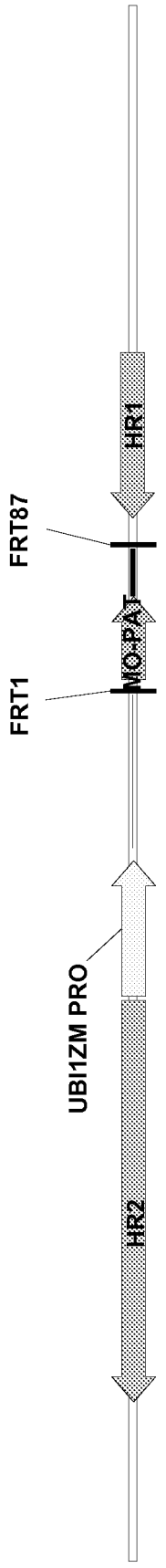


Figure 5

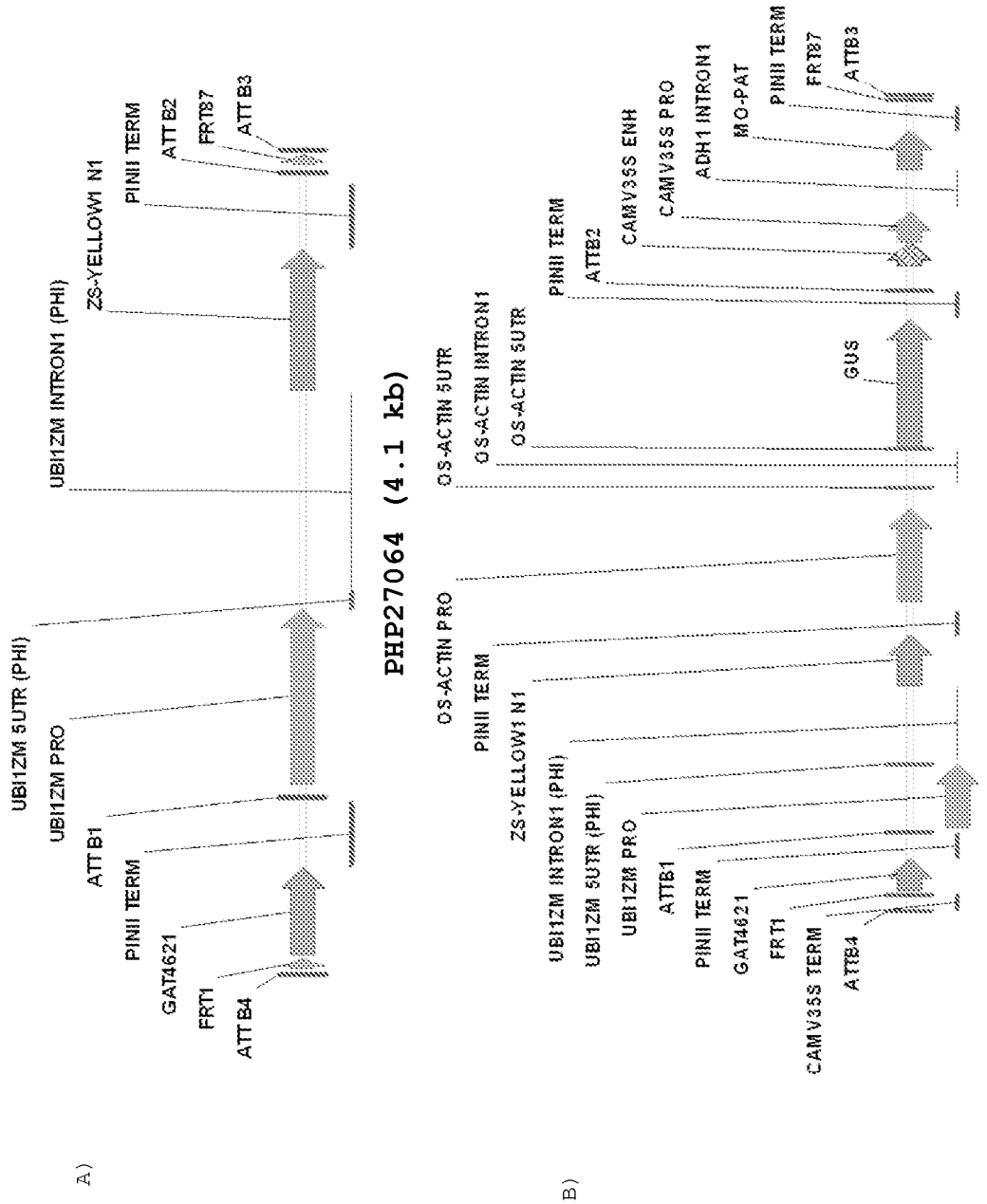
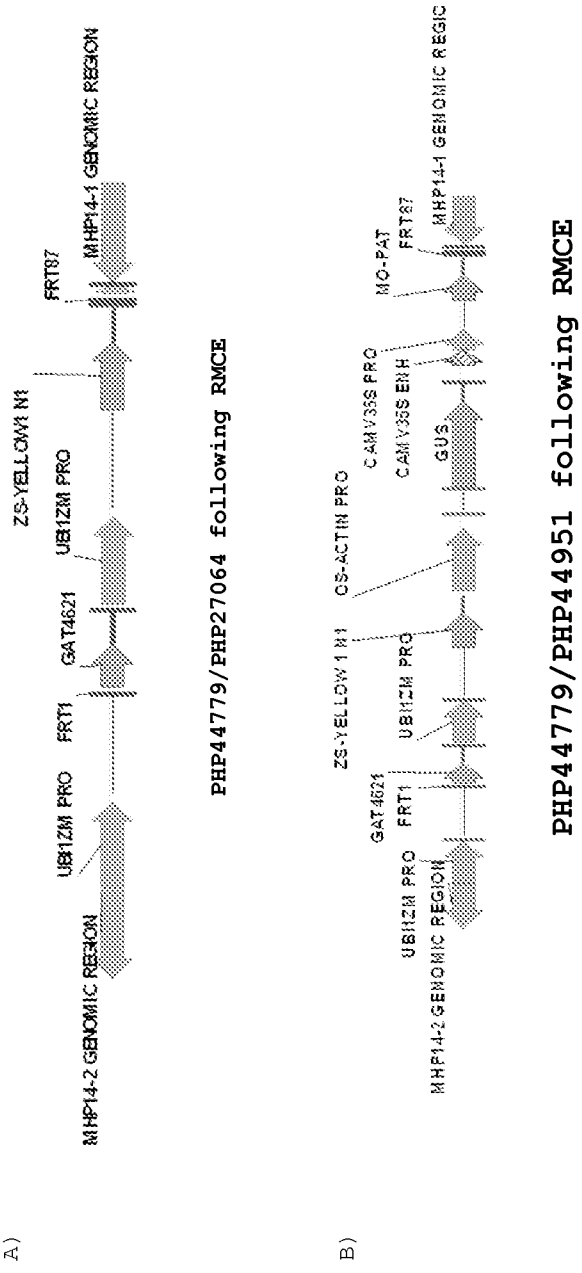

A)
UBI1ZM 5UTR (PHI)
UBI1ZM PRO
ATT B1
PINII TERM
GAT4621
FRT1
ATT B4
UBI1ZM INTRON1 (PHI)
ZS-YELLOW1 N1
PINII TERM
ATT B2
FRT87
ATT B3
PHP27064 (4.1 kb)
B)
OS-ACTIN PRO
PINII TERM
ZS-YELLOW1 N1
UBI1ZM INTRON1 (PHI)
UBI1ZM 5UTR (PHI)
UBI1ZM PRO
ATTB1
PINII TERM
GAT4621
FRT1
CAMV35S TERM
ATTB4
OS-ACTIN 5UTR
OS-ACTIN INTRON1
OS-ACTIN 5UTR
PINII TERM
ATTB2
CAMV35S ENH
CAMV35S PRO
ADH1 INTRON1
MO-PAT
PINII TERM
FRT87
ATTB3
GUS
PHP44951 (11.2 kb

Figure 6


A)
MHP14-2 GENOMIC REGION
UBI1ZM PRO
FRT1
GAT4621
UBI1ZM PRO
ZS-YELLOW1 N1
FRT87
MHP14-1 GENOMIC REGION
PHP44779/PHP27064 following RMCE
B)
MHP14-2 GENOMIC REGION
UBI1ZM PRO
FRT1
GAT4621
UBI1ZM PRO
ZS-YELLOW1 N1
CS-ACTIN PRO
CAMV35S ENH
CAMV35S PRO
GUS
MO-PAT
FRT87
MHP14-1 GENOMIC REGIC
PHP44779/PHP44951 following RMCE

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/047202

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/82 A01H5/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2007/011733 A2 (PIONEER HI BRED INT [US]; TAO YUMIN [US]; BIDNEY DENNIS [US]; GORDON-K) 25 January 2007 (2007-01-25)	33-44
Y	the whole document	1-32

Y	WO 2009/006297 A2 (PIONEER HI BRED INT [US]; LYZNIK L ALEKSANDER [US]; TAO YUMIN [US]; GA) 8 January 2009 (2009-01-08) the whole document	1-32

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 October 2012	19/10/2012

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Maddox, Andrew

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2012/047202

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2007011733 A2	25-01-2007	BR PI0612862 A2	30-11-2010
		CA 2615797 A1	25-01-2007
		EP 1907553 A2	09-04-2008
		US 2007015195 A1	18-01-2007
		US 2008047031 A1	21-02-2008
		US 2010173801 A1	08-07-2010
		US 2010192263 A1	29-07-2010
		US 2011047655 A1	24-02-2011
		WO 2007011733 A2	25-01-2007
WO 2009006297 A2	08-01-2009	CA 2691440 A1	08-01-2009
		CN 101849010 A	29-09-2010
		EP 2167666 A2	31-03-2010
		US 2009133152 A1	21-05-2009
		WO 2009006297 A2	08-01-2009