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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 June 2016 (30.06.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/106121 A1

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

(21) International Application Number:
PCT/US2015/066619

(22) International Filing Date:
18 December 2015 (18.12.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/096,442 23 December 2014 (23.12.2014) US

(71) Applicant: **SYNGENTA PARTICIPATIONS AG** [CH/CH]; Schwarzwaldallee 215, 4058 Basel (CH).

(71) Applicant *(for US only):* **SYNGENTA CROP PROTECTION, LLC** [US/US]; 3054 E. Cornwallis Road, Durham, North Carolina 27709 (US).

(72) Inventors: **CHEN, Zhongying;** 3054 E. Cornwallis Road, Durham, North Carolina 27709 (US). **KIM, Myoung;** 3054 E. Cornwallis Road, Durham, North Carolina 27709 (US). **ZHONG, Heng;** 3054 E. Cornwallis Road, Durham, North Carolina 27709 (US). **GU, Weining;** 3054 E. Cornwallis Road, Durham, North Carolina 27709 (US). **JIANG, Yaping;** 3054 E. Cornwallis Road, Durham, North Carolina 27709 (US). **QUE, Qiudeng;** 3054 E. Cornwallis Road, Durham, North Carolina 27709 (US). **CHILTON, Mary-Dell;** 3054 E. Cornwallis Road, Durham, North Carolina 27709 (US).

(74) Agent: **BARRON, Yoshimi D.;** Syngenta Crop Protection, LLC, 3054 E. Cornwallis Road, Durham, North Carolina 27709 (US).

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

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

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

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

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

WO 2016/106121 A1

(54) Title: METHODS AND COMPOSITIONS FOR IDENTIFYING AND ENRICHING FOR CELLS COMPRISING SITE SPECIFIC GENOMIC MODIFICATIONS

(57) Abstract: The present invention relates to methods and compositions for modifying a target site in the genome of a plant cell. Such modifications include integration of a transgene and mutations. The present invention also relates to methods and compositions for identifying and enriching for cells which comprise the modified target site.

METHODS AND COMPOSITIONS FOR IDENTIFYING AND ENRICHING FOR CELLS COMPRISING SITE SPECIFIC GENOMIC MODIFICATIONS

RELATED APPLICATIONS

This application claims the benefit of provisional application 62/096,442 filed December 23, 2014 and incorporated by reference in its entirety herein.

SEQUENCE LISTING

A Sequence Listing in ASCII text format, submitted under 37 C.F.R. § 1.821, entitled "80484_ST25.txt", 409 kilobytes in size, generated on December 15, 2015 and filed via EFS-Web is provided in lieu of a paper copy. This Sequence Listing is hereby incorporated by reference into the specification for its disclosures.

FIELD OF THE INVENTION

The present invention relates to methods and compositions for modifying a target site in the genome of a plant cell. Such modifications include transgene integration and mutations. The present invention further relates to methods and compositions for identifying and enriching for a cell with one or more transgenes integrated at a target site within the genome of the cell, as well as for identifying and enriching for a cell comprising a mutation introduced at a target site within the genome of the cell without integration into the genome of a heterologous nucleotide sequence encoding a nuclease for site specific cleavage at the target site within the genome.

BACKGROUND OF THE INVENTION

Recent advances in the field of targeted modifications of a genome have made is so that routine targeted modifications may soon be possible. Significant advances have been made in the last few years towards the development of methods and compositions to target and cleave genomic DNA by site specific nucleases (e.g., Zinc Finger Nucleases (ZFNs), Meganucleases, Transcription Activator-Like Effector Nucelases (TALENS) and Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated nuclease (CRISPR/Cas) with an engineered crRNA/tracr RNA), to induce targeted mutagenesis, induce targeted deletions of cellular DNA sequences, and facilitate targeted recombination of an exogenous donor DNA polynucleotide, such as a transgene, within a predetermined genomic locus. This predetermined genomic locus is not obvious. Many sites in the genome

are non-ideal for, for example, transgene insertion, due to highly repetitive nucleotide sequence, methylation, and other characteristics that result in a very high or very low level of recombination or poor expression of genes on introduced transgenes. Therefore, there is a need in the art to identify ideal target sites within a genome for targeted modifications, such as transgene insertion.

Once a target site has been used for targeted modification, there is a need to determine if the desired targeted modification was successfully created. Existing methods of screening for targeted genomic modifications in cells are primarily based on polymerase chain reaction (PCR) protocols, nucleic acid sequencing and Southern analysis. In the case of PCR amplification, the screening process of handling the complexity of gene insertion or modification at a specific site is inefficient due to the complexity of PCR primer settings and inherent ambiguity of PCR amplification due to the resulting complexity of genome rearrangement and genome ploidy. Some of the problems with PCR include: 1) no clear distinction between one copy and two copy insertions due to ploidy of the genome; 2) a requirement for complex primer design and large sets of primer combinations to deal with the complexity of gene insertion or modification at the specific site(s); and 3) low throughput of gel electrophoresis and ambiguity of amplification bands. Although subsequent sequencing can help in identifying the characteristics of PCR amplification products, there are problems with large scale sequencing efforts and interpretation of results for large sample numbers. Further gene segregation analysis is required to isolate homozygous progeny for further screening. These steps require large scale operations for screening of commercial crops in order to capture less than 2% of potential candidates and the inventory scale of plants in greenhouses require commercial scales of space and operational costs until the plant growth stage is mature enough to carry out Southern analyses.

The present invention addresses these shortcomings in the art by providing an ideal target site for a maize genome. The present invention also provides a more strategic and efficient approach to identify and enrich for cells with a targeted genomic insertion or a targeted genomic mutation, which reduces the number of candidate plants with high accuracy at the very early stages of the screening process, avoiding a large scale sequencing effort and reducing greenhouse operational costs for plant maintenance.

SUMMARY OF THE INVENTION

In one aspect, the present invention provides a method of integrating a transgene into a genomic nuclease cleavage site in a maize genome, comprising introducing into a maize

cell: a) a first nucleic acid molecule comprising at least about 100 contiguous nucleotides, wherein said contiguous nucleotides have at least about 90% identity with a target site in the nucleotide sequence of SEQ ID NO:l or the nucleotide sequence of SEQ ID NO:2, and further comprising a transgene; and b) a second nucleic acid molecule comprising a nucleotide sequence encoding a nuclease for site-directed cleavage at a genomic nuclease cleavage site adjacent to the nucleotide sequence of SEQ ID NO: 1 or the nucleotide sequence of SEQ ID NO:2 that corresponds to the contiguous nucleotides of (a), under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave the nucleotide sequence at the genomic nuclease cleavage site, whereby the transgene is integrated at the genomic nuclease target cleavage site in the maize genome. The present invention also provides a method of producing a maize plant, plant part, or progeny thereof comprising a transgene integrated into a genomic nuclease cleavage site in the maize genome, comprising regenerating a maize plant from the maize cell produced by the method described above. The present invention further provides a maize plant, plant part, or progeny thereof comprising a transgene integrated into a genomic nuclease cleavage site in the maize genome, produced by the method described.

In a further aspect, the present invention provides a method of enriching for a cell comprising a transgene inserted into a nuclease cleavage site in a genome of the cell, comprising: a) introducing into a plurality of cells: i) a first nucleic acid molecule comprising at least 100 contiguous nucleotides, wherein the at least 100 contiguous nucleotides have at least 90% identity with a target site in the genome of the cell, and further comprising a transgene; and ii) a second nucleic acid molecule encoding a nuclease for site-directed cleavage at a nuclease cleavage site in the genome of the cell adjacent to the nucleotide sequence in the genome of the cell that corresponds to the at least 100 contiguous nucleotides of (a), under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave at the nuclease cleavage site in the genome of the cell and integrate the transgene into the nuclease cleavage site in the genome of the cell; b) culturing the cells of (a) to produce a cell line or tissue; c) extracting a genomic DNA sample from the cell line or tissue of (b); d) performing real-time quantitative polymerase chain reaction (qPCR) assays T and G on the sample of (c), wherein the assays T and G respectively comprise the following probes: i) a first probe comprising a nucleotide sequence that is complementary to a nucleotide sequence of the target site, at least five base pairs away from the nuclease cleavage site for carrying out assay T, and ii) a second probe comprising a nucleotide sequence that is complementary to a nucleotide sequence of the transgene for

carrying out assay G; e) obtaining a DNA copy number of the target site from the results of assay T and a DNA copy number of the transgene from the results of assay G; and f) enriching for a cell line or tissue that has reduced copy number in assay T relative to a reference and a copy number greater than zero for assay G, thereby enriching for the cell comprising the transgene inserted into the nuclease cleavage site in the genome of the cell.

Furthermore, the present invention provides a method of identifying a cell comprising a transgene inserted into a nuclease cleavage site in a genome of the cell, comprising: a) introducing into a plurality of cells: i) a first nucleic acid molecule comprising at least 100 contiguous nucleotides having at least 90% identity with a target site in the genome of the cell, and further comprising a transgene; and ii) a second nucleic acid molecule encoding a nuclease for site-directed cleavage at a nuclease cleavage site in the genome of the cell adjacent to the nucleotide sequence in the genome corresponding to the at least 100 contiguous nucleotides of (a), under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave at the nuclease cleavage site in the genome of the cell and integrate the transgene into the nuclease cleavage site in the genome of the cell; b) culturing the cells of (a) to produce a cell line or tissue; c) extracting a genomic DNA sample from the cell line or tissue of (b); d) performing real-time quantitative polymerase chain reaction (qPCR) assays T and G on the sample of (c), wherein the assays T and G respectively comprise the following probes: i) a first probe comprising a nucleotide sequence that is complementary to a nucleotide sequence of the target site, at least five base pairs away from the nuclease cleavage site for carrying out assay T, and ii) a second probe comprising a nucleotide sequence that is complementary to a nucleotide sequence of the transgene for carrying out assay G; e) obtaining a DNA copy number of the target site from the results of assay T and a DNA copy number of the transgene from the results of assay G; and f) identifying a cell line or tissue that has reduced copy number in assay T relative to a reference and a copy number greater than zero for assay G, thereby identifying the cell comprising the transgene inserted into the nuclease cleavage site in the genome of the cell. The present invention also provides for a cell line or tissue that is enriched for or identified by the described methods, and further provides for a plant, plant part, or progeny thereof derived from the cell line or tissue.

In further aspects of this invention, a method is provided of enriching for a cell comprising a mutation introduced into a nuclease cleavage site in a genome of the cell and lacking integration of a heterologous nucleotide sequence encoding a nuclease for site-directed cleavage of a nucleotide sequence at the nuclease cleavage site into the genome of

the cell, comprising: a) introducing a nucleic acid molecule comprising a heterologous sequence encoding a nuclease for site-directed cleavage of the nucleotide sequence at the nuclease cleavage site in the genome of the cell into a plurality of cells under conditions wherein expression of the nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave the nucleotide sequence at the nuclease cleavage site in the genome of the cell, thereby introducing a mutation at the nuclease cleavage site in the genome of the cell without integration of the heterologous nucleotide sequence encoding the nuclease into the genome of the cell; b) culturing the plurality of cells of (a) to produce a cell line or tissue; c) extracting a genomic DNA sample from the cell line or tissue of (b); d) performing real-time quantitative polymerase chain reaction (qPCR) assays 1 and 2 on the sample of (c), wherein the assays respectively comprise the following probes: i) a first probe comprising a nucleotide sequence that is complementary to the nucleotide sequence comprising the nuclease cleavage site to carry out assay 1, and ii) a second probe comprising a nucleotide sequence that is complementary to the heterologous nucleotide sequence encoding the nuclease to carry out assay 2; e) obtaining a DNA copy number of the nuclease cleavage site from the results of assay 1 and a DNA copy number of the heterologous nucleotide sequence encoding the nuclease from the results of assay 2; and f) enriching for a cell line or tissue that has a reduced copy number for assay 1 relative to a reference and a copy number equal to zero for assay 2, thereby enriching for the cell comprising the mutation introduced into the nuclease cleavage site in the genome of the cell and lacking integration of the heterologous nucleotide sequence encoding the nuclease into the genome of the cell.

Also provided as an aspect of this invention is a method of identifying a cell comprising a mutation introduced into a nuclease cleavage site in a genome of the cell and lacking integration of a heterologous nucleotide sequence encoding a nuclease for site-directed cleavage of a nucleotide sequence at the nuclease cleavage site into the genome of the cell, comprising: a) introducing a nucleic acid molecule comprising a heterologous sequence encoding a nuclease for site-directed cleavage of the nucleotide sequence at the nuclease cleavage site in the genome of the cell into a plurality of cells under conditions wherein expression of the nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave the nucleotide sequence at the nuclease cleavage site in the genome of the cell, thereby introducing a mutation at the nuclease cleavage site in the genome of the cell without integration of the heterologous nucleotide sequence encoding the nuclease into the genome of the cell; b) culturing the plurality of cells of (a) to produce a cell line or tissue; c) extracting a genomic DNA sample from the cell line or tissue of (b); d) performing real-time

quantitative polymerase chain reaction (qPCR) assays 1 and 2 on the sample of (c), wherein the assays respectively comprise the following probes: i) a first probe comprising a nucleotide sequence that is complementary to the nucleotide sequence comprising the nuclease cleavage site to carry out assay 1, and ii) a second probe comprising a nucleotide sequence that is complementary to the heterologous nucleotide sequence encoding the nuclease to carry out assay 2; e) obtaining a DNA copy number of the nuclease cleavage site from the results of assay 1 and a DNA copy number of the heterologous nucleotide sequence encoding the nuclease from the results of assay 2; and f) identifying a cell line or tissue that has a reduced copy number for assay 1 relative to a reference and a copy number equal to zero for assay 2, thereby identifying the cell comprising the mutation introduced into the nuclease cleavage site in the genome of the cell and lacking integration of the heterologous nucleotide sequence encoding the nuclease into the genome of the cell. The present invention also provides for a cell line or tissue that is enriched for or identified by the described methods, and further provides for a plant, plant part, or progeny thereof derived from the cell line or tissue.

In additional aspects, the present invention provides a method of producing a plant, plant part, or progeny thereof comprising a mutation introduced at a nuclease cleavage site in a genome of a plant cell and lacking integration of a heterologous nucleotide sequence encoding a nuclease for site-directed cleavage of a nucleotide sequence at the nuclease cleavage site in the genome of the plant cell, comprising: a) introducing into the plant cell a nucleic acid molecule comprising a heterologous nucleotide sequence encoding a nuclease for site-directed cleavage of the nucleotide sequence at the nuclease cleavage site in the genome of the plant cell under conditions wherein expression of the nucleic acid molecule occurs transiently to produce the nuclease and the nuclease can cleave the nucleotide sequence at the nuclease cleavage site in the genome of the plant cell, thereby introducing a mutation at the nuclease cleavage site in the genome of the plant cell without integration of the heterologous nucleotide sequence encoding the nuclease into the genome of the plant cell; and b) regenerating a plant, plant part, or progeny thereof from the plant cell of (a). The present invention further provides the plant, plant part, or progeny thereof produced by the method described.

The present invention also provides a method for modifying a target site in the genome of a plant cell, comprising: a) introducing into the plant cell a first nucleic acid comprising at least 100 contiguous nucleotides, wherein the at least 100 contiguous

nucleotides have at least 90% identity with a target site in the genome of the cell, and further comprising a transgene; and b) a second nucleic acid molecule encoding nuclease for site-directed cleavage at a nuclease cleavage site in the genome of the cell adjacent to the nucleotide sequence in the genome of the cell that corresponds to the at least 100 contiguous nucleotides of (a), wherein the nuclease is a modified Cas9 nuclease comprising SEQ ID NO: 30, under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave at the nuclease cleavage site in the genome of the cell and modify the target site in the genome of the plant cell.

The present invention also provides a method of producing a maize plant, plant part, or progeny thereof comprising a modification at a target site in the genome of the plant cell, comprising: a) introducing into the plant cell a first nucleic acid comprising at least 100 contiguous nucleotides, wherein the at least 100 contiguous nucleotides have at least 90% identity with a target site in the genome of the cell, and further comprising a transgene; b) a second nucleic acid molecule encoding nuclease for site-directed cleavage at a nuclease cleavage site in the genome of the cell adjacent to the nucleotide sequence in the genome of the cell that corresponds to the at least 100 contiguous nucleotides of (a), wherein the nuclease is a modified Cas9 nuclease comprising SEQ ID NO: 30, under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave at the nuclease cleavage site in the genome of the cell and modify the target site in the genome of the plant cell; and c) regenerating a plant, plant part, or progeny thereof from the plant cell of (a). The present invention further provides the plant, plant part, or progeny thereof produced by the method described.

The present invention also provides a method of integrating a transgene into a genomic nuclease cleavage site in an event MTR604 transgenic maize genome, comprising introducing into an event MIR604 maize cell: a) a first nucleic acid molecule comprising at least 100 contiguous nucleotides, wherein said at least 100 contiguous nucleotides have at least 90% identity with a target site in a nucleotide sequence selected from the group comprising SEQ ID NO: 133, SEQ ID NO: 134, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 137, SEQ ID NO: 138, and SEQ ID NO: 139, and further comprising a transgene; and b) a second nucleic acid molecule comprising a nucleotide sequence encoding a nuclease for site-directed cleavage at a genomic nuclease cleavage site adjacent to a nucleotide sequence with at least 90% identity to a nucleotide sequence selected from the group comprising SEQ ID NO: 133, SEQ ID NO: 134, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 137, SEQ ID NO: 138, and SEQ ID NO: 139, that corresponds to the at least 100

contiguous nucleotides of (a), under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave the nucleotide sequence at the genomic nuclease cleavage site, whereby the transgene is integrated at the genomic nuclease target cleavage site in the maize genome. The present invention further provides a method of producing a maize plant, plant part, or progeny thereof comprising a transgene integrated into a genomic nuclease cleavage site in an event MIR604 maize genome, comprising regenerating a maize plant from the maize cell produced by the method described. The present invention further provides a maize plant, plant part, or progeny thereof comprising comprising a transgene integrated into a genomic nuclease cleavage site in the event MTR604 maize genome, produced by the method described.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1. Schematic diagram showing MTR604 insertion site flanking sequences. 88 base pairs of sequences between MIR604RBFS 1 and MTR604LBFS1 are deleted in MIR604 event during T-DNA integration. This MIR604 insertion site does not contain the event MIR604 transgene.

Fig. 2. Schematic representation of targeted insertion into MIR604 insertion site safe harbour locus and PCR reactions to identify potential targeted integration events with two primer pairs: P I (FE4706)/P2 (FE4705) and P3 (FE4708)/P4 (FE4707). P I (FE4706) and P4 (FE4707) only binds to chromosomal regions outside the homology arms present in the donor and target region, whereas P2 (FE4705) and P3 (FE4708) only binds to donor molecules. Primer pair P1(FE4706) and P2 (FE4705) produces a fragment of 2.87 Kbp and primer pair P3(FE4708)/ and P4(FE4707) amplifies a fragment of 2.0 Kbp only if targeted insertion is present at the safe harbor locus # 1 (MIR604 insertion site). The approximate position of Bsu36I restriction sites and probes used in Southern DNA blot analysis **(Fig. 5)** are indicated in the targeted insertion event.

Fig. 3. An example of PCR screening assay as outlined in **Fig. 2.** In the left panel, PCR is done with P3(FE4708)/ and P4(FE4707) which amplifies a fragment of 2.0 Kbp from 2 events (lane 25, MZET141320A250A and lane 42, event MZET141606A097A). In the right panel, PCR is done with pair P I (FE4706) and P2 (FE4705) produces a fragment of 2.87 Kbp from only 1 event (lane 25, MZET141320A250A).

Fig. 4. Number of GUS spots in maize immature embryos bombarded with vectors containing GUUS repeat intra-molecular recombination substrate with MIR604FR1 target

sequence (5'-TACAC GTACT AATCG TGCTT CACGC ACAGG CACAG CACGT AGTAG ACAGG A-3', SEQ ID NO:66) along with single TALEN vector (Fl, cTNmir604Fwl-01 or R2, cTNmir604Rv2-01) or a pair of TALEN genes (FR1, cTNmir604Fwl-01 and cTNmir604Rvl-01) under the control of maize ubiquitin promoter (prZmUbi 1-10) or without TALEN (ctl, blank control). cTNmir604Rv2-0 1 does not recognize MTR604FR1 sequence and results in background level of GUS activity (R2, negative control).

Fig. 5. DNA blot analysis of targeted insertion events at the safe harbor locus #1 (MTR604 insertion site). DNA Probe 1: against flanking native genomic sequences; Probe 2: probe against prCMP; Probe 3: Probe against cPMI (See **Fig. 2** for probe locations in the schematic map). Lane 1: DIG-labeled markers; Lane 2: Wild type maize transformation line NP2222; Lane 3: NP2222 spiked with 21942; digested with *Hindlll* (releasing a 8553 bp fragment); Lane 4: MZET134207E056A; Lane 5: MZET134300A679A; Lane 6: MZET1 34505 A104A; Lane 7: MZET141322A015A; Lane 8: MZET141322B143A; All maize genomic DNAs in lane 2 to 8 were digested with *Bsu36I* restriction enzyme. Note: Probe 1 also hybridizes weakly to homologous sequences in other parts of the genome. WT safe harbor locus has the dominant 17.5 Kb band, whereas targeted insertion events have the fragment size increased to 28 Kb. For probe 2 and 3, the 28 Kb Bsu36I bands contain targeted insertion of donor DNA sequences through homologous recombination. In lane 7, the event likely contains an insertion of the rearranged donor DNA molecule.

Fig. 6. Schematic diagram showing reduction of target sequence copy number in a plant with a mutation in the target sequence (M) generated by cleavage with a site-directed nuclease.

Fig. 7. Schematic representation of Taqman assay probe design for a target sequence in the MIR604 insertion site and interpretation of Taqman assay results in regard to targeted mutation.

Figs. 8A-B. Strategies to enrich for potential targeted insertion events based on copy number reduction of target sequences. **(A)** Schematic representation of potential types of mutations and targeted insertion as a result of targeted nuclease cleavage at the target locus in a targeted insertion experiment. M is the site-directed nuclease cleavage site; T is a sequence located away from M by at least 5 nucleotides in the region of the target locus and it should be as far away as possible from M but within the region replaced by targeted insertion. However, T can sit within the same amplicon as assay for M. G is an assay target for transgenic sequences (gene of interest (GOI)). (B) Copy number call of different assays in

plants with different kinds of mutations or insertions in the target site as shown in (A) using real-time qPCR assays.

Fig.9. Schematic drawing of MIR604 transgenic event T-DNA insertions and flanking regions. MIR604 RB FS: maize genomic region flanking the T-DNA right border; MIR604 LB FS: maize genomic region flanking the T-DNA left border; prUbil : maize ubiquitin-1 promoter; cPMI-01: PMI coding sequence; tNOS: Nopaline synthase terminator; mCry3A: synthetic form of Cry3A gene sequence (mCry3A) from *Bacillus thuringiensis.* (U.S. Patent No. 7,897,748)

Fig. 10. Targeted insertion of an insecticidal gene (IC) expression cassette (Exp. Cass.) and an expression cassette comprising the selectable marker ZmEPSPS (EPSPS Exp. Cass.) from donor vector 22872 the into MTR604 transgene locus **(Fig. 9)** mediated by TALENs expressed from vector 22840. A pair of TALENs is expressed from 22840 and cleaves the cPMI target sequence. t: tNOS-05; LBFS: maize genomic sequences flanking the T-DNA Left Border; RBFS: maize genomic sequences flanking the T-DNA Right Border; P1(FE4796): SEQ ID NO: 127; P2 (FE4793): SEQ ID NO: 128; P3 (FE35035): SEQ ID NO: 132; P4 (FE35034): SEQ ID NO: 131.

Fig. 11. Targeted insertion of transgene expression cassettes into MTR604 transgene locus **(Fig. 9)** mediated by site-directed nuclease to replace the whole PMI marker gene cassette.

Fig. 12. Targeted insertion of the donor transgene expression cassettes into MTR604 transgene locus **(Fig. 9)** mediated by site-directed nuclease to replace the whole MTR604 T-DNA insert.

BRIEF DESCRIPTION OF THE SEQUENCES IN THE SEQUENCE LISTING

SEQ ID NO: 1 is a nucleotide sequence of the MTR604 insertion site sequence from maize line A188. This the MTR604 insertion site without an event MTR604 transgene.

SEQ ID NO: 2 is maize elite line NP2222 genomic sequences corresponding to the A188 MTR604 insertion site and its flanking sequences.

SEQ ID NO: 3-27 are nucleotide sequences that are potential target sequences for Cas9-mediated cleavage proximal to the MTR604 insertion site.

SEQ ID NO: 28 is a maize genomic target sequence, MIR604FR2.

SEQ ID NO: 29 is a nucleotide sequence encoding a Type II Cas9 gene from *Streptococcus pyogenes* SF370 optimized with maize-preferred codons.

SEQ ID NO: 30 is an amino acid sequence comprising a modified Cas9 protein.

SEQ ID NO: 31-34 are nucleotide sequences that can be used to guide Cas9 cleavage of the MIR604 insertion site.

SEQ ID NO: 35 is a nucleotide sequence encoding tracRNA scaffold and PolIII termination sequences.

SEQ ID NO: 36 is a nucleotide sequence encoding a single guide RNA (sgRNA).

SEQ ID NO: 37 is a nucleotide sequence comprising an expression cassette comprising prOsU3 and coding sequences for the sgRNA of SEQ ID NO: 36.

SEQ ID NO: 38 is a nucleotide sequence comprising xJHAX-03.

SEQ ID NO: 39 is a nucleotide sequence comprising xJHAX-04.

SEQ ID NO: 40-65 are nucleotide sequences selected as TALEN target sequences based on NP2222 genomic sequences (SEQ ID NO: 2).

SEQ ID NO: 66 is a nucleotide sequence comprising the TALEN target sequence MIR604FR1.

SEQ ID NO: 67 is a nucleotide sequence comprising the TALEN target sequence MIR604FR2.

SEQ ID NO: 68 is an amino acid sequence of the artificial nuclease cTNmir604Fwl-01 which recognizes target sequence SEQ ID NO: 42.

SEQ ID NO: 69 is an amino acid sequence of the artificial nuclease cTNmir604Fwl-02 which recognizes target sequence SEQ ID NO: 42.

SEQ ID NO: 70 is an amino acid sequence of the artificial nuclease cTNmirFwl-03 which recognizes target sequence SEQ ID NO: 42.

SEQ ID NO: 71 is an amino acid sequence of the artificial nuclease cTNmir604Rvl-01 which recognizes target sequence SEQ ID NO: 43.

SEQ ID NO: 72 is an amino acid sequence of the artificial nuclease cTNmir604Rvl-02 which recognizes target sequence SEQ ID NO: 43.

SEQ ID NO: 73 is an amino acid sequence of the artificial nuclease cTNmir604Rvl-03 which recognizes target sequence SEQ ID NO: 43.

SEQ ID NO: 74 is an amino acid sequence of the artificial nuclease cTNmir604Fw2-01 which recognizes target sequence SEQ ID NO: 53.

SEQ ID NO: 75 is an amino acid sequence of the artificial nuclease cTNmir604Fw2-02 which recognizes target sequence SEQ ID NO: 53.

SEQ ID NO: 76 is an amino acid sequence of the artificial nuclease cTNmir604Fw2-03 which recognizes target sequence SEQ ID NO: 53.

SEQ ID NO: 77 is an amino acid sequence of the artificial nuclease cTNmire604RV2-01 which recognizes target sequence SEQ ID NO: 54.

SEQ ID NO: 78 is an amino acid sequence of the artificial nuclease cTNmir604RV2-02 which recognizes target sequence SEQ ID NO: 54.

SEQ ID NO: 79 is an amino acid sequence of the artificial nuclease cTNmir604Rv2-03 which recognizes target sequence SEQ ID NO: 54.

SEQ ID NO: 80 is an amino acid sequence of the artificial nuclease cTNmir604Fw2-05 which recognizes target sequence SEQ ID NO: 53.

SEQ ID NO: 81 is an amino acid sequence of the artificial nuclease cTNmir604Rv2-04 which recognizes target sequence SEQ ID NO: 65.

SEQ ID NO: 82 is a nucleotide sequence encoding for the full length artificial nuclease molecule cTNmir604Fwl-01 (SEQ ID NO:68).

SEQ ID NO: 83 is a nucleotide sequence encoding for the truncated artificial nuclease molecule cTNmir604Fwl-03 (SEQ ID NO:70).

SEQ ID NO: 84 is a nucleotide sequence encoding for the full length artificial nuclease molecule cTNmir604Rvl-01 (SEQ ID NO:71).

SEQ ID NO: 85 is a nucleotide sequence encoding for the truncated artificial nuclease molecule cTNmir604Rvl-03 (Seq.ID No.72).

SEQ ID NO: 86 is a nucleotide sequence encoding for the full length artificial nuclease molecule cTNmir604Fw2-01 (SEQ ID NO: 72).

SEQ ID NO: 87 is a nucleotide sequence encoding for the truncated artificial nuclease molecule cTNmir604Fw2-03 (SEQ ID NO:73).

SEQ ID NO: 88 is a nucleotide sequence encoding for the truncated artificial nuclease molecule cTNmir604Fw2-05 (SEQ ID NO: 80).

SEQ ID NO: 89 is a nucleotide sequence encoding for the full length artificial nuclease molecule cTNmir604Rv2-01 (SEQ ID NO:77).

SEQ ID NO: 90 is a nucleotide sequence encoding for the truncated artificial nuclease molecule cTNmir604Rv2-03 (SEQ ID NO:79).

SEQ ID NO: 91 is a nucleotide sequence encoding for the truncated artificial nuclease molecule cTNmir604Rv2-04 (SEQ ID NO.81).

SEQ ID NO: 92-97 are nucleotide sequences useful for using qPCR for the detection of mutations within the SEQ ID NO: 67 target sequence.

SEQ ID NO: 98 is a nucleotide sequence comprising a gene encoding phosphomannose isomerase (cPMI-01).

SEQ ID NO: 99-101 are nucleotide sequences comprising PMI target sequences for genomic modification meditated by TALENs

SEQ ID NO: 102-107 are nucleotide sequences comprising TALEN sequence targets within SEQ ID NO: 98.

SEQ ID NO: 108 is an amino acid sequence of the artificial nuclease protein TLN_PMIFW 1a which recognizes SEQ ID NO: 102

SEQ ID NO: 109 is an amino acid sequence of the artificial nuclease protein TLN_PMIRVla which recognizes SEQ ID NO: 103.

SEQ ID NO: 110 is an amino acid sequence of the artificial nuclease protein TLN_PMIFW3 which recognizes SEQ ID NO: 106

SEQ ID NO: 111 is an amino acid sequence of the artificial nuclease protein TLN_PMIRV3 which recognizes SEQ ID NO: 107.

SEQ ID NO: 112 is a nucleotide sequence which encodes for the artificial nuclease protein TLN_PMIFWla .

SEQ ID NO: 113 is a nucleotide sequence which encodes for the artificial nuclease protein TLN_PMIRVla.

SEQ ID NO: 114 is a nucleotide sequence which encodes for the artificial nuclease protein TLN_PMIFW3 .

SEQ ID NO: 115 is a nucleotide sequence which encodes for the artificial nuclease protein TLN_PMIRV3.

SEQ ID NO: 116-1 18 are nucleotide sequences comprising the artificial nuclease target sequences.

SEQ ID NO: 119 is an amino acid sequence of the artificial nuclease protein TLN_rPMIFWl-01 which recognizes SEQ ID NO: 117.

SEQ ID NO: 120 is an amino acid sequence of the artificial nuclease protein TLN_rPMIRv 1-01 which recognizes SEQ ID NO: 118.

SEQ ID NO: 121 is an amino acid sequence of the artificial nuclease protein TLN_rPMIFwl -02 which recognizes SEQ ID NO: 117.

SEQ ID NO: 122 is an amino acid sequence of the artificial nuclease protein TLN_rPMIRv 1-02 which recognizes SEQ ID NO: 118.

SEQ ID NO: 123 is a nucleotide sequence encoding the artificial nuclease protein TLN_rPMIFWl-01.

SEQ ID NO: 124 is a nucleotide sequence encoding the artificial nuclease protein TLN_rPMIRvl-01.

SEQ ID NO: 125 is a nucleotide sequence encoding the artificial nuclease protein TLN_rPMIFWl-02.

SEQ ID NO: 126 is a nucleotide sequence encoding the artificial nuclease protein TLN_rPMIRvl-02.

SEQ ID NO: 127-132 are nucleotide sequences useful for the detection of targeted integration.

SEQ ID NO: 133 is a nucleotide sequence of the PMI expression cassette (prZmUbil-cPMI-tNOS) present in the T-DNA insert of event MIR604 transgenic plants **(Fig. 9)**.

SEQ ID NO: 134 is a nucleotide sequence of the T-DNA insert present in event MIR604 and of the right and left border regions **(Fig. 9).**

SEQ ID NO: 135 is a nucleotide sequence of event MIR604 transgene locus including the whole T-DNA insert and the flanking genomic DNA regions, including RBFS and LBFS **(Fig. 9).**

SEQ ID NO: 136 is a nucleotide sequence of the B73 maize genomic region proximal to the MIR604 T-DNA insertion right border (RB) region (RBFS in **Fig. 9).**

SEQ ID NO: 137 is a nucleotide sequence of the B73 maize genomic region proximal to the MIR604 T-DNA insertion left border (LB) region (LBFS in **Fig. 9).**

SEQ ID NO: 138 is a nucleotide sequence of the elite maize line NP2222 genomic sequence corresponding to the B73 MTR604 insertion site locus sequences proximal to the RB region including the RBFS **(Fig. 9).**

SEQ ID NO: 139 is a nucleotide sequence of the elite maize line NP2222 genomic sequence corresponding to the B73 MTR604 T-DNA insertion site locus sequences proximal to the LB region including the LBFS **(Fig. 9).**

DETAILED DESCRIPTION OF THE INVENTION

This description is not intended to be a detailed catalog of all the different ways in which the invention may be implemented, or all the features that may be added to the instant invention. For example, features illustrated with respect to one embodiment may be incorporated into other embodiments, and features illustrated with respect to a particular embodiment may be deleted from that embodiment. In addition, numerous variations and additions to the various embodiments suggested herein will be apparent to those skilled in the art in light of the instant disclosure, which do not depart from the instant invention. Hence, the following descriptions are intended to illustrate some particular embodiments of the

invention, and not to exhaustively specify all permutations, combinations and variations thereof.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The terminology used in the description of the invention herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the invention. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety.

The following definitions and methods are provided to better define the present invention and to guide those of ordinary skill in the art in the practice of the present invention. Unless otherwise noted, terms used herein are to be understood according to conventional usage by those of ordinary skill in the relevant art. Definitions of common terms in molecular biology may also be found in Rieger et al., Glossary of Genetics: Classical and Molecular, 5th edition, Springer-Verlag: New York, 1994.

"Accuracy" of an amplification method such as a polymerase chain reaction (PCR) method (e.g., TaqMan) means the closeness of agreement between a test result and an accepted reference value.

As used herein, the term "amplified" means the construction of multiple copies of a nucleic acid molecule or multiple copies complementary to the nucleic acid molecule using at least one of the nucleic acid molecules as a template. See, e.g., Diagnostic Molecular Microbiology: Principles and Applications, D. H. Persing et al., Ed., American Society for Microbiology, Washington, D.C. (1993). The product of amplification is termed an amplicon.

A "coding sequence" is a nucleic acid sequence that is transcribed into RNA such as mRNA, rRNA, tRNA, snRNA, sense RNA or antisense RNA. In some embodiments, the RNA is then translated in an organism to produce a protein.

The "coefficient of linearity (R^2)" is the correlation coefficient of a standard curve obtained by linear regression analysis.

"Dynamic range" as used herein means the range of DNA concentrations over which the method of the invention performs in a linear manner with an acceptable level of accuracy and precision.

"Detection kit" as used herein refers to a kit used to detect target DNA from the events of interest in a sample comprising nucleic acid probes and primers of the present invention, which will be processed specifically under optimum conditions to a target DNA

sequence, and other materials necessary to enable nucleic acid hybridization and/or amplification methods.

As used herein the term transgenic "event" refers to a recombinant plant produced by transformation and regeneration of a single plant cell with heterologous DNA, for example, an expression cassette that includes one or more genes of interest (e.g., transgenes). The term "event" refers to the original transformant and/or progeny of the transformant that include the heterologous DNA. The term "event" also refers to progeny produced by a sexual outcross between the transformant and another line. Even after repeated backcrossing to a recurrent parent, the inserted DNA and the flanking DNA from the transformed parent is present in the progeny of the cross at the same chromosomal location. Normally, transformation of plant tissue produces multiple events, each of which represent insertion of a DNA construct into a different location in the genome of a plant cell. Based on the expression of the transgene or other desirable characteristics, a particular event is selected. Thus, "event MIR604," "MTR604" or "MTR604 event" as used herein, means the original MTR604 transformant and/or progeny of the MIR604 transformant (U.S. Patent Nos. 7,361,813, 7,897,748, 8,354,519, and 8,884,102, incorporated by references herein).

The insertion site of event MTR604 has many characteristics which make it a good candidate for a target site for genomic modifications. Such characteristics include that the site does not interrupt native genes, the site is not in a highly repetitive region of nucleotide sequence, the nucleotide sequence of the site is not significantly repeated elsewhere in the maize genome, and transgenes introduced at this site are known to have good expression levels, both in the initially transformed plant, in other maize varieties into which event MTR604 has been introduced, and in the progeny of event MTR604 plants, for multiple generations. Additionally, the success of event MTR604 as a commercial product and in a successful commercial-level breeding program, where event MTR604 is introduced into at least dozens of maize varieties and has shown excellent expression of the transgenes in multiple environmental conditions, indicates that the event MTR604 insertion site is a good candidate for targeted insertion.

"Expression cassette" as used herein means a nucleic acid molecule capable of directing expression of a particular nucleotide sequence in an appropriate host cell, comprising a promoter operably linked to the nucleotide sequence of interest, typically a coding region, which is operably linked to termination signals. It also typically comprises sequences required for proper translation of the nucleotide sequence. The coding region usually codes for a protein of interest but may also code for a functional RNA of interest, for

example antisense RNA or a nontranslated RNA, in the sense or antisense direction. The expression cassette may also comprise sequences not necessary in the direct expression of a nucleotide sequence of interest but which are present due to convenient restriction sites for removal of the cassette from an expression vector. The expression cassette comprising the nucleotide sequence of interest may be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components. The expression cassette may also be one that is naturally occurring but has been obtained in a recombinant form useful for heterologous expression. Typically, however, the expression cassette is heterologous with respect to the host, i.e.,, the particular nucleic acid sequence of the expression cassette does not occur naturally in the host cell and must have been introduced into the host cell or an ancestor of the host cell by a transformation process known in the art. The expression of the nucleotide sequence in the expression cassette may be under the control of a constitutive promoter or of an inducible promoter that initiates transcription only when the host cell is exposed to some particular external stimulus. In the case of a multicellular organism, such as a plant, the promoter can also be specific to a particular tissue, or organ, or stage of development. An expression cassette, or fragment thereof, can also be referred to as "inserted sequence" or "insertion sequence" when transformed into a plant.

A "gene" is a defined region that is located within a genome and that, besides the aforementioned coding nucleic acid sequence, comprises other, primarily regulatory, nucleic acid sequences responsible for the control of the expression, that is to say the transcription and translation, of the coding portion. A gene may also comprise other 5' and 3' untranslated sequences and termination sequences. Further elements that may be present are, for example, introns.

"Gene of interest" refers to any gene which, when transferred to a plant, confers upon the plant a desired characteristic such as antibiotic resistance, virus resistance, insect resistance, disease resistance, or resistance to other pests, herbicide tolerance, improved nutritional value, improved performance in an industrial process or altered reproductive capability. The "gene of interest" may also be one that is transferred to plants for the production of commercially valuable enzymes or metabolites in the plant.

"Genotype" as used herein is the genetic material inherited from parent plants not all of which is necessarily expressed in the descendant plants. By way of example, the MTR604 genotype refers to the heterologous genetic material transformed into the genome of a plant as well as the genetic material flanking the inserted sequence.

As used herein, "heterologous" refers to a nucleic acid molecule or nucleotide sequence not naturally associated with a host cell into which it is introduced, that either originates from another species or is from the same species or organism but is modified from either its original form or the form primarily expressed in the cell, including non-naturally occurring multiple copies of a naturally occurring nucleic acid sequence. . Thus, a nucleotide sequence derived from an organism or species different from that of the cell into which the nucleotide sequence is introduced, is heterologous with respect to that cell and the cell's descendants. In addition, a heterologous nucleotide sequence includes a nucleotide sequence derived from and inserted into the same natural, original cell type, but which is present in a non-natural state, e.g., present in a different copy number, and/or under the control of different regulatory sequences than that found in the native state of the nucleic acid molecule. A nucleic acid sequence can also be heterologous to other nucleic acid sequences with which it may be associated, for example in a nucleic acid construct, such as e.g., an expression vector. As one nonlimiting example, a promoter may be present in a nucleic acid construct in combination with one or more regulatory element and/or coding sequences that do not naturally occur in association with that particular promoter, i.e.,, they are heterologous to the promoter.

A "homologous" nucleic acid sequence is a nucleic acid sequence naturally associated with a host cell into which it is introduced. A homologous nucleic acid sequence can also be a nucleic acid sequence that is naturally associated with other nucleic acid sequences that may be present, e.g., in a nucleic acid construct. As one nonlimiting example, a promoter may be present in a nucleic acid construct in combination with one or more regulatory elements and/or coding sequences that naturally occur in association with that particular promoter, i.e.,, they are homologous to the promoter.

"Operably-linked" refers to the association of nucleic acid sequences on a single nucleic acid sequence so that the function of one affects the function of the other. For example, a promoter is operably-linked with a coding sequence or functional RNA when it is capable of affecting the expression of that coding sequence or functional RNA (i.e.,, the coding sequence or functional RNA is under the transcriptional control of the promoter). Coding sequences in sense or antisense orientation can be operably-linked to regulatory sequences.

"Primers" as used herein are isolated nucleic acids that are annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, then extended along the target DNA strand by a

polymerase, such as DNA polymerase. Primer pairs or sets can be used for amplification of a nucleic acid molecule, for example, by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods.

A "probe" is an isolated nucleic acid molecule that is complementary to a portion of a target nucleic acid molecule and is typically used to detect and/or quantify the target nucleic acid molecule. Thus, in some embodiments, a probe can be an isolated nucleic acid molecule to which is attached a detectable moiety or reporter molecule, such as a radioactive isotope, ligand, chemiluminescenc agent, fluorescence agent or enzyme. Probes according to the present invention can include not only deoxyribonucleic or ribonucleic acids but also polyamides and other probe materials that bind specifically to a target nucleic acid sequence and can be used to detect the presence of and/or quantify the amount of, that target nucleic acid sequence.

A TaqMan probe is designed such that it anneals within a DNA region amplified by a specific set of primers. As the Taq polymerase extends the primer and synthesizes the nascent strand from a single-strand template from 3' to 5' of the complementary strand, the 5' to 3' exonuclease of the polymerase extends the nascent strand through the probe and consequently degrades the probe that has annealed to the template. Degradation of the probe releases the fluorophore from it and breaks the close proximity to the quencher, thus relieving the quenching effect and allowing fluorescence of the fluorophore. Hence, fluorescence detected in the quantitative PCR thermal cycler is directly proportional to the fluorophore released and the amount of DNA template present in the PCR.

Primers and probes are generally between 5 and 100 nucleotides or more in length. In some embodiments, primers and probes can be at least 20 nucleotides or more in length, or at least 25 nucleotides or more, or at least 30 nucleotides or more in length. Such primers and probes hybridize specifically to a target sequence under optimum hybridization conditions as are known in the art. Primers and probes according to the present invention may have complete sequence complementarity with the target sequence, although probes differing from the target sequence and which retain the ability to hybridize to target sequences may be designed by conventional methods according to the invention.

Methods for preparing and using probes and primers are described, for example, in Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, ed. Sambrook et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989. PCR-primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose.

The polymerase chain reaction (PCR) is a technique for "amplifying" a particular piece of DNA. In order to perform PCR, at least a portion of the nucleotide sequence of the DNA molecule to be replicated must be known. In general, primers or short oligonucleotides are used that are complementary (e.g., substantially complementary or fully complementary) to the nucleotide sequence at the 3' end of each strand of the DNA to be amplified (known sequence). The DNA sample is heated to separate its strands and is mixed with the primers. The primers hybridize to their complementary sequences in the DNA sample. Synthesis begins (5' to 3' direction) using the original DNA strand as the template. The reaction mixture must contain all four deoxynucleotide triphosphates (dATP, dCTP, dGTP, dTTP) and a DNA polymerase. Polymerization continues until each newly-synthesized strand has proceeded far enough to contain the sequence recognized by the other primer. Once this occurs, two DNA molecules are created that are identical to the original molecule. These two molecules are heated to separate their strands and the process is repeated. Each cycle doubles the number of DNA molecules. Using automated equipment, each cycle of replication can be completed in less than 5 minutes. After 30 cycles, what began as a single molecule of DNA has been amplified into more than a billion copies ($2^{30} = 1.02 \times 10^9$).

The oligonucleotides of an oligonucleotide primer pair are complementary to DNA sequences located on opposite DNA strands and flanking the region to be amplified. The annealed primers hybridize to the newly synthesized DNA strands. The first amplification cycle will result in two new DNA strands whose 5' end is fixed by the position of the oligonucleotide primer but whose 3' end is variable ('ragged' 3' ends). The two new strands can serve in turn as templates for synthesis of complementary strands of the desired length (the 5' ends are defined by the primer and the 3' ends are fixed because synthesis cannot proceed past the terminus of the opposing primer). After a few cycles, the desired fixed length product begins to predominate.

A quantitative polymerase chain reaction (qPCR), also referred to as real-time polymerase chain reaction, monitors the accumulation of a DNA product from a PCR reaction in real time. qPCR is a laboratory technique of molecular biology based on the polymerase chain reaction (PCR), which is used to amplify and simultaneously quantify a targeted DNA molecule. Even one copy of a specific sequence can be amplified and detected in PCR. The PCR reaction generates copies of a DNA template exponentially. This results in a quantitative relationship between the amount of starting target sequence and amount of PCR product accumulated at any particular cycle. Due to inhibitors of the polymerase reaction found with the template, reagent limitation or accumulation of pyrophosphate

molecules, the PCR reaction eventually ceases to generate template at an exponential rate (i.e.,, the plateau phase), making the end point quantitation of PCR products unreliable. Therefore, duplicate reactions may generate variable amounts of PCR product. Only during the exponential phase of the PCR reaction is it possible to extrapolate back in order to determine the starting quantity of template sequence. The measurement of PCR products as they accumulate (i.e., real-time quantitative PCR) allows quantitation in the exponential phase of the reaction and therefore removes the variability associated with conventional PCR. In a real time PCR assay, a positive reaction is detected by accumulation of a fluorescent signal. For one or more specific sequences in a DNA sample, quantitative PCR enables both detection and quantification. The quantity can be either an absolute number of copies or a relative amount when normalized to DNA input or additional normalizing genes. Since the first documentation of real-time PCR, it has been used for an increasing and diverse number of applications including mRNA expression studies, DNA copy number measurements in genomic or viral DNAs, allelic discrimination assays, expression analysis of specific splice variants of genes and gene expression in paraffin-embedded tissues and laser captured micro-dissected cells.

As used herein, the phrase "Ct value" refers to "threshold cycle," which is defined as the "fractional cycle number at which the amount of amplified target reaches a fixed threshold." In some embodiments, it represents an intersection between an amplification curve and a threshold line. The amplification curve is typically in an "S" shape indicating the change of relative fluorescence of each reaction (Y-axis) at a given cycle (X-axis), which in some embodiments is recorded during PCR by a real-time PCR instrument. The threshold line is in some embodiments the level of detection at which a reaction reaches a fluorescence intensity above background. *See* Livak & Schmittgen (2001) 25 *Methods* 402-408. It is a relative measure of the concentration of the target in the PCR. Generally, good Ct values for quantitative assays such as qPCR are in some embodiments in the range of 10-40 for a given reference gene. Ct levels are inversely proportional to the amount of target nucleic acid in the sample (i.e.,, the lower the Ct level the greater the amount of detectable target nucleic acid in the sample). Additionally, good Ct values for quantitative assays such as qPCR show a linear response range with proportional dilutions of target gDNA.

In some embodiments, qPCR is performed under conditions wherein the Ct value can be collected in real-time for quantitative analysis. For example, in a typical qPCR experiment, DNA amplification is monitored at each cycle of PCR during the extension stage. The amount of fluorescence generally increases above the background when DNA is

in the log linear phase of amplification. In some embodiments, the Ct value is collected at this time point.

The term "transformation" as used herein refers to the transfer of a nucleic acid molecule into the genome of a host cell, resulting in genetically stable inheritance. In some embodiments, the introduction into a plant, plant part and/or plant cell is via bacterial-mediated transformation, particle bombardment transformation, calcium-phosphate-mediated transformation, cyclodextrin-mediated transformation, electroporation, liposome-mediated transformation, nanoparticle-mediated transformation, polymer-mediated transformation, virus-mediated nucleic acid delivery, whisker-mediated nucleic acid delivery, microinjection, sonication, infiltration, polyethylene glycol-mediated transformation, protoplast transformation, or any other electrical, chemical, physical and/or biological mechanism that results in the introduction of nucleic acid into the plant, plant part and/or cell thereof, or any combination thereof.

Procedures for transforming plants are well known and routine in the art and are described throughout the literature. Non-limiting examples of methods for transformation of plants include transformation via bacterial-mediated nucleic acid delivery *(e.g.*, via bacteria from the genus *Agrobacterium)*, viral-mediated nucleic acid delivery, silicon carbide or nucleic acid whisker-mediated nucleic acid delivery, liposome mediated nucleic acid delivery, microinjection, microparticle bombardment, calcium-phosphate-mediated transformation, cyclodextrin-mediated transformation, electroporation, nanoparticle-mediated transformation,, sonication, infiltration, PEG-mediated nucleic acid uptake, as well as any other electrical, chemical, physical (mechanical) and/or biological mechanism that results in the introduction of nucleic acid into the plant cell, including any combination thereof. General guides to various plant transformation methods known in the art include Miki et al. ("Procedures for Introducing Foreign DNA into Plants" in Methods in Plant Molecular Biology and Biotechnology , Glick, B. R. and Thompson, J. E., Eds. (CRC Press, Inc., Boca Raton, 1993), pages 67-88) and Rakowoczy-Trojanowska *(Cell Mol Biol Lett* 7:849-858 (2002)).

Agrobacterium-mediated transformation is a commonly used method for transforming plants because of its high efficiency of transformation and because of its broad utility with many different species. *Agrobacterium-mediated* transformation typically involves transfer of the binary vector carrying the foreign DNA of interest to an appropriate *Agrobacterium* strain that may depend on the complement of vir genes carried by the host *Agrobacterium* strain either on a co-resident Ti plasmid or chromosomally (Uknes et al. 1993, *Plant Cell*

5:159-169). The transfer of the recombinant binary vector to *Agrobacterium* can be accomplished by a tri-parental mating procedure using *Escherichia coli* carrying the recombinant binary vector, a helper *E. coli* strain that carries a plasmid that is able to mobilize the recombinant binary vector to the target *Agrobacterium* strain. Alternatively, the recombinant binary vector can be transferred to *Agrobacterium* by nucleic acid transformation (Hofgen and Willmitzer 1988, *Nucleic Acids Res* 16:9877).

Transformation of a plant by recombinant *Agrobacterium* usually involves co-cultivation of the *Agrobacterium* with explants from the plant and follows methods well known in the art. Transformed tissue is typically regenerated on selection medium carrying an antibiotic or herbicide resistance marker between the binary plasmid T-DNA borders.

Another method for transforming plants, plant parts and plant cells involves propelling inert or biologically active particles at plant tissues and cells. *See, e.g.,* US Patent Nos. 4,945,050; 5,036,006 and 5,100,792. Generally, this method involves propelling inert or biologically active particles at the plant cells under conditions effective to penetrate the outer surface of the cell and afford incorporation within the interior thereof. When inert particles are utilized, the vector can be introduced into the cell by coating the particles with the vector containing the nucleic acid of interest. Alternatively, a cell or cells can be surrounded by the vector so that the vector is carried into the cell by the wake of the particle. Biologically active particles *{e.g.,* dried yeast cells, dried bacteria or a bacteriophage, each containing one or more nucleic acids sought to be introduced) also can be propelled into plant tissue.

Thus, in particular embodiments of the present invention, a plant cell can be transformed by any method known in the art and as described herein and intact plants can be regenerated from these transformed cells using any of a variety of known techniques. Plant regeneration from plant cells, plant tissue culture and/or cultured protoplasts is described, for example, in Evans et al. (Handbook of Plant Cell Cultures, Vol. 1, MacMilan Publishing Co. New York (1983)); and Vasil I. R. (ed.) (Cell Culture and Somatic Cell Genetics of Plants, Acad. Press, Orlando, Vol. I (1984), and Vol. II (1986)). Methods of selecting for transformed transgenic plants, plant cells and/or plant tissue culture are routine in the art and can be employed in the methods of the invention provided herein.

By "stably introducing" or "stably introduced" in the context of a polynucleotide introduced into a cell is intended the introduced polynucleotide is stably incorporated into the genome of the cell, and thus the cell is stably transformed with the polynucleotide.

"Stable transformation" or "stably transformed" as used herein means that a nucleic acid is introduced into a cell and integrates into the genome of the cell. As such, the integrated nucleic acid is capable of being inherited by the progeny thereof, more particularly, by the progeny of multiple successive generations. "Genome" as used herein also includes the nuclear and the plastid genome, and therefore includes integration of the nucleic acid into, for example, the chloroplast genome. Stable transformation as used herein can also refer to a transgene that is maintained extrachromasomally, for example, as a minichromosome.

Stable transformation of a cell can be detected by, for example, a Southern blot hybridization assay of genomic DNA of the cell with nucleic acid sequences which specifically hybridize with a nucleotide sequence of a transgene introduced into an organism (e.g., a plant). Stable transformation of a cell can be detected by, for example, a Northern blot hybridization assay of RNA of the cell with nucleic acid sequences which specifically hybridize with a nucleotide sequence of a transgene introduced into a plant or other organism. Stable transformation of a cell can also be detected by, *e.g.,* a polymerase chain reaction (PCR) or other amplification reactions as are well known in the art, employing specific primer sequences that hybridize with target sequence(s) of a transgene, resulting in amplification of the transgene sequence, which can be detected according to standard methods Transformation can also be detected by direct sequencing and/or hybridization protocols well known in the art.

The "transformation and regeneration process" refers to the process of stably introducing a transgene into a plant cell and regenerating a plant from the transgenic plant cell. As used herein, transformation and regeneration includes the selection process, whereby a transgene comprises a selectable marker and the transformed cell has incorporated and expressed the transgene, such that the transformed cell will survive and developmentally flourish in the presence of the selection agent. "Regeneration" refers to growing a whole plant from a plant cell, a group of plant cells, or a plant piece such as from a protoplast, callus, or tissue part.

As used in the description of the embodiments of the invention and the appended claims, the singular forms "a," "an," and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise.

As used herein, "and/or" refers to and encompasses any and all possible combinations of one or more of the associated listed items.

The term "about," as used herein when referring to a measurable value such as an amount of a compound, dose, time, temperature, and the like, is meant to encompass variations of 20%, 10%, 5%, 1%, 0.5%, or even 0.1% of the specified amount.

The terms "comprise," "comprises" and/or "comprising," when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof.

As used herein, the transitional phrase "consisting essentially of means that the scope of a claim is to be interpreted to encompass the specified materials or steps recited in the claim and those that do not materially affect the basic and novel characteristic(s) of the claimed invention. Thus, the term "consisting essentially of when used in a claim of this invention is not intended to be interpreted to be equivalent to "comprising."

The terms "nucleotide sequence" "nucleic acid," "nucleic acid sequence," "nucleic acid molecule," "oligonucleotide" and "polynucleotide" are used interchangeably herein to refer to a heteropolymer of nucleotides and encompass both RNA and DNA, including cDNA, genomic DNA, mRNA, synthetic *(e.g.,* chemically synthesized) DNA or RNA and chimeras of RNA and DNA. The term nucleic acid molecule refers to a chain of nucleotides without regard to length of the chain. The nucleotides contain a sugar, phosphate and a base which is either a purine or pyrimidine. A nucleic acid molecule can be double-stranded or single-stranded. Where single-stranded, the nucleic acid molecule can be a sense strand or an antisense strand. A nucleic acid molecule can be synthesized using oligonucleotide analogs or derivatives *(e.g.,* inosine or phosphorothioate nucleotides). Such oligonucleotides can be used, for example, to prepare nucleic acid molecules that have altered base-pairing abilities or increased resistance to nucleases. Nucleic acid sequences provided herein are presented herein in the 5' to 3' direction, from left to right and are represented using the standard code for representing the nucleotide characters as set forth in the U.S. sequence rules, 37 CFR §§1.821 - 1.825 and the World Intellectual Property Organization (WIPO) Standard ST.25.

A "nucleic acid fragment" is a fraction of a given nucleic acid molecule. In higher plants, deoxyribonucleic acid (DNA) is the genetic material while ribonucleic acid (RNA) is involved in the transfer of information contained within DNA into proteins. A "genome" is the entire body of genetic material contained in each cell of an organism. Unless otherwise indicated, a particular nucleic acid sequence of this invention also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences and as well as the sequence explicitly indicated. Specifically,

degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al., *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka et al., *J. Biol. Chem.* 260:2605-2608 (1985); and Rossolini et al., *Mol. Cell. Probes* 8:91-98 (1994)).

The term nucleic acid molecule is used interchangeably with gene, cDNA, and mRNA encoded by a gene.

As used herein, the term "gene" refers to a nucleic acid molecule capable of being used to produce mRNA, antisense RNA, miRNA, and the like. Genes may or may not be capable of being used to produce a functional protein. Genes can include both coding and non-coding regions (e.g., introns, regulatory elements, promoters, enhancers, termination sequences and 5' and 3' untranslated regions). In some embodiments, a gene refers to only the coding region. A gene may be "isolated" by which is meant a nucleic acid molecule that is substantially or essentially free from components normally found in association with the nucleic acid molecule in its natural state. Such components include other cellular material, culture medium from recombinant production, and/or various chemicals used in chemically synthesizing the nucleic acid molecule.

As used herein "sequence identity" refers to the extent to which two optimally aligned polynucleotide or peptide sequences are invariant throughout a window of alignment of components, *e.g.*, nucleotides or amino acids. "Identity" can be readily calculated by known methods including, but not limited to, those described in: *Computational Molecular Biology* (Lesk, A. M., ed.) Oxford University Press, New York (1988); *Biocomputing: Informatics and Genome Projects* (Smith, D. W., ed.) Academic Press, New York (1993); *Computer Analysis of Sequence Data, Part I* (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, New Jersey (1994); Sequence Analysis in Molecular Biology (von Heinje, G., ed.) Academic Press (1987); and Sequence Analysis Primer (Gribskov, M. and Devereux, J., eds.) Stockton Press, New York (1991).

As used herein, the term "percent sequence identity" or "percent identity" refers to the percentage of identical nucleotides in a linear polynucleotide sequence of a reference ("query") polynucleotide molecule (or its complementary strand) as compared to a test ("subject") polynucleotide molecule (or its complementary strand) when the two sequences are optimally aligned. In some embodiments, "percent identity" can refer to the percentage of identical amino acids in an amino acid sequence.

As used herein, the phrase "substantially identical," in the context of two nucleic acid molecules, nucleotide sequences or protein sequences, refers to two or more sequences or

subsequences that have at least about 70%, least about 75%, at least about 80%, least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. In some embodiments of the invention, the substantial identity exists over a region of the sequences that is at least about 50 residues to about 150 residues in length. Thus, in some embodiments of this invention, the substantial identity exists over a region of the sequences that is at least about 50, about 60, about 70, about 80, about 90, about 100, about 110, about 120, about 130, about 140, about 150, or more residues in length. In some particular embodiments, the sequences are substantially identical over at least about 150 residues. In a further embodiment, the sequences are substantially identical over the entire length of the coding regions. Furthermore, in representative embodiments, substantially identical nucleotide or protein sequences perform substantially the same function (e.g., conferring increased resistance to a nematode plant parasite, reducing the growth of a nematode plant parasite, reducing cyst development).

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

Optimal alignment of sequences for aligning a comparison window are well known to those skilled in the art and may be conducted by tools such as the local homology algorithm of Smith and Waterman, the homology alignment algorithm of Needleman and Wunsch, the search for similarity method of Pearson and Lipman, and optionally by computerized implementations of these algorithms such as GAP, BESTFIT, FASTA, and TFASTA available as part of the GCG® Wisconsin Package® (Accelrys Inc., San Diego, CA). An "identity fraction" for aligned segments of a test sequence and a reference sequence is the number of identical components which are shared by the two aligned sequences divided by the total number of components in the reference sequence segment, *i.e.*,, the entire reference sequence or a smaller defined part of the reference sequence. Percent sequence identity is represented as the identity fraction multiplied by 100. The comparison of one or more polynucleotide sequences may be to a full-length polynucleotide sequence or a portion thereof, or to a longer polynucleotide sequence. For purposes of this invention "percent

identity" may also be determined using BLASTX version 2.0 for translated nucleotide sequences and BLASTN version 2.0 for polynucleotide sequences.

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al,* 1990). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix *(see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89: 10915 (1989)).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences *(see, e.g.,* Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90: 5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleotide sequence to the reference nucleotide sequence is less than about 0.1 to less than about 0.001. Thus, in some embodiments of the invention, the smallest sum probability in a comparison of the test nucleotide sequence to the reference nucleotide sequence is less than about 0.001.

Two nucleotide sequences can also be considered to be substantially identical when the two sequences hybridize to each other under stringent conditions. In some representative embodiments, two nucleotide sequences considered to be substantially identical hybridize to each other under highly stringent conditions.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern hybridizations are sequence dependent, and are different under different environmental parameters. An extensive guide to the hybridization of nucleic acids is found in Tijssen *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes* part I chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays" Elsevier, New York (1993). Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH.

The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent hybridization conditions for hybridization of complementary nucleotide sequences which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formamide with 1 mg of heparin at 42°C, with the hybridization being carried out overnight. An example of highly stringent wash conditions is 0.15M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes *{see,* Sambrook, *infra,* for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example of a medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1x SSC at 45°C for 15 minutes. An example of a low stringency wash for a duplex of, *e.g.,* more than 100 nucleotides, is 4-6x SSC at 40°C for 15 minutes. For short probes *{e.g.,* about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30°C. Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Nucleotide sequences that do not hybridize to each other under stringent conditions are still substantially identical if the proteins that they encode

are substantially identical. This can occur, for example, when a copy of a nucleotide sequence is created using the maximum codon degeneracy permitted by the genetic code.

The following are examples of sets of hybridization/wash conditions that may be used to clone homologous nucleotide sequences that are substantially identical to reference nucleotide sequences of the present invention. In one embodiment, a reference nucleotide sequence hybridizes to the "test" nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaP0_4$, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C. In another embodiment, the reference nucleotide sequence hybridizes to the "test" nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaP0_4$, 1 mM EDTA at 50°C with washing in IX SSC, 0.1% SDS at 50°C or in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaP0_4$, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C. In still further embodiments, the reference nucleotide sequence hybridizes to the "test" nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaP0_4$, 1 mM EDTA at 50°C with washing in 0.IX SSC, 0.1% SDS at 50°C, or in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaP0_4$, 1 mM EDTA at 50°C with washing in 0.IX SSC, 0.1% SDS at 65°C.

An "isolated" nucleic acid molecule or nucleotide sequence or an "isolated" polypeptide is a nucleic acid molecule, nucleotide sequence or polypeptide that, by the hand of man, exists apart from its native environment and/or has a function that is different, modified, modulated and/or altered as compared to its function in its native environment and is therefore not a product of nature. An isolated nucleic acid molecule or isolated polypeptide may exist in a purified form or may exist in a non-native environment such as, for example, a recombinant host cell. Thus, for example, with respect to polynucleotides, the term isolated means that it is separated from the chromosome and/or cell in which it naturally occurs. A polynucleotide is also isolated if it is separated from the chromosome and/or cell in which it naturally occurs and is then inserted into a genetic context, a chromosome, a chromosome location, and/or a cell in which it does not naturally occur. The recombinant nucleic acid molecules and nucleotide sequences of the invention can be considered to be "isolated" as defined above.

Thus, an "isolated nucleic acid molecule" or "isolated nucleotide sequence" is a nucleic acid molecule or nucleotide sequence that is not immediately contiguous with nucleotide sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally occurring genome of the organism from which it is derived. Accordingly, in one embodiment, an isolated nucleic acid includes some or all of the 5' non-coding (e.g., promoter) sequences that are immediately contiguous to a coding sequence. The

term therefore includes, for example, a recombinant nucleic acid that is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment), independent of other sequences. It also includes a recombinant nucleic acid that is part of a hybrid nucleic acid molecule encoding an additional polypeptide or peptide sequence. An "isolated nucleic acid molecule" or "isolated nucleotide sequence" can also include a nucleotide sequence derived from and inserted into the same natural, original cell type, but which is present in a non-natural state, e.g., present in a different copy number, and/or under the control of different regulatory sequences than that found in the native state of the nucleic acid molecule.

The term "isolated" can further refer to a nucleic acid molecule, nucleotide sequence, polypeptide, peptide or fragment that is substantially free of cellular material, viral material, and/or culture medium (e.g., when produced by recombinant DNA techniques), or chemical precursors or other chemicals (e.g., when chemically synthesized). Moreover, an "isolated fragment" is a fragment of a nucleic acid molecule, nucleotide sequence or polypeptide that is not naturally occurring as a fragment and would not be found as such in the natural state. "Isolated" does not necessarily mean that the preparation is technically pure (homogeneous), but it is sufficiently pure to provide the polypeptide or nucleic acid in a form in which it can be used for the intended purpose.

In representative embodiments of the invention, an "isolated" nucleic acid molecule, nucleotide sequence, and/or polypeptide is at least about 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% pure (w/w) or more. In other embodiments, an "isolated" nucleic acid, nucleotide sequence, and/or polypeptide indicates that at least about a 5-fold, 10-fold, 25-fold, 100-fold, 1000-fold, 10,000-fold, 100,000-fold or more enrichment of the nucleic acid (w/w) is achieved as compared with the starting material.

"Wild-type" nucleotide sequence or amino acid sequence refers to a naturally occurring ("native ") or endogenous nucleotide sequence or amino acid sequence. Thus, for example, a "wild-type mRNA" is an mRNA that is naturally occurring in or endogenous to the organism. A "homologous" nucleotide sequence is a nucleotide sequence naturally associated with a host cell into which it is introduced.

By the term "express" or "expression" of a polynucleotide coding sequence, it is meant that the sequence is transcribed, and optionally translated.

"Nucleotide sequence of interest" refers to any nucleotide sequence which, when introduced into a plant, confers upon the plant a desired characteristic such as antibiotic resistance, virus resistance, insect resistance, disease resistance, or resistance to other pests, herbicide tolerance, improved nutritional value, improved performance in an industrial process or altered reproductive capability. The "nucleotide sequence of interest" may also be one that is transferred to plants for the production of commercially valuable enzymes or metabolites in the plant.

As used herein, the phrases "operably linked," "operatively linked," "operatively associated" or "in operative association" and the like, mean that elements of a nucleic acid construct such as an expression cassette or nucleic acid molecule are configured so as to perform their usual function. Thus, regulatory or control sequences *(e.g.,* promoters) operatively associated with a nucleotide sequence are capable of effecting expression of the nucleotide sequence. For example, a promoter in operative association with a nucleotide sequence encoding miR396c would be capable of effecting the expression of that miR396c nucleotide sequence.

The control sequences need not be contiguous with the nucleotide sequence of interest, as long as they function to direct the expression thereof. Thus, for example, intervening untranslated, yet transcribed, sequences can be present between a promoter and a coding sequence, and the promoter sequence can still be considered "operably linked" to the coding sequence.

As used herein, the terms "transformed" and "transgenic" refer to any plant, plant cell, callus, plant tissue, or plant part that contains all or part of at least one recombinant (e.g., heterologous) polynucleotide. In some embodiments, all or part of the recombinant polynucleotide is stably integrated into a chromosome or stable extra-chromosomal element, so that it is passed on to successive generations. For the purposes of the invention, the term "recombinant polynucleotide" refers to a polynucleotide that has been altered, rearranged, or modified by genetic engineering. Examples include any cloned polynucleotide, or polynucleotides, that are linked or joined to heterologous sequences. The term "recombinant" does not refer to alterations of polynucleotides that result from naturally occurring events, such as spontaneous mutations, or from non-spontaneous mutagenesis followed by selective breeding.

The term "introducing" or "introduce" in the context of a plant cell, plant and/or plant part means contacting a nucleic acid molecule with the plant, plant part, and/or plant cell in such a manner that the nucleic acid molecule gains access to the interior of the plant cell

and/or a cell of the plant and/or plant part. Where more than one nucleic acid molecule is to be introduced these nucleic acid molecules can be assembled as part of a single polynucleotide or nucleic acid construct, or as separate polynucleotide or nucleic acid constructs, and can be located on the same or different nucleic acid constructs. Accordingly, these polynucleotides can be introduced into plant cells in a single transformation event, in separate transformation events, or, e.g., as part of a breeding protocol. Thus, the term "transformation" as used herein refers to the introduction of a heterologous nucleic acid into a cell. Transformation of a cell may be stable or transient. Thus, a transgenic plant cell, plant and/or plant part of the invention can be stably transformed or transiently transformed.

The term "plant part," as used herein, includes but is not limited to embryos, pollen, ovules, seeds, leaves, stems, shoots, flowers, branches, fruit, kernels, ears, cobs, husks, stalks, roots, root tips, anthers, plant cells including plant cells that are intact in plants and/or parts of plants, plant protoplasts, plant tissues, plant cell tissue cultures, plant calli, plant clumps, and the like. As used herein, "shoot" refers to the above ground parts including the leaves and stems. Further, as used herein, "plant cell" refers to a structural and physiological unit of the plant, which comprises a cell wall and also may refer to a protoplast. A plant cell of the present invention can be in the form of an isolated single cell or can be a cultured cell or can be a part of a higher-organized unit such as, for example, a plant tissue or a plant organ.

"Transient transformation" in the context of a polynucleotide means that a polynucleotide is introduced into the cell and does not integrate into the genome of the cell.

As used herein, "stably introducing," "stably introduced," "stable transformation" or "stably transformed" in the context of a polynucleotide introduced into a cell, means that the introduced polynucleotide is stably integrated into the genome of the cell, and thus the cell is stably transformed with the polynucleotide. As such, the integrated polynucleotide is capable of being inherited by the progeny thereof, more particularly, by the progeny of multiple successive generations. "Genome" as used herein includes the nuclear and/or plastid genome, and therefore includes integration of a polynucleotide into, for example, the chloroplast genome. Stable transformation as used herein can also refer to a polynucleotide that is maintained extrachromasomally, for example, as a minichromosome.

Transient transformation may be detected by, for example, an enzyme-linked immunosorbent assay (ELISA) or Western blot, which can detect the presence of a peptide or polypeptide encoded by one or more nucleic acid molecules introduced into an organism. Stable transformation of a cell can be detected by, for example, a Southern blot hybridization assay of genomic DNA of the cell with nucleic acid sequences which specifically hybridize

with a nucleotide sequence of a nucleic acid molecule introduced into an organism (e.g., a plant). Stable transformation of a cell can be detected by, for example, a Northern blot hybridization assay of RNA of the cell with nucleic acid sequences which specifically hybridize with a nucleotide sequence of a nucleic acid molecule introduced into a plant or other organism. Stable transformation of a cell can also be detected by, *e.g.,* a polymerase chain reaction (PCR) or other amplification reaction as are well known in the art, employing specific primer sequences that hybridize with target sequence(s) of a nucleic acid molecule, resulting in amplification of the target sequence(s), which can be detected according to standard methods. Transformation can also be detected by direct sequencing and/or hybridization protocols well known in the art.

The terms "open reading frame" and "ORF" refer to the amino acid sequence encoded between translation initiation and termination codons of a coding sequence. The terms "initiation codon" and "termination codon" refer to a unit of three adjacent nucleotides ('codon') in a coding sequence that specifies initiation and chain termination, respectively, of protein synthesis (mRNA translation).

"Promoter" refers to a nucleotide sequence, usually upstream (5') to its coding sequence, which controls the expression of the coding sequence by providing the recognition for RNA polymerase and other factors required for proper transcription. "Promoter regulatory sequences" consist of proximal and more distal upstream elements. Promoter regulatory sequences influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences include enhancers, promoters, untranslated leader sequences, introns, and polyadenylation signal sequences. They include natural and synthetic sequences as well as sequences that may be a combination of synthetic and natural sequences. An "enhancer" is a DNA sequence that can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue specificity of a promoter. It is capable of operating in both orientations (normal or flipped), and is capable of functioning even when moved either upstream or downstream from the promoter. The meaning of the term "promoter" includes "promoter regulatory sequences."

"Primary transformant" and "TO generation" refer to transgenic plants that are of the same genetic generation as the tissue that was initially transformed (i.e.,, not having gone through meiosis and fertilization since transformation). "Secondary transformants" and the "Tl, T2, T3, etc. generations" refer to transgenic plants derived from primary transformants through one or more meiotic and fertilization cycles. They may be derived by self-

fertilization of primary or secondary transformants or crosses of primary or secondary transformants with other transformed or untransformed plants.

"Gene" refers to a nucleic acid fragment that expresses mRNA, functional RNA, or specific protein, including regulatory sequences. The term "native gene" refers to a gene as found in nature. The term "chimeric gene" refers to any gene that contains 1) DNA sequences, including regulatory and coding sequences that are not found together in nature, or 2) sequences encoding parts of proteins not naturally adjoined, or 3) parts of promoters that are not naturally adjoined. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or comprise regulatory sequences and coding sequences derived from the same source, but arranged in a manner different from that found in nature.

A "transgene" refers to a nucleic acid molecule that has been introduced into the genome by transformation and is stably maintained. A transgene may comprise at least one expression cassette, typically comprises at least two expression cassettes, and may comprise ten or more expression cassettes. Transgenes may include, for example, genes that are either heterologous or homologous to the genes of a particular plant to be transformed. Additionally, transgenes may comprise native genes inserted into a non-native organism, or chimeric genes. The term "endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism but one that is introduced into the organism by gene transfer.

"Intron" refers to an intervening section of DNA which occurs almost exclusively within a eukaryotic gene, but which is not translated to amino acid sequences in the gene product. The introns are removed from the pre- mature mRNA through a process called splicing, which leaves the exons untouched, to form an mRNA. For purposes of the present invention, the definition of the term "intron" includes modifications to the nucleotide sequence of an intron derived from a target gene, provided the modified intron does not significantly reduce the activity of its associated 5' regulatory sequence.

"Exon" refers to a section of DNA which carries the coding sequence for a protein or part of it. Exons are separated by intervening, non- coding sequences (introns). For purposes of the present invention, the definition of the term "exon" includes modifications to the nucleotide sequence of an exon derived from a target gene, provided the modified exon does not significantly reduce the activity of its associated 5' regulatory sequence.

Accordingly, in one embodiment, the present invention provides a method of integrating a transgene into a genomic nuclease cleavage site in a maize genome, comprising introducing into a maize cell: a) a first nucleic acid molecule comprising at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, or at least 150 contiguous nucleotides, wherein said contiguous nucleotides have at least 80% , at least 85%, at least 90%, at least 93%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity with a target site in the nucleotide sequence of SEQ ID NO: 1 or the nucleotide sequence of SEQ ID NO:2, and further comprising a transgene; and b) a second nucleic acid molecule comprising a nucleotide sequence encoding a nuclease for site-directed cleavage at a genomic nuclease cleavage site adjacent to the nucleotide sequence of SEQ ID NO: 1 or the nucleotide sequence of SEQ ID NO:2 that corresponds to the contiguous nucleotides of (a), under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave the nucleotide sequence at the genomic nuclease cleavage site, whereby the transgene is integrated at the genomic nuclease target cleavage site in the maize genome.

As used herein, a "target site" means a region of nucleotides in the genome that is the selected or preferred site for insertion of a nucleotide sequence (e.g., one or more transgenes, expression cassettes, or nucleotide sequences of interest) into the genome as well as a selected or preferred site for introducing a mutation (e.g., a substitution and/or a deletion, and/or an insertion such as an INDEL) into the genome. In some embodiments, a target site can comprise a nuclease cleavage site, also referred to as a genomic nuclease cleavage site. A nonlimiting example of a target site of this invention is the chromosome interval on chromosome 1 defined by and including base pair (bp) position 38,860,000 to base pair (bp) position 39,105,000 as defined by Maize B73 RefGen_V2 available in the Maize Genome Database.

As used herein, the terms "adjacent" or "adjacent to" with regard to one or more nucleotide sequences of this invention means immediately next to (e.g., with no intervening sequence) or separated by from about 1 base to about 10,000 bases (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 100, 200, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10,000 bases), including any values included within this range but not explicitly recited herein.

A "nuclease cleavage site" or "genomic nuclease cleavage site" is a region of nucleotides that comprise a nuclease cleavage sequence that is recognized by a specific nuclease, which acts to cleave the nucleotide sequence of the genomic DNA in one or both

strands. Such cleavage by the nuclease enzyme initiates DNA repair mechanisms within the cell, which establishes an environment for homologous recombination to occur. In the methods herein wherein the first nucleic acid molecule comprises, for example, at least about 100 contiguous nucleotides having, for example, at least 90% identity with a target site in the genome of the cell, the first nucleic acid molecule is integrated into the genome of the cell via homologous recombination, thereby integrating the one or more transgenes into the genome of the cell.

In some embodiments of the above method, the first nucleic acid molecule can comprise at least about 100, 150, 200, 250, 300, 250, 400, 450, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 7500, 10000, 15,000 or 20,000 nucleotides, including any value within this range not explicitly recited herein.

In some embodiments of the above method, the nucleotide sequence comprising the genomic nuclease cleavage site in the maize genome can be the nucleotide sequence of SEQ ID NO:l, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:28, SEQ ID NO:66, or SEQ ID NO:67.

In some embodiments of the above method, the genomic nuclease cleavage site is located within a chromosome interval on chromosome 1 defined by and including base pair (bp) position 38,860,000 to base pair (bp) position 39,015,000 as defined by Maize B73 RefGen_V2, available in the Maize Genome Database.

In some embodiments of the method above, the nuclease has cleavage specificity for a nuclease cleavage site in the nucleotide sequence selected from the group consisting of SEQ ID NO:l (HiII-MIR604), SEQ ID NO:2 (AX-MIR604), SEQ ID NO:3, SEQ ID NO:28, SEQ ID NO: 66, SEQ ID NO: 67 and any combination thereof.

In some embodiments, the first nucleic acid molecule and the second nucleic acid molecule can be introduced into the maize cell by biolistic nucleic acid delivery, via an *Agrobacterium*, by co-transformation, and/or with a T-DNA vector in any combination and/or order.

In some embodiments, the first nucleic acid molecule and the second nucleic acid molecule can be present on a single nucleic acid construct and in some embodiments, the first nucleic acid molecule and the second nucleic acid molecule can be present on separate nucleic acid constructs.

In some embodiments, the first nucleic acid molecule and/or the second nucleic acid molecule can be transiently expressed in the maize cell.

In some embodiments, the first nucleic acid molecule and/or the second nucleic acid molecule can be stably integrated into the maize genome in the maize cell.

The present invention further provides a method of producing a maize plant, plant part, or progeny thereof comprising a transgene integrated into the genomic nuclease cleavage site in the maize genome, comprising regenerating a maize plant from the maize cell produced by the method described herein. Accordingly, the present invention provides a maize plant, plant part, or progeny thereof comprising the transgene integrated into the genomic nuclease cleavage site in the maize genome, produced by the method of this invention.

The present invention is based in some embodiments on the unexpected discovery and development of rapid (e.g., high throughput) methods to identify and enrich for cells that comprise one or more transgenes integrated into the genome at a target site that employ selective combinations of quantitative polymerase chain reaction (qPCR) assays.

The present invention further provides a method of identifying a cell and/or enriching for a cell comprising a transgene inserted into a nuclease cleavage site in a genome of the cell, comprising: a) introducing into a plurality of cells: i) a first nucleic acid molecule comprising at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, or at least 150 contiguous nucleotides,, wherein the contiguous nucleotides have at least 80% , at least 85%, at least 90%, at least 93%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity with a target site in the genome of the cell, and further comprising a transgene; and ii) a second nucleic acid molecule encoding a nuclease for site-directed cleavage at a nuclease cleavage site in the genome of the cell adjacent to the nucleotide sequence in the genome of the cell that corresponds to the contiguous nucleotides of (i), under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave at the nuclease cleavage site in the genome of the cell and integrate the transgene into the nuclease cleavage site in the genome of the cell; b) culturing the cells of (a) to produce at least one cell line or tissue; c) extracting a genomic DNA sample from each of the cell lines or tissues of (b); d) performing real-time quantitative polymerase chain reaction (qPCR) assays T and G on the samples of (c), wherein the assays T and G respectively comprise the following probes: i) a first probe comprising a nucleotide sequence that is complementary to a nucleotide sequence of the target site, at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least twelve, or at least fifteen base pairs away from the nuclease cleavage site for carrying out assay T, and ii) a second probe comprising a nucleotide sequence that is complementary to a nucleotide sequence of the transgene for carrying out assay G; e) obtaining a DNA copy number of the target site from the results of

assay T and a DNA copy number of the transgene from the results of assay G; and f) identifying and/or enriching for a cell line or tissue that has reduced copy number in assay T relative to a reference and a copy number greater than zero for assay G, thereby identifying and/or enriching for the cell comprising the transgene inserted into the nuclease cleavage site in the genome of the cell.

In the methods described above directed to identifying and/or enriching for cells that comprise one or more transgenes inserted into a nuclease cleavage site in a genome of the cell, the qPCR assays can be performed in a high-throughput format as is well known in the art, such that a large volume of samples can be assayed rapidly and simultaneously. Such rapid and efficient screening allows for the identification and enrichment for the small percentage of cells (e.g., around 2%) among the plurality of cells employed in these methods, which would typically be a large volume of cells.

In the methods described above, the first probe (for carrying out assay T) can comprise, consist essentially of or consist of a nucleotide sequence that is complementary (e.g., at least about 90%, 95%, 98%, 99% or 100% complementary) to nucleotide sequence at least five (e.g., 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20) base pairs away from the nuclease cleavage site and the second probe (for carrying out assay G) can comprise a nucleotide sequence that is complementary (e.g., at least about 90%, 95%, 98%, 99% or 100% complementary) to at least one of the one or more transgenes.

In some embodiments of the enriching and identifying methods described above, in addition to the step of identifying and/or enriching for a cell line or tissue that has reduced copy number in assay T relative to a reference and a copy number greater than zero (e.g., a copy number of about one, a copy number of about 2, or a copy number of about 3) for assay G, the methods in some embodiments can further comprise the step of discarding a cell line or tissue that has no change in the DNA copy number of assay T in comparison with a reference, and in some embodiments, can further comprise the step of discarding a cell line or tissue that has a copy number of zero (e.g., a copy number of less than one) for assay G.

As used herein, being "positive" or a positive result for an assay (e.g., assay G) means that the copy number is greater than zero and being "negative" for an assay (e.g., assay G) means that the copy number is zero or less than one.

As also used herein, a "reference" is a genome that has a fixed gene copy number. In some embodiments, the reference can be a "wild type" genome (e.g., a genome of a cell that has not had the first and second nucleic acid molecules of this invention introduced into it according to the methods of this invention)

In particular embodiments of the invention, the first and second probes are fluorescence probes and in some embodiments, the first and second probes are Taqman probes.

In some embodiments of the invention, the qPCR assays are performed in the same mixture and in some embodiments, the qPCR assays are performed in different mixtures, in any combination.

In embodiments in which the plant is a maize plant, the nuclease cleavage site is a maize MIR604 transgene insertion site, namely a nucleotide sequence with at least 80%, at least 85%, at least 90%, at least 93%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO:2.

In some embodiments of the methods described herein, the nuclease can be a non-engineered nuclease (e.g., a nuclease in its "native" form or "wild type" form without modifications). In some embodiments, the nuclease can be an engineered nuclease with programmable cleavage target specificity. Non-limiting examples of a nuclease of this invention include CRISPR gRNA-Cas9 nuclease, zinc finger nuclease, engineered meganuclease and/or TAL effector nuclease, singly or in any combination.

The present invention also provides a cell line or tissue that is identified and/or enriched by the methods described herein, wherein the cell line or tissue is derived from a plant or a plant part. In some embodiments, the cell line or tissue is derived from a monocot plant or monocot plant part. In some embodiments, the cell line or tissue is derived from a dicot plant or plant part. In some embodiments, the cell line or tissue is derived from a cereal plant or cereal plant part. In further embodiments, the cell line or tissue is derived from a maize plant or maize plant part. Other nonlimiting examples of a plant of this invention include rice, sugarcane, barley, sugarbeet, potato, tobacco, soybean, tomato, wheat and sunflower.

Further provided herein is a cell line or tissue that is identified and/or enriched by the methods described herein, wherein the cell line or tissue is derived from a eukaryotic organism.

In some embodiments of the enriching and identifying methods described above, in addition to the step of identifying and/or enriching for a cell line or tissue that has reduced copy number in assay T relative to a reference and is positive for assay G, the methods in some embodiments can further comprise the step of discarding a cell line or tissue that has no change in the DNA copy number of assay T in comparison with a reference, and in some

embodiments, can further comprise the step of discarding a cell line or tissue that is negative for assay G.

As used herein, being "positive" for an assay (e.g., assay G) means that the copy number is greater than zero and being "negative" for an assay (e.g., assay G) means that the copy number is equal to zero.

As also used herein, a "reference" is a genome or other nucleic acid molecule that has a fixed gene copy number. In some embodiments, the reference can be a "wild type" genome (e.g., a genome of a cell that has not had the first and second nucleic acid molecules of this invention introduced into it according to the methods of this invention)

In particular embodiments of the invention, the first and second probes are fluorescence probes and in some embodiments the first and second probes are Taqman probes.

In some embodiments of the invention, the qPCR assays are performed in the same mixture and in some embodiments, the qPCR assays are performed in different mixtures, in any combination.

In embodiments in which the plant produced is a maize plant, the nuclease cleavage site is a maize MIR604 transgene insertion site, namely a nucleotide sequence with at least 80%, at least 85%, at least 90%, at least 93%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO:2

Further provided herein is a method of identifying a cell and/or for enriching for a cell comprising a mutation introduced into a nuclease cleavage site in a genome of the cell and lacking integration of a heterologous nucleotide sequence encoding a nuclease for site-directed cleavage of a nucleotide sequence at the nuclease cleavage site into the genome of the cell, comprising: a) introducing a nucleic acid molecule comprising a heterologous sequence encoding a nuclease for site-directed cleavage of the nucleotide sequence at the nuclease cleavage site in the genome of the cell into a plurality of cells under conditions wherein expression of the nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave the nucleotide sequence at the nuclease cleavage site in the genome of the cell, thereby introducing a mutation at the nuclease cleavage site in the genome of the cell without integration of the heterologous nucleotide sequence encoding the nuclease into the genome of the cell; b) culturing the plurality of cells of (a) to produce at least one cell line or tissue; c) extracting a genomic DNA sample from each of the cell lines or tissues of (b); d) performing real-time quantitative polymerase chain reaction (qPCR) assays 1 and 2 on the

samples of (c), wherein the assays respectively comprise the following probes: i) a first probe comprising a nucleotide sequence that is complementary to the nucleotide sequence comprising the nuclease cleavage site to carry out assay 1, and ii) a second probe comprising a nucleotide sequence that is complementary to the heterologous nucleotide sequence encoding the nuclease to carry out assay 2; e) obtaining a DNA copy number of the nuclease cleavage site from the results of assay 1 and a DNA copy number of the heterologous nucleotide sequence encoding the nuclease from the results of assay 2; and f) identifying and/or enriching for a cell line or tissue that has a reduced copy number for assay 1 relative to a reference and a copy number equal to zero for assay 2, thereby identifying and/or enriching for the cell comprising the mutation introduced into the nuclease cleavage site in the genome of the cell and lacking integration of the heterologous nucleotide sequence encoding the nuclease into the genome of the cell.

In some embodiments of the methods described herein, the cell line or tissue may be derived from a plant or plant part, for example a plant derived from tissue culture or germinated seeds. In some embodiments the plant can be a monocot and in some embodiments, the plant can be a dicot. In some embodiments, the plant can be a cereal. In particular embodiments, the plant can be a maize plant. Other nonlimiting examples of a plant of this invention include rice, sugarcane, barley, sugarbeet, potato, tobacco, soybean, tomato, wheat and sunflower. In some embodiments, the cell line or tissue that is identified and/or enriched by the methods described herein is derived from a eukaryotic organism.

In some embodiments of the enriching and identifying methods described above, in addition to the step of identifying and/or enriching for a cell line or tissue that has a reduced copy number in assay 1 relative to a reference and a copy number equal to zero (e.g., is less than one) for assay 2, the methods in some embodiments can further comprise the step of discarding a cell line or tissue that has no change in the DNA copy number of assay 1 relative to a reference, and in some embodiments, can further comprise the step of discarding a cell line or tissue that has a copy number greater than zero (e.g., a copy number of about 1, a copy number of about 2, or a copy number of about 3) for assay 2.

As used herein, being "positive" or a positive result for an assay (e.g., assay 2) means that the copy number is greater than zero (e.g., a copy number of about 1, a copy number of about 2, or a copy number of about 3) and being "negative" for an assay (e.g., assay 2) means that the copy number is equal to zero (e.g., is less than one).

As also used herein, a "reference" is a genome or other nucleic acid molecule that has a fixed gene copy number. In some embodiments, the reference can be a "wild type" genome

(e.g., a genome of a cell that has not had the first and second nucleic acid molecules of this invention introduced into it according to the methods of this invention).

In particular embodiments of the invention, the first and second probes are fluorescence probes and in some embodiments, the first and second probes are Taqman probes.

In some embodiments of the invention, the qPCR assays are performed in the same mixture and in some embodiments, the qPCR assays are performed in different mixtures, in any combination.

In some embodiments of the methods described herein, the tissue can be a plant derived from tissue culture or germinated seeds. In some embodiments the plant can be a monocot and in some embodiments, the plant can be a dicot. In particular embodiments, the plant can be a maize plant. Other nonlimiting examples of a plant of this invention include rice, sugarcane, barley, sugarbeet, potato, tobacco, soybean, tomato, wheat and sunflower.

In some embodiments of the methods described herein, the nuclease can be a non-engineered nuclease (e.g., a nuclease in its "native" form or "wild type" form without modifications). In some embodiments, the nuclease can be an engineered nuclease with programmable cleavage target specificity. Non-limiting examples of a nuclease of this invention include CRISPR gRNA-Cas9 nuclease (for example, a Cas9 nuclease comprising SEQ ID NO: 30) zinc finger nuclease, engineered meganuclease and/or TAL effector nuclease, singly or in any combination.

In embodiments in which the plant is a maize plant, the nuclease cleavage site is a maize MIR604 transgene insertion site, namely a nucleotide sequence with at least 80%, at least 85%, at least 90%, at least 93%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO:2

The present invention additionally provides a kit of reagents and instructions for carrying out the methods and assay of this invention. In some embodiments, a kit or a package comprising the compositions, formulations and/or agents for carrying out the methods of the present invention is provided. For example, a kit may include means for obtaining a cell or tissue, as well as means for obtaining a nucleic acid sample. The kit may also contain reagents for carrying out the steps of the methods of this invention. Such reagents can include site-specific probes and/ or primers that facilitate isolation and biochemical characterization of nucleic acid molecules of this invention. The kit can contain one or more separate containers.

Although the instructional materials, when present, typically comprise written or

printed materials they are not limited to such. Any medium capable of storing such instructions and communicating them to an end user is contemplated by this invention. Such media include, but are not limited to electronic storage media (e.g., magnetic discs, tapes, cartridges, chips), optical media (e.g., CD ROM), and the like. Such media may include addresses to internet sites that provide such instructional materials.

When the components of the kit are provided in one or more liquid solutions, the liquid solution is preferably an aqueous solution, with a sterile aqueous solution being particularly preferred. However, the components of the kit may be provided as dried powder(s). When reagents or components are provided as a dry powder, the powder can be reconstituted by the addition of a suitable solvent. It is envisioned that the solvent may also be provided in another container. For example, wherein the components of the kit are in lyophilized form, the kit may optionally contain a sterile and physiologically acceptable reconstitution medium such as water, saline, buffered saline, and the like.

In some embodiments, the containers of the kit can include at least one vial, test tube, flask, bottle, syringe or other containers, into which the compositions/formulations of the present invention, and any other desired agent, may be placed and suitably aliquoted.

In additional embodiments, the present invention provides a method of producing a plant, plant part, or progeny thereof comprising a mutation introduced at a nuclease cleavage site in a genome of a plant cell and lacking integration of a heterologous nucleotide sequence encoding a nuclease for site-directed cleavage of a nucleotide sequence at the nuclease cleavage site in the genome of the plant cell, comprising: a) introducing into the plant cell a nucleic acid molecule comprising a heterologous nucleotide sequence encoding a nuclease for site-directed cleavage of the nucleotide sequence at the nuclease cleavage site in the genome of the plant cell under conditions wherein expression of the nucleic acid molecule occurs transiently to produce the nuclease and the nuclease can cleave the nucleotide sequence at the nuclease cleavage site in the genome of the plant cell, thereby introducing a mutation at the nuclease cleavage site in the genome of the plant cell without integration of the heterologous nucleotide sequence encoding the nuclease into the genome of the plant cell; and b) regenerating a plant, plant part, or progeny thereof from the plant cell of (a). In another embodiment, the present invention provides for the plant cell produced by the method described above. In a further embodiment, the present invention provides for a plant or plant part regenerated or derived from the plant cell produced by the method described above.

In some embodiments of the method described above, the plant is a monocot. In other embodiments, the plant is a dicot. In some embodiments, the plant is a cereal. In further

embodiments the plant is maize. Other nonlimiting examples of a plant of this invention include rice, sugarcane, barley, sugarbeet, potato, tobacco, soybean, tomato, wheat and sunflower.

In some embodiments of the methods described above, the mutation comprises at least one nucleotide substitution, the deletion of at least one nucleotide, or a combination of substitution, deletion, and/or insertion, such as for example an INDEL.

In some embodiments of the methods described above, the nucleic acid molecule is biolistic nucleic acid delivery, Agrobacterium -mediated transformation, or any method of plant transformation known in the art.

In some embodiments of the methods described above, the nuclease for site-directed cleavage is a non-engineered nuclease. In some embodiments, the nuclease is an engineered nuclease with programmable cleavage target specificity. In some embodiments, the nuclease is a Cas9. In some embodiments, the nuclease is a Cas9 comprising SEQ ID NO: 30.

The present invention additionally provides a method of producing a plant, plant part, or progeny thereof comprising a transgene introduced at a nuclease cleavage site in a genome of a plant cell and lacking integration of a heterologous nucleotide sequence encoding a nuclease for site-directed cleavage of a nucleotide sequence at the nuclease cleavage site in the genome of the plant cell, comprising: a) introducing into the plant cell a nucleic acid molecule comprising a heterologous nucleotide sequence encoding a nuclease for site-directed cleavage of the nucleotide sequence at the nuclease cleavage site in the genome of the plant cell under conditions wherein expression of the nucleic acid molecule occurs transiently to produce the nuclease and the nuclease can cleave the nucleotide sequence at the nuclease cleavage site in the genome of the plant cell, thereby introducing a mutation at the nuclease cleavage site in the genome of the plant cell without integration of the heterologous nucleotide sequence encoding the nuclease into the genome of the plant cell; and b) regenerating a plant, plant part, or progeny thereof from the plant cell of (a). In another embodiment, the present invention provides for the plant cell produced by the method described above. In a further embodiment, the present invention provides for a plant or plant part regenerated or derived from the plant cell produced by the method described above.

In some embodiments of the method described above, the transgene may comprise at least one, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, or ten or more expression cassettes.

In some embodiments of the method described above, the nuclease cleavage site is or is adjacent to a nucleotide sequence that has at least 80%, at least 85%, at least 90%, at least

93%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO:2.

In some embodiments of the methods described above, the plant is a monocot. In other embodiments, the plant is a dicot. In some embodiments, the plant is a cereal. In further embodiments the plant is maize. Other nonlimiting examples of a plant of this invention include rice, sugarcane, barley, sugarbeet, potato, tobacco, soybean, tomato, wheat and sunflower.

In some embodiments of the methods described above, the first nucleic acid molecule and the second nucleic acid molecule are introduced at the same time, for example by co-transformation, biolistic nucleic acid delivery, or Agrobacterium-mediated transformation. In some embodiments, the first nucleic acid molecule and the second nucleic acid molecule are separate molecules. In some embodiments, a single nucleic acid molecule or construct comprises the first nucleic acid molecule and the second nucleic acid molecule described above.

In some embodiments of the methods described above, the nuclease for site-directed cleavage is a non-engineered nuclease. In some embodiments, the nuclease is an engineered nuclease with programmable cleavage target specificity. In some embodiments, the nuclease is a Cas9. In some embodiments, the nuclease is a Cas9 comprising SEQ ID NO: 30.

The present invention additionally provides a method for modifying a target site in the genome of a plant cell, comprising: a) introducing into the plant cell a first nucleic acid comprising at least 100 contiguous nucleotides, wherein the at least 100 contiguous nucleotides have at least 90% identity with a target site in the genome of the cell, and further comprising a transgene; and b) a second nucleic acid molecule encoding nuclease for site-directed cleavage at a nuclease cleavage site in the genome of the cell adjacent to the nucleotide sequence in the genome of the cell that corresponds to the at least 100 contiguous nucleotides of (a), wherein the nuclease is a modified Cas9 nuclease comprising SEQ ID NO: 30, under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave at the nuclease cleavage site in the genome of the cell and modify the target site in the genome of the plant cell. In another embodiment, the present invention provides for the plant cell produced by the method described above. In a further embodiment, the present invention provides for a plant or plant part regenerated or derived from the plant cell produced by the method described above.

In some embodiments of the method described above, the plant is a monocot. In other embodiments, the plant is a dicot. In some embodiments, the plant is a cereal. In further

embodiments the plant is maize. In some embodiments, the maize is transgenic. In further embodiments, the transgenic maize is event MIR604. Other nonlimiting examples of a plant of this invention include rice, sugarcane, barley, sugarbeet, potato, tobacco, soybean, tomato, wheat and sunflower.

In some embodiments of the methods described above, the modification of the target site comprises at least one nucleotide substitution, the deletion of at least one nucleotide, or a combination of substitution, deletion, and/or insertion, such as for example an INDEL. In other embodiments, the modification of the target site is an insertion, such as a transgene insertion.

In some embodiments of the methods described above, the nucleic acid molecule is biolistic nucleic acid delivery, Agrobacterium -mediated transformation, or any method of plant transformation known in the art.

The present invention additionally provides a method of integrating a transgene into a genomic nuclease cleavage site in an event MTR604 transgenic maize genome, comprising introducing into an event MIR604 maize cell: a) a first nucleic acid molecule comprising at least 100 contiguous nucleotides, wherein said at least 100 contiguous nucleotides have at least 90% identity with a target site in a nucleotide sequence selected from the group comprising SEQ ID NO: 133, SEQ ID NO: 134, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 137, SEQ ID NO: 138, and SEQ ID NO: 139, and further comprising a transgene; and b) a second nucleic acid molecule comprising a nucleotide sequence encoding a nuclease for site-directed cleavage at a genomic nuclease cleavage site adjacent to a nucleotide sequence with at least 90% identity to a nucleotide sequence selected from the group comprising SEQ ID NO: 133, SEQ ID NO: 134, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 137, SEQ ID NO: 138, and SEQ ID NO: 139, that corresponds to the at least 100 contiguous nucleotides of (a), under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave the nucleotide sequence at the genomic nuclease cleavage site, whereby the transgene is integrated at the genomic nuclease target cleavage site in the maize genome.

The present invention further provides a method of producing a maize plant, plant part, or progeny thereof comprising a transgene integrated into a genomic nuclease cleavage site in an event MIR604 maize genome, comprising regenerating a maize plant from the maize cell produced by the method described in the proceeding paragraph. The present invention further provides a maize plant, plant part, or progeny thereof comprising a

transgene integrated into a genomic nuclease cleavage site in an event MIR604 maize genome, produced by the method described above.

The present invention will now be described with reference to the following examples. It should be appreciated that these examples are not intended to limit the scope of the claims to the invention, but are rather intended to be exemplary of certain embodiments. Any variations in the exemplified methods that occur to the skilled artisan are intended to fall within the scope of the present invention.

EXAMPLES

Example 1. Regions around the MIR604 transgene insertion site as a potential safe harbor

The following artificially defined criteria are used to identify potential maize genomic safe harbor regions that are suitable for targeted transgene integration and stable expression: (1) Regions that contain mostly unique sequences so it is suitable for performing targeted integration mediated by homologous recombination; (2) Regions that are not part of a known functional gene including those encoding for miRNAs; Ideally, these regions should be at least 2 Kb upstream of any known open reading frame or 1 Kb downstream from the 3'-untranslated region (3'-UTR) of a gene; thus integration of transgene will not interrupt any endogenous gene sequences or affect function of neighboring endogenous genes; (3) Regions that are not close to heterochromatic regions with highly repetitive sequences such as pericentromeric regions that may result in unstable expression of transgenes or potential silencing of inserted transgenes; (4) Regions that do not contain known *cis*-acting elements such as enhancers or repressors so that transgene expression pattern and level is altered unexpectedly when inserted. (5) Regions that have empirical data showing good transgene expression.

Several candidate regions are identified using the above criteria in the maize genome, for example, in chromosome 1 between position 38,555,000 and 38,605,000, between position 38,640,000 and 38,715,000, and between position 38,860,000 and 39,015,000 (Maize B73 RefGen_V2). Since commercial transgenic events usually have good transgene expression, insertion sites of commercial events are also examined for their potential to serve as candidate safe harbors. However, almost all of them fail to meet the above criteria except for the root-worm resistance trait event MIR604. Interestingly, the transgene insert in MTR604 happens to be located at Chromosome 1 between position 39,014,056 and

39,014,148 close to the end of position 39,015,000. Regions flanking the MIR604 insertion site is unique in that it is the only one out of the many examined to meet all of the safe harbor criteria. Since MIR604 event has been on market for several years, the region around the insertion site is an ideal candidate as safe harbor for insertion of additional transgenes. However, it is shown before that transgene inserted into the previously generated transgene loci may also lead to expression variation (Day et al. "Transgene integration into the same chromosome location can produce alleles that express at a predictable level or alleles that are differentially silenced *Genes and Develop.* 14:2869-2880 (2000)). It is important to verify the hypothesis that the MTR604 insertion site region is a good safe harbor for expression of new transgene alleles at the same locus created via site-directed transformation using different site-directed nucleases and delivery approaches.

Example 2. Cloning of the genomic sequences flanking the MIR604 insertion site in Hill

The MTR604 transgenic event was generated from binary vector pNOV2130 using *Agrobacterium-mediated* transformation of A188 maize immature embryos using mannose as selection. Maize varieties containing MIR604 transgene are widely grown in the United States. MIR604 event contains single copy insertion of pNOV2130 T-DNA in the maize genome. The sequences of MIR604 insertion site and its flanking regions are described in US Patent No. 8,354,519, incoroporated in its entirety herein, and are as in SEQ ID NO:l.

Example 3. Cloning of chromosomal sequences corresponding to the safe harbor locus 1 (MIR604 insertion site) sequences from a transformable elite maize variety NP2222

The original MIR604 insertion site sequences (SEQ ID. NO:l) were derived from non-elite transformation variety A188. It is desirable to insert the transgene directly into an elite transformation variety. However, the sequences from the elite transformation target variety might be different than from A188 and thus will not be recognized by site-directed nucleases designed using the A188 genomic sequences. To obtain genomic sequences corresponding to the MIR604 insertion site flanking regions in the elite maize transformation variety NP2222 (US Patent No. 9,133,474, incorporated by reference herein), PCR primers were designed based on A188 MTR604 insertion site flanking sequences and used to amplify corresponding regions from NP2222. Amplified sequences were sequenced and assembled into a contig which was used for assembly of Hi-Seq whole genome deep sequencing reads around the insertion site. Finally, the NP2222 genomic sequences, named AX_MTR604, which corresponds to the A188 MTR604 insertion site were obtained and are as in SEQ ID

NO:2. Sequence comparison shows that there are significant differences in the genomic sequences between NP2222 and A188, including many InDels (insertions/deletions) and nucleotide substitutions.

Example 4. Targeted insertion of transgenes at the MIR604 insertion site safe harbor mediated by programmable CRISPR-Cas9 nuclease

Example 4.1. Introduction to CRISPR-Cas9 nucleases for mediating targeted insertion

Targeted insertion of transgenic sequences for replacing short stretches of DNA sequences (allele replacement) or inserting large DNA fragments (transgene insertion) can be mediated by DNA breaks introduced by CRISPR-Cas9 nucleases via homologous recombination (Shan et al., *Nature Biotechnology* 31:686-688 (2013); Wang et al., *Cell* 153:910-918 (2013); Yang et al., *Cell* 154:1370-1379 (2013); Puchta and Fauser, *Plant Journal* 78:727-741 (2014); Chen and Gao, *Plant Cell Rep.* 33:575-583 (2014)). In this example, CRISPR-Cas9 nucleases are used to mediate the insertion of large DNA molecules into the desired chromosomal safe harbor target in corn plants. The MTR604 event insertion site in NP2222 corn line was chosen as the tentative transgene expression safe harbor for studying Cas9/gRNA- mediated transgene insertion.

Example 4.2. Candidate safe harbor (MIR604) target sequence selection

The putative safe harbor regions at and surrounding the MIR604 insertion site are scanned for potential Cas9 cleavage sites by using the rule of 5'-G/A-(N)i8-2o -NGG-3' in both strands so that the target template sequences $A(N)_{18-20}$ and $G(N)i_{8-2o}$ preceding the 5'-NGG-3' sequence motif can be conveniently placed under the control of a DNA PolIII promoter such as rice prOsU3 and prOsU6, respectively. Many sequences can be identified as potential Cas9-gRNA cleavage targets around the MTR604 insertion site. For example, the following potential target sequences were identified for Cas9-mediated cleavage: 5'-AGTGC AGTGC AGTGC AGGAC AGG- 3' (SEQ ID. NO:3), 5'-ACTAA TCGTG CTTCA CGCAC AGG-3'(SEQ ID. NO:4), 5'-AGGCA CAGCA CGTAG TAGAC AGG- 3'(SEQ ID. NO:5); 5'-ACATG TCGAT CCGAC GACGA CGG-3'(SEQ ID. NO:6), 5'-AGTTT TATTA TAATC CGAA ACGG- 3'(SEQ ID. NO:7), 5'-AATCC GAAAC GGAGC ACGCA CGG-3[1](SEQ ID. NO: 8), 5'-AAACG GAGCA CGCAC GGCGG TGG- 3'(SEQ ID. NO:9), 5'-GGAGC ACGCA CGGCG GTGG AGG-3'(SEQ ID. NO: 10), 5'-ATCCA AAGCT ACATC CGTGC

AGG-3'(SEQ ID. NO: 11), 5'-GTGCA GTGCA GTGCA GTGC AGG-3'(SEQ ID. NO: 12), 5'-GGACA GGACC TCCTT TGTTT AGG-3'(SEQ ID. NO: 13), 5'-GCGTG CGCAG AGCGC CTGCT CGG-3'(SEQ ID. NO: 14), 5'-GCGTC ATCCA TGTGT TC TGG- 3'(SEQ ID. NO:15), 5'-GTCCA TCTCC ATTCA CTGGT T CGG-3'(SEQ ID. NO: 16), 5'-AATGC CTGCA GAAGA GGCCG TGG-3'(SEQ ID. NO: 17). Similarly, target sequences from the other strand were also identified, for example: 5'-GCGGC CGGCA CGTTG CTAAC C AGG-3'(SEQ ID. NO: 18), 5'-AGAGA AGAAA AATTC GTCCA TGG-3'(SEQ ID. NO:19) , 5'-GGCCT CTTCT GCAGG CATT TGG-3'(SEQ ID. NO:20), 5'-AAGGA ACCCG AACCA GTGAA TGG-3'(SEQ ID. NO:21), 5'-ATCGG TCCTAA ACAAA GG AGG-3'(SEQ ID. NO:22), 5'-GGATG CAGCT TTGGC AACG AGG-3'(SEQ ID. NO:23), 5'-GTCGC GCAGC GCTCC TGCA CGG-3'(SEQ ID. NO:24), 5'-GCTCC TGCAC GGATG TAGCT T TGG-3'(SEQ ID. NO:25), 5'-GGATG TAGCT TTGGA TTGC TGG-3'(SEQ ID. NO:26), 5'-AAATA AAAAA ATCGG ATTAA AGG-3'(SEQ ID. NO:27).

One of the above listed sequences, 5'-AGTGC AGTGC AGTGC AGGAC AGG-3'(SEQ ID NO:3), which is located very close to the MTR604 insert site, was chosen as a target sequence for testing Cas9-gRNA mediated transgene insertion. Sequences (20 bp) preceding the Cas9 recognition PAM motif (5'-NGG-3'), 5'-AGTGC AGTGC AGTGC AGGAC-3' (SEQ ID NO:28, aka. xMIR604FR2) were used to construct sgRNA expression vector using the rice PolIII promoter prOsU3 in the example below.

Example 4.3. CRISPR-Cas9 and guide RNA design and expression vectors

Example 4.3.1. Optimization of Cas9 for expression in maize cells

In order to achieve good expression in maize cells, Type II Cas9 gene from *Streptococcus pyogenes* SF370 was optimized with maize-preferred codons (cBCas9Nu-01, SEQ ID NO:29). A nuclear localization signal was also incorporated into the C-terminus of Cas9 to improve its targeting to nucleus (Cas9Nuc, SEQ ID NO:30). To express the modified Cas9 protein (Cas9Nuc) in maize cells, the maize-optimized Cas9 gene (cBCas9Nu-01, SEQ ID NO:29) was placed under the control of maize ubiquitin-1 promoter (prUbil-10) followed by a terminator sequence (tNOS).

Example 4.3.2. Guide RNAs (gRNAs) for mediating the MIR604 insertion site safe harbor modification: gRNA design and its expression

For targeted cleavage of the safe harbor # 1 (MIR604 insertion site) target sequence (5'-AGTGC AGTGC AGTGC AGGAC AGG-3', SEQ ID NO:3), crRNAs of at least 17

nucleotides (nt) long were designed against the maize genomic target sequence (5'-AGTGC AGTGC AGTGC AGGAC-3', SEQ ID NO:28,) preceding the 5'-NGG-3' for Cas9-mediated target recognition. For example, crRNAs of 17-nt (5'-GC AGTGC AGTGC AGGAC-3', SEQ ID NO:31), 18-nt (5'-TGC AGTGC AGTGC AGGAC-3', SEQ ID NO:32), 19- nt (5'-GTGC AGTGC AGTGC AGGAC-3', SEQ ID NO:33), 20-nt (5'-AGTGC AGTGC AGTGC AGGAC - 3', SEQ ID NO:28) or 21-nt (5'-C AGTGC AGTGC AGTGC AGGAC - 3', SEQ ID NO:34) can be used to guide Cas9 cleavage of the safe harbor #1(MTR604 insertion site). The target crRNA is co-delivered with tracRNA and Cas9 protein or mRNA to mediate target site cleavage. Preferably, the crRNA molecule is fused with tracRNA molecule covalently into a single guide RNA (sgRNA). sgRNAs can be synthesized chemically or produced by *in vitro* transcription. *In vitro* produced sgRNAs can be used directly for physical delivery such as biolistic bombardment with Cas9 RNA or protein to mediate target cleavage and homology-directed target modification if repair donor oligonucleotide is co-delivered. More preferably, sgRNA is produced *in planta* from a DNA expression cassette comprising a RNA polymerase III (PolIII) promoter, for example the rice U3 or U6 promoters (prOsU3 and prOsU6). For prOsU3, the transcriptional start site begins with nucleotide A, whereas for prOsU6, the transcriptional start site begins with nucleotide G (Shan et al., (2013) *Nature Biotechnology* 31: 686-688; Xie and Yang, (2013) *Molecular Plant* 6:1975-1983). For example, to produce sgRNA targeting the safe harbor #1(MTR604 insertion site) sequence (5'-AGTGC AGTGC AGTGC AGGAC AGG-3', SEQ ID NO:3), 19-nt DNA oligonucleotides (5'-GTGC AGTGC AGTGC AGGAC-3', SEQ ID NO:33) or 20-nt oligonucleotides (5'-AGTGC AGTGC AGTGC AGGAC - 3', SEQ ID NO:28) were fused to the DNA sequences encoding tracRNA scaffold and PolIII termination sequences (5'-GTTTT AGAGC TAGAA ATAGC AAGTT AAAAT AAGGC TAGTC CGTTA TCAAC TTGAA AAAGT GGCAC CGAGT CGGTG CTTTT TTTTT-3', SEQ ID NO:35) (Mali et al. (2013). Science 339:823-826) to form coding sequence for a single guide RNA (sgRNA) named rBsgRNA-01 (Seq. ID. NO:36) which was placed under the control of rice polymerase III promoter U3 (prOsU3) or U6 (prOsU6). For this example, the expression cassette comprised prOsU3 and coding sequences for the sgRNA rBsgRNA-01, comprising the 20-nt xMIR604FR2 (SEQ ID NO:28) target RNA fused with tracRNA (SEQ ID NO:37). The expression cassette comprising prOsU3 promoter and rBsgRNA-01 sgRNA was cloned into a biolistic transformation vector along with the Cas9 expression cassette. This biolistic transformation vector is referred to as 22169.

Example 4.4. Generation of targeted insertion events at the MIR604 insertion site safe harbor

Example 4.4.1. Construction donor vector for targeted insertion via homologous recombination

A gene targeting donor vector (referred to as 21942) was constructed by inserting expression cassettes for 2 insect control genes (eCry3. 1Ab and mCry3A) and the PMI selectable marker gene between two homology arms (xJHAX-03, SEQ ID NO:38 and xJHAX-04, SEQ ID NO:39). From the 5' end, the donor nucleic acid sequence comprises xJHAX-03 operably linked to an eCry3. 1Ab expression cassette, which is operably linked to a mCry3A expression cassette, which is operably linked to a cPMI expression cassette, which is operably linked to xJHAX-04 **(Fig. 2).** The two homology arms (xJHAX-03 and xJHAX-04) have sequences identical to part of the safe harbor # 1 (MTR604 insertion site) sequences (SEQ ID NO:2) and are for guiding the targeted insertion of the donor sequences to the Cas9 cleavage site at the target locus using homologous recombination **(Fig. 2).**

Example 4.4.2. Generation of targeted insertion events at the MIR604 insertion site safe harbor with biolistic bombardment

For target gene sequence modification mediated by homology-directed repair, a donor DNA molecule needs to be co-delivered with Cas9 and sgRNA. To generate potential events carrying targeted insertion events at the safe harbor locus #1, plasmid DNA of a vector (22169) carrying an expression cassette for Cas9Nuc and sgRNA was mixed with a fragment of vector 21942 comprising the donor nucleic acid sequence **(Fig. 2),** which comprises the expression cassettes and homology arms described in example 4.4.1. The DNA (Cas9Nuc and sgRNA vector with donor nucleic acid sequence) was then precipitated onto gold particles and used to bombard immature maize embryos (line NP2222). Methods for maize immature embryo bombardment, callus induction tissue regeneration and rooting methods have been described previously (Wright et al., *Plant Cell Reports* 20:429-436 (2001)). Briefly, immature embryos were isolated from harvested immature ears at about 9-1 1 days after pollination and pre-cultured for 1 to 3 days on osmoticum media. Pre-cultured embryos were then bombarded with the DNA described above using BioRad PDS-1000 Biolistic particle delivery system. Bombarded embryos were then incubated in callus induction media and then moved onto mannose selection media. Mannose resistant calli were transferred to regeneration media to induce shoot formation. Shoots were then sub-cultured onto rooting

media. Samples were then harvested from rooted plants for Taqman assays to detect mutations in the target site to enrich for potential targeted insertion events (described herein) and junction PCRs were performed to identify potential plants containing the targeted insertion **(Fig. 2** and **Fig. 3).** Identified putative targeted insertion events were further characterized by more detailed PCR, sequencing and Southern analysis for confirmation **(Fig. 5). Table 1** shows an experiment (MZET134300) that resulted in the recovery of a targeted insertion event MZET134300A679A. In this experiment, more than 80% of transgenic events positive for donor nucleic acid expression cassettes (384 out of 473 events) contain modifications at the target site sequence xMIR604FR2 (SEQ ID NO:28). PCR reactions were performed on a subset of events and identified one clean targeted insertion event through double crossover homologous recombination at both homology arms. Additional DNA sequencing and Southern blot analysis confirmed that the event was a clean targeted insertion event, meaning that this event comprises a single copy of the donor nucleic acid sequence described in example 4.4.1, specifically the eCry3.1Ab, mCry3A, and PMI expression cassettes, is backbone free, shows evidence of a double-crossover homologous recombination event, and has no integration of the vector DNA comprising the nuclease. This Example shows that the MIR604 insertion site is a good target site for targeted insertion.

Table 1. Targeting experiments in corn with sgRNA-Cas9 nuclease at the safe harbor locus #1(MIR604 insertion site) target sequence xMIR604FR2 (SEQ ID NO:28)

Experiment ID	DNA used for bombardment	No. of embryos	Total transgenic events	Events with target site mutation	Events with potential targeted insertion	Events with confirmed clean targeted insertion
MZET134300	22169, 21942 (1:1, 8 $x10^{10}$ molecules of each)	3620	473	384	29	1

To determine the efficiency of sgRNA-Cas9 mediated genome modification, we assayed for the presence of mutations in all 473 transgenic plants described in Table 1, using high throughput Taqman assays as described in the subsequent Examples. Since the transformation is done through co-delivery of repair donor and Cas9 nuclease constructs, we

expect to see donor nucleic acid sequence in transgenic plants that do not contain the Cas9Nuc expression vector. Indeed, out of the 473 PMI-positive plants for donor nucleic acid sequence, 301 of them (63.6%) have and 172 of them (36.4%) do not have co-integrated Cas9 nuclease expression vector, respectively (Table 2). 83 plants (17.5%) without a co-integrated Cas9Nuc nuclease expression vector (22169) have their target site (xMIR604FR2, SEQ ID NO:28) modified either in one allele (7 plants) or both alleles (76 plants) of the maize genome **(Table 2).**

In addition, we assayed for the presence of mutations in regenerated plants that escaped the mannose selection process or transformation escapes that do not contain donor nucleic acid sequence expression cassettes. As expected, out of 471 escapes, only 2 plants are positive for the Cas9Nuc nuclease expression vector and both of these 2 plants have biallelic mutations in the genomic target **(Table 2).** Surprisingly, a high percentage of escape plants (23.9%, 112 out of 469 plants) negative for any transgene (donor nucleic acid sequence expression cassettes or Cas9Nuc expression vector) have mutations at the safe harbor locus #1 (MIR604 insertion site) target sequence xMIR604FR2 (SEQ ID NO:28). 37 of these 112 events have biallelic mutations, i.e., both copies of the xMIR604FR2 sequence (SEQ ID NO:28) in the maize genome are mutated. The remaining 75 events have mutation in one of the copies of the sequence. This surprising result indicates that transient expression of Cas9 nuclease and sgRNA in the maize cells is sufficient for generating mutations at the chromosome targets. Also, selection is optional to obtain mutant plants. If sufficient number of regenerated plants is screened, targeted mutants can be easily identified through transient delivery and expression of Cas9Nuc protein and gRNA or gRNAs in plant cells.

Table 2. Breakdown of different types of events in regenerated plants with gRNA-Cas9 mediated targeted mutagenesis at the safe harbor locus #1(MIR604 insertion site) target sequence xMIR604FR2 (SEQ ID NO:28)

Experiment MZET13430	Number	Percentage
Total immature embryo targets	3620	
Total regenerated plants	944	
Donor nucleic acid (PMI) positive plants (Transformants)	473	13.1%[1]
- Events with no target site modification	89	
- Events with target site modification	384	81.2%

➢ Events with monoallelic modification	20	
o Monoallelic modification with co-integration of Cas9 vector	13	
o Monoallelic modification without co-integration of Cas9 vector	7	
➢ Events with biallelic modification	364	
o Biallelic modification with co-integration of Cas9 vector	288	
o Biallelic modification without co-integration of Cas9 vector	76	
Donor nucleic acid (PMI) negative plants (Escapes)	471	13.0%[2]
- Events with no target site modification	357	75.8%
- Events with target site modification	114	24.2%
➢ Events with monoallelic modification	75	15.9%
o Monoallelic modification with co-integration of Cas9 vector (22169)	0	
o Monoallelic modification without co-integration of Cas9 vector (22169)	75	15.9%
➢ Events with biallelic modification	39	8.3%
o Biallelic modification with co-integration of Cas9 vector (22169)	2	
o Biallelic modification without co-integration of Cas9 vector (22169)	37	7.9%
Total number of events with mutations at the target site	498	52.7%

transformation frequency is 13.1%
[2]Escape frequency is 13.0%

Example 4.4.3. Generation of targeted insertion events at the MIR604 insertion site safe harbor with *Agrobacterium-mtdiattd* transformation

Targeted insertion of transgenes into the safe harbor locus can also be generated with DNA donor and expression vectors for Cas9 nuclease and sgRNA delivered via *Agrobacterium. Agrobacterium-mediated* transformation methods have been described

elsewhere (Ishida et al., *Nat. Biotechnol.* 14:745-750 (1996)). Briefly, binary vectors for delivering donor DNA and expression cassettes of Cas9 and sgRNA are constructed. Donor DNA may be introduced in the same binary vector as expression cassettes of Cas9 and sgRNA, or may be introduced into separate T-DNA in the same binary vector, or may be introduced into separate binary vectors which can be transformed into the same *Agrobacterium* strain or separate *Agrobacterium* strains and delivered together through co-transformation. To construct a binary vector for *Agrobacterium-mediated* delivery of Cas9 and sgRNA, a DNA fragment containing the Cas9 and sgRNA expression cassettes is inserted into binary vector backbone to form pB-Cas9-U3-xMTR604FR2.

Similarly, a binary donor vector is constructed by inserting a nucleic acid fragment containing homology arms (xJHAX-03 and xJHAX-04), an eCry3.1Ab expression cassette, a mCry3A expression cassette, and a PMI marker expression cassettes into a binary vector. Both binary vectors are introduced into *Agrobacterium* strain LBA4404 containing a helper plasmid through electroporation. *Agrobacterium* strains containing these binary vectors are mixed and then used to co-infect maize immature embryos. Infected embryos are co-cultivated with *Agrobacterium* cells for 2-4 days and then used to induce calli. Calli are selected with mannose-containing media and mannose-resistant calli are regenerated into plantlets using a method similar to Negrotto et al. *Plant Cell Rep.* 19:798-803 (2000). Samples are taken from rooted plantlets for qPCR Taqman assays to enrich for potential targeted insertion events as described in the subsequent Examples and then junction PCR analyses are carried out to identify targeted insertion events as shown in **Fig. 2** and **Fig. 3**. Identified putative targeted insertion events are further characterized in detail by Southern analysis and sequencing of PCR products.

Example 5. Targeted insertion of transgene sequences to the MIR604 insertion site safe harbor mediated by TALE nucleases (TALENs)

Example 5.1. Selection of TALEN recognition target against AX-MIR604 sequences

Target sequences were selected from the AX_MIR604 (SEQ ID NO:2) for TALEN design. **Table 3** lists the selected sequences, their names and identifier numbers.

Table 3. Selected TALEN target sequences based on NP2222 genomic sequences (SEQ ID NO:2)

TALEN target name	Sequence (5' to 3')	Length	Sequence identifier
MIR604A1FW1	TTGCT ACTCC ATGTG ACT	18	SEQ ID NO:40
MIR604A1RV1	TTGTC ATATT CTTTT T	16	SEQ ID NO:41
MIR604A2FW1; aka. mir604Fw1	TACAC GTACT AATCG TGCT	19	SEQ ID NO:42
MIR604A2RV1; aka. mir604Rv1	TCCTG TCTAC TACGT GCT	18	SEQ ID NO:43
MIR604A2RV2	TTGTT CCTGT CTACT ACGT	19	SEQ ID NO:44
MIR604A3FW1	TTGGT CTTTG ATGAG GTGAT	20	SEQ ID NO:45
MIR604A3RV1	TCGAC ATGTA CAAAG TAGGT	20	SEQ ID NO:46
MIR604A4FW1	TTCGG AAACA TCCTT TAAT	19	SEQ ID NO:47
MIR604A4RV1	TTATA ATAAA ACTAA TATT	19	SEQ ID NO:48
MIR604A5FW1	TAATA AATAA ATAAA TAAAT	20	SEQ ID NO:49
MIR604A5RV1	TTGGA TTGCT GGATA ATGT	19	SEQ ID NO:50
MIR604A6FW1	TCGTT GCCAA AGCTG CAT	18	SEQ ID NO:51
MIR604A6RV1	TCCTG TCCTG CACTG CACT	19	SEQ ID NO:52
MIR604A7FW1; aka. mir604Fw2	TGCAT CCGTG CAGTG CAGT	19	SEQ ID NO:53
MIR604A7RV1; aka. mir604Rv2	TCCTA AACAA AGGAG GT	17	SEQ ID NO:54
MIR604A8FW1	TAGGA CGCGA TGCTG CT	17	SEQ ID NO:55
MIR604A8RV1	TGCGC ACGCA AGTGT CGT	18	SEQ ID NO:56
MIR604A9FW1	TCCAT CTCCA TTCAC TGGT	19	SEQ ID NO:57
MIR604A9RV1	TTCTG CAGGC ATTTG GCAT	19	SEQ ID NO:58
MIR604A10FW1	TTTTC TTCTC TTCTC GAT	18	SEQ ID NO:59
MIR604A10RV1	TAACC AGGCT AGCTT CGTT	19	SEQ ID NO:60
MIR604A11FW1	TAAGC TACAA AAGAA CGC	18	SEQ ID NO:61
MIR604A11RV1	TGTTT CGCGG CCGGC CCT	18	SEQ ID NO:62
MIR604A12FW1	TTTCC GTCCT GGCCT GTC	18	SEQ ID NO:63
MIR604A12RV1	TCGTC CGACG ACGAT CGAT	19	SEQ ID NO:64
MIR604Rv2-LT	TCCTA AACAA AGGAG GTCC	19	SEQ ID NO:65

5

Example 5.2. Design of TALEN fusion nucleases against selected the MIR604 insertion site safe harbor sequences

DNA binding specificity of TALENs is designed against the target sequences in **Table 3**. As an example, here is the design of two pairs of heterodimeric TALENs to cleave target sequences MIR604AXA2 ***(aka.*** MIR604FR1, SEQ ID NO:66, 5'-TACAC GTACT AATCG TGCTT CACGC ACAGG CACAG CACGT AGTAG ACAGG **A**-3') and MIR604AXA7 (aka. MIR604FR2, SEQ ID NO: 67, 5'-TGCAT CCGTG CAGTG CAGTG CAGTG CAGGA CAGGA CCTCC TTTGT TTAGG A-*3* '). Individual TALEN monomers recognizing 2 targets, MIR604A2FW1 (*aka*. mir604Fwl, 5'-TACAC GTACT AATCG TGCT-3', SEQ ID NO:42) and MIR604A2RV1 ***(aka.*** mir604Rvl, 5'-TCCTG TCTAC TACGT GCT-3', SEQ ID NO:43) within the MIR604AXA2 sequence, were assembled individually. For TALEN against MIR604A2FW1 ***(aka.*** mir604Fwl, 5'-TACAC GTACT AATCG TGCT-3', SEQ ID NO:42), the specificity determining di-residues within the RVD (Repeat-Variable Di-residue) repeats are as the following,

RVD position	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
RVD residue	N/A	NI	HD	NI	HD	NN	NG	NI	HD	NG	NI	NI	NG	HD	NN	NG	NN	HD	NG
Target nucleotide	T	A	C	A	C	G	T	A	C	T	A	A	T	C	G	T	G	C	T

For TALEN against MIR604A2RV1 ***(aka.*** mir604Rvl, 5'-TCCTG TCTAC TACGT GCT-3', SEQ ID NO:43), the specificity determining di-residues within the DVR repeats are as the following,

RVD position	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
RVD residue	N/A	HD	HD	NG	NN	NG	HD	NG	NI	HD	NG	NI	HD	NN	NG	NN	HD	NG
Target nucleotide	T	C	C	T	G	T	C	T	A	C	T	A	C	G	T	G	C	T

We constructed three versions of each TALEN containing the RVDs recognizing mir604Fwl (SEQ ID NO:42), a first full-length version that keeps most of the TAL effector protein sequences such as the N-terminal T3SS and the NLSs after the RVD repeat region (cTNmir604Fwl-01, SEQ ID NO:68), a second shorter version that has removed the N-

terminal T3SS (cTNmir604Fwl-02, SEQ ID NO:69), and a third short version with deletions in the N-terminal T3SS and also NLSs after the RVD repeat region (cTNmir604Fwl-03, SEQ ID NO:70). Similarly, we constructed three versions of each TALEN containing the RVDs recognizing mir604Rvl (SEQ ID NO:43), a first full-length version that keeps most of the TAL effector protein sequences such as the N-terminal T3SS and the NLSs after the RVD repeat region (cTNmir604Rvl-01, SEQ ID NO:71), a second shorter version that has removed the N-terminal T3SS (cTNmir604Rvl-02, SEQ ID NO: 72), and a third short version with deletions in the N-terminal T3SS and also NLSs after the DVR repeat region (cTNmir604Rvl-03, SEQ ID NO:73). The amino acid sequences of these engineered nucleases are shown in SEQ ID NO:68 (cTNmir604Fwl-01), SEQ ID NO:69 (cTNmir604Fwl-02), SEQ ID NO:70 (cTNmir604Fwl-03), SEQ ID NO:71 (cTNmir604Rvl-01), SEQ ID NO:72 (cTNmir604Rvl-02) and SEQ ID NO:73 (cTNmir604Rvl-03).

Individual TALEN monomers recognizing another 2 target sequences, MIR604A7FW1 (aka. mir604Fw2, 5'-TGCAT CCGTG CAGTG CAGT-3', SEQ ID. NO:53) and MIR604A7RV1 (aka. mir604Rv2, 5'-TCCTA AACAA AGGAG GT-3', SEQ ID NO:54) within the MIR604AXA7 (aka. mir604FR2, SEQ ID. NO:67) sequence, were also assembled individually. For TALENs against MIR604A7FW1 (aka. mir604Fw2, 5'-TGCAT CCGTG CAGTG CAGT-3', SEQ ID NO:53), the specificity determining di-residues within the RVD repeats are as the following,

RVD position	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
RVD residue	N/A	NN	HD	NI	NG	HD	HD	NN	NG	NN	HD	NI	NN	NG	NN	HD	NI	NN	NG
Target nucleotide	T	G	C	A	T	C	C	G	T	G	C	A	G	T	G	C	A	G	T

For TALENs against MIR604A7RV1 (aka. mir604Rv2, 5'-TCCTA AACAA AGGAG GT-3', SEQ ID NO: 54), the specificity determining di-residues within the DVR repeats are as the following,

RVD position	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
RVD residue	N/A	HD	HD	NG	NI	NI	NI	HD	NI	NI	NI	NN	NN	NI	NN	NN	NG
Target nucleotide	T	C	C	T	A	A	A	C	A	A	A	G	G	A	G	G	T

We constructed three versions of each TALEN containing the RVDs recognizing mir604Fw2 (SEQ ID NO:53), a first full-length version (cTNmir604Fw2-01, SEQ ID N0.74) that keeps most of the TAL effector protein sequences such as the N-terminal T3SS and the NLSs after the RVD repeat region, a second shorter version (cTNmir604Fw2-02, SEQ ID NO. 75) that has removed the N-terminal T3SS, and a third short version (cTNmir604Fw2-03, SEQ ID N0.76) that has deletions in the N-terminal T3SS and also NLSs after the RVD repeat region. Similarly, we constructed three versions of each TALEN containing the RVDs recognizing MIR604Rv2 (SEQ ID NO:50), a first full-length version (cTNmir604Rv2-01, SEQ ID NO. 77) that keeps most of the TAL effector protein sequences such as the N-terminal T3SS and the NLSs after the RVD repeat region, a second shorter version (cTNmir604Rv2-02, SEQ ID N0.78) that has removed the N-terminal T3SS, and a third short version (cTNmir604Rv2-03, SEQ ID N0.79) with deletions in the N-terminal T3SS and also NLSs after the RVD repeat region.

For MIR604AXA7 (aka. mir604FR2, SEQ ID NO:67) sequence cleavage, another pair of TALENS were assembled that have slightly different amino acid sequences and recognition specificity: cTNmir604Fw2-05 (SEQ ID NO.80) containing the RVDs recognizing mir604Fw2 (SEQ ID NO:53) and cTNmir604Rv2-04 (SEQ ID NO.81) containing the RVDs recognizing MIR604Rv2-LT (SEQ ID NO:65, 5'-TCCTA AACAA AGGAG GTCC-3'), respectively. The amino acid sequences of these engineered nucleases are in SEQ ID N0.74 (cTNmir604Fw2-01), SEQ ID N0.75 (cTNmir604Fw2-02), SEQ ID N0.76 (cTNmir604Fw2-03), SEQ ID N0.77 (cTNmir604Rv2-01), SEQ ID N0.78 (cTNmir604Rv2-02), SEQ ID N0.79 (cTNmir604Rv2-03), SEQ ID NO.80 (cTNmir604Fw2-05) and SEQ ID NO.81 (cTNmir604Rv2-04).

Example 5.3. Assembly of TALEN against AX-MIR604 insertion locus sequences

Artificial TALE fusion nuclease protein sequences (SEQ ID NO: 68 to SEQ ID NO.81) were back-translated into DNA coding sequences using plant-preferred codons for maximizing expression in corn and other monocot plants. Some of examples are shown here. For example, SEQ ID NO:82 is the DNA coding sequence for cTNmir604Fwl-01 protein sequence (SEQ ID NO:68) and SEQ ID NO:84 is the DNA coding sequence for cTNmir604Rvl-01 protein sequence (SEQ ID NO:71). Artificial fusion nuclease DNA sequences were then assembled from library of fragments containing different RVD repeats, promoter and terminator to form TALEN expression cassettes directly after Type lis enzyme

digestion and ligation as described (Cermak et al., *Nucleic Acid Research* 39(12):e82 (201 1); Zhang et al., *Nature Biotech* 29: 149-154 (201 1)). For example, the assembled reporter construct MIRA2R1FLA-GUUS contains the assembled TALEN sequence TLNMIR604 A2RV 1 (SEQ ID NO:84) encoding cTNmir604Rvl-01 (SEQ ID NO:71) under the control of maize ubiquitin promoter (prZmUbil-10) and also has a nonfunctional GUS recombination assay substrate cassette containing a direct repeat of GUS fragment and an inverted repeat of the 18 bp TALEN recognition sequence MTR604A2RV1 *(aka.* mir604Rvl, 5'-TCCTG TCTAC TACGT GCT-3', SEQ ID NO:43). Similarly, expression constructs containing other assembled TALENs are assembled in similar fashion. In many cases, expression cassettes for a pair of TALENs, *e.g.,* cTNmir604Fwl-01 (SEQ ID NO:68) and cTNmir604Rvl-01 (SEQ ID NO:71) which recognize and cleave a target sequence MIR604AXA2 (aka. MIR604FR1, 5'-TACAC GTACT AATCG TGCT T CACGC ACAGG CAC AG CACGT AGTAG ACAGG A-3', SEQ ID NO:66, only the upper strand is shown), are placed in the same transformation vector in order to coordinate their simultaneous expression in the target tissue during transformation.

Example 5.4. Transient assay for TALEN activity against AX_MIR604 DNA sequences

Assembled construct MIRA2R1FLA-GUUS containing the assembled TALEN sequence (SEQ ID NO: 84) encoding for cTNmir604Rvl-01 (SEQ ID NO:71) under the control of maize ubiquitin promoter (prZmUbil-10) and the nonfunctional GUS recombination assay substrate cassette were bombarded into immature maize embryos. The direct repeat of GUS fragment also contains an inverted repeat of the cTNmir604Rvl-01 TALEN recognition sequence mir604Rvl (5'-TCCTG TCTAC TACGT GCT-3', SEQ ID. NO:43). Similarly, expression constructs containing DNA sequences encoding for cTNmir604Fwl-01, cTNmir604Fwl-02, cTNmir604Fwl-03, cTNmir604Rvl-02, cTNmir604Rvl-03, or the corresponding pairs of them were bombarded into maize embryos along with their target substrate(s). In many cases, expression cassettes for a pair of TALENs recognizing and cleaving a target sequence, e.g., cTNmir604Fwl-01and cTNmir604Rvl-01 for MIR604AXA2 (aka. mir604FRl, SEQ ID NO: 66), were placed in the same transformation vector in order to coordinate their simultaneous expression in the target tissue. 1 to 4 days after bombardment, transformed maize embryos were placed in X-Gluc solution overnight to detect GUS activity histochemically. GUS activity is only visible when the GUUS repeat undergoes intramolecular recombination. Co-expression of a pair of TALENs (cTNmir604Fwl-01 and cTNmir604Rvl-01) recognizing MIR604FR1 (SEQ ID NO:66)

target greatly increases the number of blue spots **(Fig. 4,** treatment FR1), suggesting the target sequence is cleaved by the pair of heterodimeric TALENs to increase the frequency of homologous recombination.

Example 5.5. Maize chromosomal locus containing the target recognition sites is cleaved at high frequency by artificial TALENs

To test cleavage of chromosomal target sequence mir604FR2 (SEQ ID NO: 67) by TALENs expressed in maize cells, two different pairs of TALENs were used. The first pair of TALENs were in a single expression vector (21321) comprising nucleic acid sequences encoding for the expression of cTNmir604Fw2-03 and cTNmir604Rv2-03, and the second pair of TALENs were in a single expression vector (21998) comprising nucleic acid sequences encoding for the expression of cTNmir604Fw2-05 and cTNmir604Rv2-04. The expression vectors (21321 and 21998) were each co-delivered by biolistic transformation into maize embryos along with the donor vector 21942 described in Example 4.4. 1. Transformed embryos were selected on mannose to recover stable transgenic plants. Stable transgenic plants were analyzed for the presence of mutations in the target region using qPCR Taqman assay and/or sequencing of PCR products. Results in **Table 4** show that for both pairs of TALENs for target site MIR604FR2 (5'-TGCAT CCGTG CAGTG CAGTG CAGTG CAGGA CAGGA CCTCC TTTGT TTAGG A- 3', SEQ ID NO:67) resulted in high percentage of mutation in stable transformants when TALEN expression vectors are delivered into plant cells with biolistic method. Both the full length and truncated version of TALENs can mediate targeted mutagenesis at the target loci efficiently.

Interestingly, we also detected mutations of MIR604 insertion site locus mir604FR2 target site (5'-TGCAT CCGTG CAGTG CAGTG CAGTG CAGGA CAGGA CCTCC TTTGT TTAGG A- 3', SEQ ID NO:67) in many regenerated mannose selection escape plants. For example, in co-transformation experiments with vector TALEN vector 21321 and donor 21942 **(Table 4),** 14 of the plants generated, namely MZET130501B017A, MZET130501B038A, MZET130501B027A, MZET130501B031A, MZET130501A012A, MZET130501B041A, MZET130501B096A, MZET130402A030A, MZET130501B044A, MZET130501B057A, MZET130501B084A, MZET130501B130A, MZET130501B045A, MZET130704C003A, contained a mutation in the mir604FR2 target sequence, but they did not harbor any detectable transgenes from either donor or TALEN expression vectors and thus were escapes of mannose selection. In these escape plants, about 5% of them have mutations in the mir604FR2 target site and some of them have both alleles of the mir604FR2

target sequences mutated. Therefore, it is a viable approach to recover plants with mutations in target sites by transiently delivering TALENs and then regenerating untransformed plants directly without selection. Mutant plants can be identified by screening population of regenerants with proper assays such as PCR.

To test cleavage of chromosomal target locus by TALENs expressed in maize cells delivered by *Agrobacterium*, 4 different binary vectors (21631, 21632, 21633 and 21634) containing expression cassettes of different pairs of TALENs were constructed. All four binary vectors comprise the donor nucleic acid sequence comprising expression cassettes for eCry3.1Ab, mCry3A, and PMI. 21631 and 21633 additionally comprise nucleic acid sequences encoding for the expression of cTNmir604Fwl-01 and cTNmir604Rvl-01; 21632 and 21634 additionally comprise nucleic acid sequences encoding for the expression of cTNmir604Fw2-01 and cTNmir604Rv2-01. 21631 and 21632 have the TALEN expression cassettes and the gene targeting donor in one T-DNA, while 21633 and 21634 have these in two separate T-DNAs. Expression of the pair of TALENs in 21631 and 21633 is expected to result in cleavage of the chromosomal target sequence MTR604AXA2 *(aka.* MTR604FR1, 5'-TACAC GTACT AATCG TGCTT CACGC ACAGG CACAG CACGT AGTAG ACAGG A-3', SEQ ID NO:66) in the maize genome. Similarly, expression of the pair of TALENs in 21632 and 21634 should result in cleavage of the chromosomal target sequence MIR604AXA7 ***(aka.*** MIR604FR2, 5'-TGCAT CCGTG CAGTG CAGTG CAGTG CAGGA CAGGA CCTCC TTTGT TTAGG A- 3', SEQ ID NO:67) in the maize genome. These vectors transformed into maize embryos by Agrobacterium-mediated transformation method. Stable transgenic plants were analyzed for the presence of mutations in the target region using Taqman assay and/or sequencing of PCR products. Results in **Table 4** show that for both pairs of TALENs for target site MIR604FR1 (SEQ ID NO:66) and MIR604FR2 (SEQ ID NO: 67) resulted in high percentage of mutation in stable transformants when delivered via *Agrobacterium-mediated* transformation **(Table 4).**

Table 4. High rate of mutagenesis of mir604FRl(SEQ ID NO:66) and mir604FR2 (SEQ ID NO:67) target sequences at the native chromosomal MIR604 insertion site locus in stable transformants derived from co-transformation of a TALEN expression vector and a donor vector containing PMI selectable marker gene

Target locus	Delivery method	Nuclease vector	Donor	No. of experi -ments	Total explants	No. of stable events	No. of events with target site mutation	Mutation frequency (% transformants)
MIR604 insertion site FR2 target	Biolistic	21321	21942	7	6279	132	46	34.8%
MIR604 insertion site FR2 target	Biolistic	21998	21942	2	7845	519	148	28.5%
MIR604 insertion site FR1 target	Agrobacterium	21631	21631	3	4521	492	134	27.2%
MIR604 insertion site FR1 target	Agrobacterium	21633	21633	3	5305	1024	218	21.3%
MIR604 insertion site FR2 target	Agrobacterium	21632	21632	3	4633	673	316	47.0%
MIR604 insertion site FR2 target	Agrobacterium	21634	21634	2	5764	990	247	24.9%

Example 5.6. Targeted insertion of transgenic sequences into NP2222 chromosomal locus corresponding to the MIR604 insertion site mediated by assembled TALENs

Cultured immature embryos Maize elite inbred line NP2222 were co-transformed with the targeting donor vector 21942 and TALEN expression vector 21321 or 21998 using particle bombardment **(Table 4** and **Table 5).** Targeting donor vector 21942 contains trait gene expression cassettes flanked by regions of homology (xJHAX-03 and xJHAX-04) flanking the TALEN cleavage site (SEQ ID NO: 67) at the MIR604 insertion site. **Table 5** shows the analysis results for potential targeted insertion at the MIR604FR2 cleavage site (SEQ ID NO:67). Four events showing PCR products as expected for double-stranded homologous recombination are obtained out of 519 PMI positive stable events **(Table 5).** Of these, a single event was identified as a clean event, meaning that it comprises a single copy of the donor nucleic acid sequence described in example 4.4.1, specifically the eCry3.1Ab,

mCry3A, and PMI expression cassettes, is backbone free, shows evidence of a double-crossover homologous recombination event, and has no integration of the vector DNA comprising the nuclease.

Table 5. Targeted insertion of mCry3A, eCry3.1Ab, and PMI expression cassettes into native safe harbor locus (native MIR604 insertion site locus) mediated by cleavage of FR2 sequence by TALEN

Delivery method	Nuclease vector	Donor	Number of Expts	Total explants	positive events	Targeted events	Intact Low Copy events
Biolistic	21998:	21942	2	7845	519	4	1

In the above experiments, transformation was done using particle bombardment of cultured immature embryos. However, immature embryos or calli derived from cultured embryos can also be used as targets. Transformation can also be done using an *Agrobacterium-mediated* gene delivery method as shown in **Table 4** using target tissues such as immature embryos, cultured embryos or calli derived from cultured embryos. For example, *Agrobacterium-mediated* transformation and recovery of events as result of targeted insertion mediated by TALEN to target site can be done using mannose selection in a fashion as described in the art (U. S. Patent No. 7,935,862, for example), where, for example, NP2222 immature embryos are used as transformation targets.

Example 6. Targeted insertion of transgenes at the safe harbor (MIR604 insertion site) mediated by engineered meganucleases

Example 6.1. Maize chromosomal target sequence selection for design of engineered meganucleases

Targeted insertion of transgenic sequences for replacing short stretch of DNA sequences (allele replacement) or inserting large DNA fragment (transgene insertion) can also be mediated by homologous recombination using DNA breaks introduced by engineered meganucleases (Puchta and Fauser, *Plant Journal* 78:727-741 (2014); Chen and Gao, *Plant Cell Rep.* 33:575-583 (2014)). The present example shows if breaks induced by engineered meganucleases can be used to mediate the insertion of large DNA molecules into the desired

chromosomal safe harbor target in corn plants. To compare its effectiveness against TALEN and CRISPR-Cas9, the safe harbor locus # 1 (MIR604 event insertion site) was chosen as the transgene insertion site. Therefore, although not to be limited by methodology, the present application teaches transgene insertion mediated by 3 nucleases platforms, namely TALEN, meganuclease and sgRNA-Cas9. Maize safe harbor locus # 1 (aka. MTR604 event insertion site) sequences (SEQ ID NO: 1 or SEQ ID NO: 2) were scanned for optimal targets for designing engineered meganucleases using technologies in the art, e.g., by using rational protein design methodology to design engineered meganucleases with altered cleavage specificity based on LAGLIDADG family meganuclease *l-Crel* (U.S. Patent No. 8,021,867). The rationally designed engineered I-Crel meganuclease variants that cleave the target sequence at high efficiency and with minimal off target cleavage are selected to mediate targeted insertion of transgenes at the safe harbor locus. DNA sequences encoding novel meganuclease variants are placed under the control of maize ubiquitin-1 promoter (prUbil-10) followed by the NOS terminator and the expression cassette is sub-cloned into a biolistic transformation vector backbone.

To test the *in planta* activity of engineered I-Crel meganuclease variant in cleaving maize chromosomal target sequence and its ability to mediate targeted insertion through homologous recombination, meganuclease expression vector is co-bombarded with targeting donor vector 21942 into immature maize embryos. Briefly, plasmid DNA vector carrying expression cassette for the engineered meganuclease is mixed with a fragment of vector 21942 which encodes the donor nucleic acid sequence and precipitated onto gold particles. The donor nucleic acid sequence of vector 21942 contains regions from xJHAX-03 to xJHAX-04, including PMI marker gene and two gene cassettes as described in Example 4.4.1. Immature embryos are isolated from harvested immature ears at about 9-1 1 days after pollination and pre-cultured for 1 to 3 days on osmoticum media. Pre-cultured embryos are then bombarded with gold particles with co-precipitated DNA vectors (21942 fragment and the meganuclease expression plasmid) using BioRad PDS-1000 Biolistic particle delivery system. Methods for maize immature embryo bombardment, callus induction tissue regeneration and rooting methods are known in the art (for example, Wright et al. 2001, *Plant Cell* Reports 20:429-436 (2001)). Bombarded embryos are then incubated in callus induction media and then moved onto mannose selection media. Mannose resistant calli are transferred to regeneration media to induce shoot formation. Shoots are then sub-cultured onto rooting media. Samples are then harvested from rooted plants for PCR and Taqman assays to identify potential plants containing the targeted insertion. Identified putative

targeted insertion events are further characterized by more detailed PCR, sequencing and Southern analysis for confirmation. In addition to the stably transformed events, we also assay for the presence of mutations in regenerated plants that escaped the mannose selection, i.e., transformation escapes that do not contain any transgene from the targeted insertion donor or the meganuclease vector. Escape plants that are negative for any transgene but have mutations at the safe harbor locus #1(MTR604 insertion site) target sequence are identified. Transient expression of the meganuclease in the maize cells is sufficient for generating mutations at the chromosome targets. Also, selection is optional to obtain mutant plants. If a sufficient number of regenerated plants is screened, targeted mutants can be easily identified through transient delivery and expression of meganuclease in plant cells.

Example 6.2. Generation of targeted insertion events at the MIR604 insertion site safe harbor locus mediated by engineered meganucleases

The two homology arms, namely xJHAX-03 (SEQ ID NO: 38) and xJHAX-04 (SEQ ID NO: 39), of donor vector 21942 have sequences identical to the safe harbor #1 (MIR604 insertion site SEQ ID NO: 1 and SEQ ID NO: 2) and are used to guide targeted insertion of donor vector sequences to the cleavage site of engineered meganuclease at the target locus using homologous recombination. PCR reactions are also performed in a subset of events that are likely to targeted insertion based on Taqman analysis. Events identified to have a targeted insertion at the target locus using PCR primer pairs spanning the recombination junctions are analyzed by detailed DNA sequencing and Southern blot analysis to confirm that targeted insertion has happened.

Example 6.3. Generation of targeted insertion events at the safe harbor locus #1 (MIR604 insertion site) with *Agrobacterium-mtdiattd* transformation mediated by engineered meganucleases

Targeted insertion of transgenes into the safe harbor locus can also be generated with DNA donor and expression vectors for meganuclease delivered via *Agrobacterium.* *Agrobacterium-mediated* transformation methods are well-known in the art (for example, Ishida et al., *Nat. Biotechnol.* 14:745-750 (1996)). Meganuclease expression cassette and donor DNA can be placed either into separate binary vectors or in the same binary vector and then co-transformed in plant cells. Donor DNA and meganuclease can be co-delivered by using separate binary vectors. Binary vector 22445 is constructed by inserting the donor nucleic acid sequence from vector 21942 (namely, the three expression cassettes operably

linked to xJHAX-03 (SEQ ID NO: 38) and xJHAX-04 (SEQ ID NO: 39), as described in Example 4.4.1), into a binary vector useful for *Agrobacterium-mediated* transformation. A binary vector is also constructed for co-delivery of both the donor nucleic acid sequence and the meganuclease expression cassette from a single binary vector, where the donor nucleic acid sequence and the meganuclease expression cassette are each operably linked to right and left border sequences, so that they comprise two separate T-DNA's in a single binary vector. These binary vectors are transformed into *Agrobacterium* strain LBA4404 (pVGW7) via electroporation and then used for transformation of maize immature embryos. For *Agrobacterium-mediated* transformation, the *Agrobacterium* strain comprising the binary vector comprising 2 T-DNA's is used to infect maize immature embryos. Alternatively, *Agrobacterium* strains containing two binary vectors are mixed and then used to co-infect maize immature embryos. Infected embryos are co-cultivated with *Agrobacterium* cells for 2-4 days and then used to induce calli. Calli are selected with mannose-containing media and mannose-resistant calli are regenerated into plantlets. Samples are taken from rooted plantlets for Taqman and PCR analysis for identifying targeted insertion events as described above for biolistic transformation. PCR reactions are also performed in a subset of events that are likely to have targeted insertion based on Taqman analysis. Events identified to have targeted insertion at the target locus using PCR primer pairs spanning the recombination junctions are analyzed by detailed DNA sequencing and Southern blot analysis to confirm that targeted insertion has occurred.

Example 7. **Molecular characterization of targeted insertion of transgenic sequences into genomic AX_MIR604 locus**

Targeted insertion events identified by PCR assays were further characterized by more detailed sequencing and Southern blot analysis for confirmation. For example, events positive for junction PCRs **(Fig. 2** and **Fig. 3)** as expected from homologous recombination occurring at one or both homologous arms were obtained from screening PMI positive stable events (as shown in **Table 1** and **Table 5).** Detailed overlapping PCR analyses were done using primers spanning targeted insertion junctions comprising the AXJVIIR604 (SEQ ID NO:2) flanking genomic regions (xJHAX-03 and xJHAX-04) and part of the transformation donor vector. Presence of positive PCR signal suggests that site-directed nucleases indeed mediate targeted insertion into the MIR604 safe harbor locus (SEQ ID NO:2) at the DNA cleavage site of MIR604FR2 (5'-TGCAT CCGTG CAGTG CAGTG CAGTG CAGGA CAGGA CCTCC TTTGT TTAGG A- 3', SEQ ID NO:67). Detailed Southern blot analysis

showed that indeed insertion of gene stacks happened at the MIR604 insertion site safe harbor target locus through double crossover homologous recombination as shown by the presence of expected size **(Fig. 5,** lanes 4, 5, 6 and 8). Lanes 4, 5, 6 and 8 have a -28 Kb band as expected for double recombination product of donor vector with -18 Kb chromosomal target fragment. Another event from the same experiment in lane 7 **(Fig. 5)** has a copy of insertion that is probably from a single crossover recombination and has additional rearrangements since the size of the recombinant band is much larger than expected size of -28 Kb.

Example 8. Gene expression and insect resistance of transgenic events obtained by targeted insertion technologies

Targeted insertion events (MZET130403A067A, MZET134406B450A, MZET134504B010A, MZET134505A104A, MZET13471 1A236A, MZET140508A344A, MZET140807A856A, MZET140913A741A, MZET140913A594A, MZET130403A067A, MZET13 1500A128A) are evaluated for transgene expression by qPCR and ELISA assays. As a control, random integration events derived from donor vector (21942 or 22445) are also assayed for trait gene expression. Expression level is also compared with a maize line (AX5707DW) with the introgressed MTR604 locus. Since the inserted transgene contains Western corn rootworm resistance genes mCry3Aa and eCry3.1Ab, transgenic events and their progeny are evaluated in respect to the performance of insect resistance by growing them in pots infected by corn rootworm.

Example 9. High throughput assay for identifying plants with targeted mutations at desirable sequences

Currently, targeted mutants are identified using one of the following methods. The first method is PCR amplification of the target region followed by restriction enzyme digestion and gel electrophoresis if the mutated sequence contains a restriction site (Lloyd et al. 2005, *Proc. Natl. Acad. Sci.* USA 102:2232-37 (2005); Zhang et al, *Proc. Natl. Acad. Sci.* USA 107:12028-33 (2010)). This method is simple, but requires the presence of a suitable restriction site and thus cannot be used for most targets. A second method is PCR amplification of the target region followed by Sanger sequencing or deep sequencing (Gross et al., *Hum. Genet.* 105:72-78 (1999);. Shukla et al., *Nature* 459:437-41 (2009); Townsend et al., *Nature* 459:442-45 (2009)). A sequencing approach is definitive and sensitive, but takes a longer time and throughput can be limited by capacity. A third approach is PCR

amplification of the target region followed by denaturation, annealing and capillary electrophoresis (Li-Sucholeik et al., *Electrophoresis* 20:1224-1232 (1999); Larsen et al., *Hum. Mutat.* **13**:318-327 (1999)) or denaturing high-performance liquid chromatography to detect base pair changes by heteroduplex analysis (McCallum et al., *Nature Biotechnology* 18:455 - 457). These methods are limited by throughput and the identified mutations need to be further verified by sequencing. A fourth method is PCR amplification of the target region followed by denaturation, heteroduplex formation/strand annealing, digestion with mismatch-specific nuclease (such as CEL1 and T7 endonuclease) and gel electrophoresis (Oleykowski et al., *Nucleic Acids Res.* **26**:597-4602 (1998); Colbert et al., *Plant Physiol.* 126:480-484 (2001); Lombardo et al., *Nat. Biotechnol.* 25: 1298-306 (2007)), for example using the commercially available Surveyer™ nuclease assay kit (Transgenomic, Gaithersburg, MD, USA; Qiu et al., *BioTechniques* 36:702-707 (2004)). However, the gel-based assays are not as sensitive as high-throughput DNA sequencing and can only detect mutation with frequency of 1% or more. Therefore, there is still a need for a simple and high throughput method for identifying induced mutations of target sequences. Additionally, all of the above approaches of identifying a potential mutant in a target site are based on the presence of a new signal in a qualitative fashion, either a new band in a gel or a new peak in a chromatogram that is different from the wild type reference sequence.

We developed an alternative approach of identifying potential mutations. The method measures the reduction of the wild type target site sequence in cells or tissues that have been treated with a site-directed nuclease in a quantitative fashion in comparison with a reference sample as shown in **Fig. 6**. In a DNA sample isolated from wild type (WT) tissues, there is no reduction of the target sequence DNA copy number. Typically, the copy number call in WT tissue is 2 copies for a single copy gene in a diploid organism. For example, ADH gene in WT maize has 2 copies. If one of the copies is mutated, only one copy of the wild type (WT) target site sequence remains. If both copies of the target sequences are mutated, the copy number of the M target sequence becomes zero **(Fig. 6).** Thus, by performing quantitative polymerase reaction assays to measure changes in the target sequence copy number, it is possible to detect if there is a mutation present in the DNA samples by comparing the result with that of a reference sample such as WT tissue. This quantitative approach significantly differs from previously known methods.

Target gene copy number can be assayed by several quantitative polymerase reaction (qPCR) techniques. Generally, qPCR is performed in such a way in that the amplified DNA is detected and measured quantitatively as the reaction progresses, or in "real time".

Therefore, qPCR is also referred to as real-time PCR. There are several potential approaches for the real-time detection of products in qPCR: (1) Measurement of PCR product with non-specific fluorescent dyes (such as SYBR® Green) that intercalate with any double-stranded DNA; this detection method is suitable when a single amplicon is being studied, as the dye will intercalate into any double-stranded DNA generated. (2) Measurement of PCR product based on target sequence-specific binding of oligonucleotide probes covalently labeled with a fluorescent reporter tag, such as in TaqMan® probes, Molecular Beacons™, or Scorpion primers. The oligonucleotide itself has no significant fluorescence, but it fluoresces either when annealed to the template (as in Molecular Beacons™) or when the dye is clipped from the oligonucleotide during extension (as in TaqMan® probes). The advantage of fluorescent probes is that they can be used in multiplex assays for detection of several target sequences in the same reaction. With TaqMan® probes, a target sequence-specific oligonucleotide probe is constructed with a fluorescent reporter at one end and a fluorescence quencher at the opposite end. The close proximity of the reporter to the quencher prevents detection of its fluorescence. The fluorescent oligonucleotide probe is broken down by the 5'- to 3'-exonuclease activity of the Taq polymerase so the fluorescent tag is no longer in proximity with the quencher and thus allows unquenched emission of fluorescence, which can be detected after excitation with a laser (Groves, *J Biomol. Tech.* 10:1 1-16 (1999)). An increase in the number of copies of PCR product at each PCR cycle results in a proportional increase in fluorescence due to the breakdown of the probe and release of the reporter.

As an example, we have designed a Taqman® probe-based method to specifically detect targeted mutation at the maize genomic MIR604 insertion site sequence that contains the cleavage site of CRISPR-Cas9 nuclease gRNA targeting SEQ ID NO:3 (5'-AGTGC AGTGC AGTGC AGGAC AGG-3') and the pair of TALENs (cTNmir604Fw2-01/cTNmir604Rv2-01) cleaving target sequence (SEQ ID NO:67, 5'-TGCAT CCGTG CAGTG CAGTG CAGTG CAGGA CAGGA CCTCC TTTGT TTAGG A-3'). As shown in **Fig.** 7, a real-time qPCR Taqman assay for detecting mutations within SEQ ID NO: 67 target sequence consists of two primers, a FW primer, 5'-CACAC CTCGT TGCCA AAGC-3' (SEQ ID NO:92) and a RV primer, 5'-CATCG CGTCC TAAAC AAAGG A-3' (SEQ ID NO:93), and a fluorescently labeled Taqman® probe (5'-CCTGT CCTGC ACTGC-3', SEQ ID NO:94) which hybridizes to the nuclease cleavage target site sequence (5'-GCAGT GCAGG ACAGG-3', SEQ ID NO:95, the target site M as shown in **Fig. 6).**

Example 10. Generation of plants with targeted mutations at desirable sequences without transgene insertion

Using the target specific assay as outlined above and in **Fig. 6** and **Fig.** 7 and qPCR Taqman assays for other target sequences, maize plants regenerated from immature embryos treated with engineered TALE nucleases or gRNA-Cas9 as described previously in Example 4 and Example 5 were assayed for copy number of different target sequences. **Table 6** shows the results.

Fluorescently labeled MGB Taqman® probe comprising of sequence 5'-CCTGT CCTGC ACTGC-3' (SEQ ID N0.94) for assay 4 (Mir604 JHAX Fw2/Rv2_MGB) is for detecting the copy number of intact nuclease cleavage site sequence (5'-GCAGT GCAGG ACAGG-3', SEQ ID NO:95) corresponding to the target sequence M in **Fig. 6**. A "low" copy number call has 1 copy. A "med" copy number call has 2 copies. A "high" copy number call has 3 or more copies. In WT maize plants and regenerated plants with no target site mutation, the copy number call with Assay 4 (the last column in Table 6, Mir604 JHAXFw2/Rv2_MGB) is "Med" (2 copies). In this set of 20 plants, 11 plants (55%) have no mutation at the genomic target sequence (SEQ ID NO:95, 5'-GCAGT GCAGG ACAGG-3'), but 6 plants (30%) have mutations in one copy of the target sequences (Low copy call), and 3 plants (15%) have both copies of the target sequences are mutated (copy call is 0). Since the qPCR assays can be multiplexed, several other assays for detecting transgene sequences are performed at the same time. In this set of plants, 7 of the 20 plants contain detectable transgene insertions (positive for Assays 1 to 3). Of the 9 plants with target sequence mutations, 5 (MZET130501B027A, MZET130501B031A, MZET130501B038A, MZET130501B044A and MZET130501B045A) of them do not contain any detectable transgene insertions, including 1 plant (MZET130501B027A) that has both copies of the target sequence mutated (biallelic or homozygous mutations). This experiment clearly demonstrated that targeted mutations at desirable sequences can be efficiently generated without transgene insertion by transiently expressing a site-directed nuclease. Additionally, the mutants can be efficiently identified using high throughput real-time qPCR assays containing at least one assay probe hybridizing to the nuclease cleavage site.

Table 6. Copy number determination of target sequence (SEQ ID NO:67)in regenerated maize plants from a Biolistic transformation experiment using qPCR Taqman assays

Plant ID	Construct ID	Assay 1[1]: cTNmir604Fw2-03	Assay 2[2]: cPMI-09	Assay 3[3]: mCry3A	Assay 4[4]: Mir604 JHAX Fw2/Rv2_MGB
MZET130501B026A	21321 21942	0	0	0	Med
MZET130501B027A	21321 21942	0	0	0	0
MZET130501B028A	21321 21942	0	High	High	Low
MZET130501B029A	21321 21942	0	0	0	Med
MZET130501B030A	21321 21942	0	0	0	Med
MZET130501B031A	21321 21942	0	0	0	Low
MZET130501B032A	21321 21942	0	Low	Low	0
MZET130501B033A	21321 21942	0	High	High	0
MZET130501B034A	21321 21942	0	Low	Low	Med
MZET130501B035A	21321 21942	0	0	0	Med
MZET130501B036A	21321 21942	0	0	0	Med
MZET130501B037A	21321 21942	0	0	0	Med
MZET130501B038A	21321 21942	0	0	0	Low
MZET130501B039A	21321 21942	0	Low	Low	Med
MZET130501B040A	21321 21942	Low	High	0	Med
MZET130501B041A	21321 21942	0	0	0	Med
MZET130501B042A	21321 21942	0	High	High	Low
MZET130501B043A	21321 21942	0	0	0	Med
MZET130501B044A	21321 21942	0	0	0	Low
MZET130501B045A	21321 21942	0	0	0	Low

[1]Assay 1 (cTNmir604Fw2-03) is for detecting insertion of site-directed TALE nuclease expression vector (21321)

[2]Assay 2 for detecting inserted selectable marker gene cPMI-09 present in the donor vector (21942)

[3]Assay 3 for detecting inserted insect control gene mCry3A present in the donor vector (21942)

[4]Assay 4 (Mir604 JHAX Fw2/Rv2_MGB)is for detecting the copy number of intact target sequence (5'-GCAGT GCAGG ACAGG-3', SEQ ID NO:95) that is hybridizing to Taqman probe comprising of sequences 5'-CCTGT CCTGC ACTGC-3', (SEQ ID NO:94)

Example 11. High throughput assays and strategies for enriching plants with potential targeted insertion at desirable genomic loci

For identifying potential transgenic events containing targeted insertion at the MIR604 insertion site safe harbor locus, we developed a high throughput approach of enriching for potential mutations. The method involves the use of one assay (Assay T in **Fig. 8A)** to identify a plant that has a reduction in the copy number of the target sequence (Target T). The fluorescent probe for assay target T is located away from the fluorescent probe of assay target M **(Fig. 8A)** which detects the copy number of the site-directed nuclease

cleavage site M (also in **Fig. 6)** by at least 5 nucleotides in the region of the target locus. It should be noted that assay T probe can sit within the same amplicon as assay M probe. However, it should be as far away from M as possible as long as it is still within the region replaced by targeted insertion of transgenic sequences (as shown in **Fig. 8A,** region containing gene of interest (GOI). Since targeted insertion usually replaces certain sequences at the target locus other than the nuclease cleavage site (M), whereas non-targeted events that are mostly likely modified at the nuclease cleavage site by NHEJ usually would have smaller target site deletions. If a plant has reduced copy number at the nuclease cleavage site (Target M), but not having a reduced copy number call (i.e., wild type) at target region further away (Target T), this plant is very likely to have only small deletion and no targeted insertion at the target locus (Event types a, b and c in **Fig. 8A** and **Fig. 8B)** and can be discarded irrespective of the Target M or Target G copy number call. Events can be further enriched by looking at the Assay G results. Any plants negative for GOI (Assay G), i.e., event types d and e in **Fig. 8A** without transgene can be further discarded. The rest of the plants, i.e., event types from d to i in **Figs. 8B** with positive GOI signal are chosen as candidate plants with potential targeted insertion at the target locus and these events are characterized further by PCR reactions specific for recombination junctions as shown in **Fig. 2.**

Example 12. Use of high throughput qPCR assays for enrichment of candidate transgenic events with targeted insertion at the genomic safe harbor locus MIR604 insertion site

Results of copy number call of different target sequences were obtained using target-specific Assay 1 **(Table** 7, corresponding to assay T in **Fig. 8),** Assay 2 for nuclease cleavage site **(Table** 7, corresponding to target M in **Fig. 8)** and other transgene sequences (Assays 3 to 7 in **Table** 7, corresponding to assay G in **Fig. 8)** from maize plants regenerated from immature embryos treated with engineered TALEN as described previously in Example 5.

Table 7 shows assay results of some representative maize plants obtained from targeting experiments with co-delivery of the TALE nuclease expression vector 21321 and donor vector 21942. In this experiment, Assay 1 which is corresponding to the assay T of **Fig. 8** has a Taqman probe sequence of 5'-CTCGT TGCCA AAGCT GCATC CGT-3' (SEQ ID NO:97) which is located 18 bases away from the nuclease cleavage site (SEQ ID NO:67 , *5'*-TGCAT CCGTG CAGTG CAGTG CAGTG CA/GGA CAGGA CCTCC TTTGT TTAGG *A-3* ' where "/" indicates potential cleavage position). All plants that have "Med" copy number call for target (Assay 1) can be discarded irrespective of other assay results since

there is no homologous recombination-mediated replacement of the target sequences (SEQ ID NO:67). In some events (MZET130501A012A and MZET130501B033A) Assay 1 has higher copy number call than Assay 2, it means that the deletion around the nuclease cleavage site is relatively small at the target region. By using results from other assays (Assay 3 to Assay 7), further enrichment can be obtained by discarding plants that do not have genes of interest (GOI). If high quality targeted insertion events are desired, any plants positive for nuclease expression vector (Assay 6), and/or vector backbone (Assay 7), and having more than one copy of the donor vector (Assay 3 to 5) can be discarded. By using this enrichment method, only a subset of the total transgenic plants from a targeted insertion experiment will need to be analyzed further by other assays such as junction PCR **(Fig. 2** and **Fig. 3)** and DNA blot analysis **(Fig. 5)** to identify truly targeted insertion events. For example, events MZET131500A1 18A and MZET131500A128A **(Fig. 5)** were identified by following the above enrichment process from a set of 334 plants in targeted insertion experiment MZET131500A.

Table 7. **Taqman assays of transgenic events and use of assay results to enrich for potential targeted insertion events from regenerated maize plants derived from a Biolistic transformation experiment using qPCR Taqman assays.**

	Assay 1	Assay 2	Assay 3	Assay 4	Assay 5	Assay 6	Assay 7	
Plant ID	**MIR604 Fw2/Rv2 insertion site**	**Mir604 JHAX Fw2/Rv2_MGB**	**prCMP-04**	**cPMI-09**	**cWrangr-01**	**cTNmir604Fw2-03**	**xprLacZ-01-01**	**Note**
MZET130402A039A	Med	Med	0	0	0	0	0	Discard#*
MZET130402A040A	0	0	High	Med	High	0	0	Keep&
MZET130402A055A	Med	Med	Low	Low	Low	Low	0	Discard#
MZET130402A056A	Med	Med	0	0	0	0	0	Discard#*
MZET130501A012A	Low	0	0	0	0	0	0	Discard*
MZET130501A013A	Med	Med	0	0	0	0	0	Discard#*
MZET130501B031A	Low	Low	0	0	0	0	0	Discard*
MZET130501B032A	0	0	Low	Low	Low	0	0	Keep&
MZET130501B033A	Low	0	High	High	High	0	0	Keep&
MZET130501B034A	Med	Med	Low	Low	Low	0	0	Discard#*
MZET130501B050A	Low	Low	Low	Low	Med	0	0	Keep&
MZET130501B061A	Med	Med	0	0	0	0	0	Discard#*
MZET130501B062A	0	0	Low	Low	Med	0	0	Keep&
MZET130501B063A	Med	Med	0	0	0	0	0	Discard*
MZET130501B064A	0	0	Low	Low	Low	0	0	Keep&

MZET130501B065A	Med	Med	Low	Low	Low	0	0	Discard#
MZET130501B066A	Med	Med	0	0	0	0	0	Discard#
MZET130501B135A	0	0	Low	Low	Med	Low	0	Keep&
MZET130501B136A	Med	Med	0	0	0	0	0	Discard*
MZET130704B006A	Med	Med	0	0	0	0	0	Discard*
MZET130704B007A	0	0	High	High	High	0	0	Keep&
MZET130704B008A	0	0	Low	Low	Low	0	0	Keep&
MZET130704B009A	Med	Med	0	Med	Med	0	0	Discard#*
MZET130704B030A	0	0	0	Low	Low	0	0	Discard*
MZET130704B031A	Med	Med	0	0	0	0	0	Discard#
MZET130704B032A	Med	Med	0	0	0	0	0	Discard#
MZET130704B033A	0	0	Low	Low	Low	0	0	Keep&
MZET130704B036A	Med	Med	0	Low	0	0	0	Discard#
Assay purpose	Target region status	Nuclease cleavage site	Donor vector GOI cassette 1	Donor vector GOI cassette 2	Donor vector GOI cassette 3	Nuclease expression vector	Vector backbone	
	# For no target change; * For no (intact) GOI insertion; & For further junction PCR assays to identify targeted insertion events							

Example 13 Targeted gene stacking and replacement of transgene sequences at the MIR604 transgene locus

Example 13.1 T-DNA insert sequences of maize commercial event MIR604

Maize event MIR604 contains a single copy insertion of pNOV2130 T-DNA in a maize genome. The T-DNA insertion and its flanking genomic sequences were cloned and shown **Fig. 9**. The PMI marker gene sequence (cPMI-01, Seq. ID No. 98) is present in the transgene T-DNA insert located next to the flanking maize genomic region MIR604LBFS1.

Example 13.2 Selection of TALEN target site sequences in MIR604 event transgene locus

In order to stack additional trait gene cassettes to the MIR604 transgene locus, we concentrated our effort on the unique regions of the transgene. The PMI gene (cPMI-01, SEQ ID No. 98) is a desirable target since it is a selectable marker gene and is no longer needed after transgenic plant generation is completed. A new selectable marker gene cassette can be used to replace the PMI cassette using MIR604_RBFS1 or the mCry3A gene expression

cassette and MIR604_LBFS1 as regions of homology. We have chosen 3 target sequences (Seq. ID No. 99 to 101) in the PMI gene to design and assemble TALENs for demonstrating feasibility of gene insertion into the MIR604 transgenic locus. PMI_Target_Sequence #1 contains the following sequences, 5'- TTAAC TCAGT GCAAA ACTAT GCCTG GGGCA GCAAA ACGGC GTTGA CTGAA - 3' (SEQ ID No.99), PMI_Target_Sequence *# 2* has the following sequences, 5'- TCTCC ATTCA GGTTC ATCCA AACAA ACACA ATTCT GAAAT CGGTT TTGCC AAA - 3', SEQ ID No. 100) and PMI_Target_Sequence # 3 contains the following sequences, 5'- TGCAC ATCCG GCGAT TGCTC ACTTT TTACA ACAGC CTGAT GCCGA ACGTT TAA -3' (SEQ ID No. 101).

Example 13.3 Design and assembly of TALEN fusion nuclease genes against the PMI gene sequences

TALENs were designed for targeted cleavage of PMI transgene at sequence targets #1 and #3 (SEQ ID No. 99 and 101). For example, a pair of TALENs for cleaving PMI target sequence #1 (SEQ ID No. 99) were designed for TsPMIFWl (5'- TTA ACT CAG TGC AAA ACT -3', SEQ ID No. 102) and TsPMIRVl (5'- TTC ACT CAA CGC CGT TTT -3', SEQ ID No. 103). TALEN molecule TLN_PMIFWla (SEQ ID No. 108) was designed to bind the TsPMIFWl sequence target (5'- TTA ACT CAG TGC AAA ACT -3', SEQ ID No. 102) and TALEN molecule TLN_PMIRVl a (5'- TTC AGT CAA CGC CGT TTT-3', SEQ ID No. 109) was designed to recognize TsPMIRVl sequence target (SEQ ID No. 103). Similarly, another pair of TALENs was designed against TsPMIFW3 (5'- TGC ACA TCC GGC GAT TGC T -3', SEQ ID No. 106) and TsPMIRV3 (5'- TTA AAC GTT CGG CAT CAG -3', SEQ ID No. 107) for cleavage of PMI Target Sequence #3 (SEQ ID No. 101). TALEN molecule TLN_PMIFW3 (SEQ ID No. 110) was designed to bind the TsPMIFW3 sequence (5'- TGC ACA TCC GGC GAT TGC T -3', SEQ ID No. 106) and TALEN molecule TLN_PMIRV3 (SEQ ID No. 111) was designed to bind the TsPMIRV3 sequence (5'- TTA AAC GTT CGG CAT CAG -3', SEQ ID No. 107). The protein coding sequences of designed TALEN proteins TLN_PMIFWla (SEQ ID No. 108), TLN_PMIRVla (SEQ ID No. 109), TLN_PMIFW3 (SEQ ID No. 110) and TLN_PMIRV3 (SEQ ID No. 111) were back-translated into DNA sequences. DNA molecules encoding these TALENs were assembled as described in previous examples. The TALEN gene DNA sequences cTNPMIFWla (SEQ ID No. 112), cTNPMIRVla (SEQ ID No. 113), cTNPMIFW3-02(SEQ ID No. 114) and cTNPMIRV3-02 (SEQ ID No.l 15) encode TLN_PMIFWla (SEQ ID

No. 108), TLN_PMIRVla (SEQ ID No. 109), TLN_PMIFW3 (SEQ ID No. 110) and TLN_PMIRV3 (SEQ ID No. 111), respectively.

Example 13.4 TALEN expression vector and targeting donor vector construction

DNA sequences, cTNPMIFW3-02 (SEQ ID No. 114) and cTNPMIRV3-02 (SEQ ID No. 115) were introduced into expression cassettes, each driven by a constitutive promoter. The two TALEN gene expression cassettes were then introduced into a binary vector backbone to form binary vector 22840. Donor vector 22842 comprises the donor nucleic acid sequence, which comprises an insecticidal gene expression cassette and a glyphosate tolerance gene cassette between two homology sequences (xMIR604-01 and xMIR604-02). The glyphosate tolerance gene cassette comprises the gene *ZmEPSPS*, whose presence can be used to identify a successful insertion of the donor nucleic acid sequence. The two homology sequences (xMTR604-01 and xMIR604-02) are identical to sequences flanking the TALEN target sequence, i.e. PMI_Target_Sequence # 3 (SEQ ID No. 101). Targeted insertion of donor sequences from vector 22872 via homologous recombination into the MTR604 transgenic locus mediated by TALEN cleavage is illustrated in **Fig. 10.**

Example 13.5 Stacking of additional trait genes into a transgenic locus of a commercial event (MIR604) and inactivation of an unneeded transgene

Maize MTR604 event is widely cultivated for controlling Western corn rootworm (WCR) (Que et al., 2010, *GM Crops.* 1, 220-229). MIR604 transgene contains a PMI selectable marker gene for the generation of the transgenic event **(Fig. 9).** PMI gene doesn't offer any agronomic benefit and is no longer needed after event generation. However, it can be used as a landing pad for insertion of other trait gene cassettes into the MIR604 locus. To demonstrate such utility, MIR604 transgene locus was introgressed into an elite corn transformation line (NP2222) to form a new transgene receptor line NP2222DW. Line NP2222DW was used as transformation host for generation of targeted insertion events through site-directed nuclease mediated insertion into the PMI gene via homologous recombination. Immature embryos derived from selfed or sib-crossed NP2222DW plants were co-infected with recA-minus *Agrobacterium* strain LBA4404 (carrying helper plasmid pVGW7) containing binary vector 22840 (comprising TALEN expression cassettes) or 22872 (comprising donor nucleic acid sequence, which comprises two expression cassettes). Generation of transgenic events from infected immature embryos was as described except

glyphosate was used as selection (Negrotto et al. (2000), *Plant Cell Rep.* 19, 798-803). Call-derived from infected immature embryos were selected on 2 mM of glyphosate. Plants were regenerated on media containing 0.2 mM glyphosate. Glyphosate selected plants were sampled determining for transgene copy number and target site cleavage with Taqman assays.

Plants with target sequence cleavage were further analyzed by PCR for targeted integration with primers spanning across recombination junctions **(Fig. 10).** For example, for amplification of recombination junction involving xMIR604-02, the primer pair (P1/P2), FE4796 (SEQ ID NO: 127)/FE4793 (SEQ ID NO: 128) was used and the reaction would produce a PCR product of 2.13Kb if recombination occurred. Another primer pair, FE35036 (SEQ ID NO: 129') /FE35037 (SEQ ID NO: 130) with a product of 2.5 kb was also used for identification of potential targeted recombinants involving homology region of xMIR604-02. For amplification of recombination junction involving xMIR604-01, a pair of primers (P3/P4), FE35034 (SEQ ID NO: 131) /FE35035 (SEQ ID NO: 132) was used and the PCR reaction is expected to produce a product of 2 Kb if there is homologous recombination. **Table 8** shows several experiments of targeted insertion that targeted events were recovered using glyphosate selection ("ZmEPSPS positive events"). These experiments demonstrated DNA sequences containing additional trait genes can be efficiently inserted into the existing commercial event MTR604 locus through homologous recombination mediated by TALEN. It should be pointed out that other site-directed nucleases including engineered meganuclease, zinc finger nuclease or CRISPR-Cas9 can be used to substitute for TALEN in the above mentioned vector 22840 for cleaving the PMI gene sequences to mediate targeted insertion. Similarly, other methods of gene delivery including biolistic particle bombardment, whisker-mediated transformation, electroporation and PEG-mediated protoplast transformation can be used to introduce the site-directed nuclease expression vector and donor DNA molecules.

Table 8. Targeted insertion of expression cassettes flanked by homologous sequences in donor vector 22872 into MIR604 transgenic locus mediated by TALEN expressed from vector 22840 delivered by Agrobacterium infection

Experiment	Target Sequence	Nuclease vector ID	Donor vector ID	Total explants	ZmEPSPS positive events	Events with cPMI-01 target site mutations*	No. of targeted events**
MZET144515	cPMI-01	22840	22872	1682	53	10	2
MZET151723	cPMI-01	22840	22872	2676	252	ND	9
MZET151818	cPMI-01	22840	22872	4500	307	ND	4
MZET152212	cPMI-01	22840	22872	3680	628	236	8
MZET152311	cPMI-01	22840	22872	4150	808	277	12

*Based on target sequence (cPMI-01) copy number call as determined by qPCR Taqman assay. **As identified by PCR reactions with primers spanning across recombination junctions **(Fig. 10)**

Example 13.6 Stacking of additional trait genes into MIR604 transgenic locus by replacing the PMI gene cassette or the whole transgene

The genomic region harboring MIR604 transgene is a preferred location for trait gene expression. In addition to inserting additional transgenes into the PMI gene, the whole MIR604 transgene locus can be used as a landing pad for insertion of other trait gene cassettes by replacing part of the transgene sequences or the whole T-DNA insert. Similar to targeted insertion into PMI gene above (Example 13.5), line NP2222DW was used as transformation host for generation of targeted insertion events through site-directed nuclease mediated insertion into the MIR604 locus via homologous recombination. For replacing only the PMI cassette, the mCry3A gene and the LBFS region were used as homology sequences in the donor vector **(Fig. 11).** The same TALEN expression vector (22840) can be delivered into the NP2222DW maize cells along with the donor containing an insecticidal (IC) expression cassette and a selectable marker (such as PMI, ZmEPSPS, or PAT) expression cassettes **(Fig. 11).** Furthermore, one or more site-directed nucleases can be used to introduce chromosomal breaks in the PMI cassette sequences. For example, two or more single-guide RNAs (sgRNAs) can be used in conjunction with the Cas9 protein to cleave PMI cassette sequence simultaneously to remove the whole PMI expression cassette **(Fig. 11).**

Immature embryos are placed on callus induction media and then calli are selected on bialaphos-containing media. Generation of transgenic events from infected immature embryos, is, for example, as described above for mannose or glyphosate, where bialaphos may also be used as selection agent. Selected plants are sampled for transgene copy number and target site cleavage with Taqman assays. Plants with target sequence cleavage are further analyzed by PCR for targeted integration with primers spanning across the recombination junctions **(Fig. 11).**

For replacing the whole MIR604 T-DNA insert, both RBFS and LBFS are inserted into the donor molecule to serve as homology sequences to mediate insertion of novel trait gene cassettes (for example, insecticidal (IC) gene expression cassettes 1, 2, and a selectable marker (PAT, for example) expression cassette as the third cassette via homologous recombination **(Fig. 12).** Immature embryos isolated from selfed or sib-crossed NP2222DW ears are co-infected with *recA*-minus *Agrobacterium* strain LBA4404 (carrying helper plasmid pVGW7) containing binary vector 22840 and the donor DNA vector comprising IC expression cassettes 1, 2, and the PAT expression cassette. Similarly, more than one site-directed nuclease can be used simultaneously to cleave more than one MTR604 transgene sequence. For example, two or more single-guide RNAs (sgRNAs) can be used in conjunction with the Cas9 protein to cleave T-DNA sequence within the MIR604 transgene (for example, LB- and RB-proximal sequences and/or PMI and mCry3A cassettes) simultaneously to remove at least one expression cassette of the MIR604 T-DNA insert **(Fig. 12).** Infected immature embryos are placed on callus induction media and then calli are selected on bialaphos-containing media. Generation of transgenic events from infected immature embryos, is, for example, as described above for mannose or glyphosate, where bialaphos may also be used as selection agent. Selected plants are sampled for transgene copy number and target site cleavage with Taqman assays. Plants with target sequence cleavage are further analyzed by PCR for targeted integration with primers spanning across the recombination junctions **(Fig. 12).** It should be obvious to those skilled in the art that other methods of gene delivery including biolistic particle bombardment, whisker-mediated transformation, electroporation and PEG-mediated protoplast transformation can be used to introduce site-directed nuclease expression vector and donor DNA molecules.

Example 14 Targeted gene stacking and replacement of transgenic loci containing a nonfunctional selectable marker gene

Example 14.1 Design and assembly of TALENs for making chromosomal breaks in transgenic loci containing a nonfunctional selectable marker gene

It is known in the art that transgene sequences can be inserted into transgenic maize and rice loci containing a truncated non-functional selectable marker gene PMI, by using *Agrobacterium-mediated* transformation and taking advantage of dsDNA breaks created by expression of native meganuclease *l-Ceul* (U.S. Patent No. 7,935,862, incorporated by reference herein). However, targeted insertion mediated by native meganucleases is limited by the fact that a previously engineered nuclease cleavage site has to be inserted first in the transgene locus. Here, we want to test if novel designer site-directed nucleases such as TALEN can be designed against randomly chosen sequences within the existing transgenic locus to mediate targeted insertion of additional transgene sequences, to overcome this limitation. To achieve this, two pairs of TALENs were designed against a randomly selected target sequence (5'- ATAGA GATCC TCTAG AGTCG ACCAT GGTGA TCACT GCAGG CATGC AAGCT TGT -3', SEQ ID. No. 116, only the upper strand is shown) within the transgene locus of pNOV5025 transgenic events. Two sequences within this stretch of DNA were chosen as TALEN binding sites, 5'- ATAGA GATCC TCTAG AGT -3' ***(oka.*** rPMIFwl, SEQ ID No. 117, only the upper strand is shown) and 5'- ACAAG CTTGC ATGCC TGC -3' ***(oka.*** rPMIRvl, SEQ ID No. 118, only the lower strand is shown). One pair of TALENs consists of one full-length TALEN (cTNrPMIFwl-01, SEQ ID No. 119) designed against target sequence rPMIFwl 5'- ATAGA GATCC TCTAG AGT -3' (SEQ. ID. No. 117) and another full-length TALEN (cTNrPMIRvl-01, SEQ ID No. 120) designed against target sequence rPMIRvl 5'- ACAAG CTTGC ATGCC TGC -3' (SED ID No.l 18). The second pair of TALENs consists of one truncated TALEN (cTNrPMIFwl-02, SEQ. ID. No. 121) designed against target sequence rPMIFwl, 5'- ATAGA GATCC TCTAG AGT -3' (SEQ. ID. No. 117) and another truncated TALEN (cTNrPMIRvl-02, SEQ. ID. No. 122) designed against target sequence rPMIRvl, 5'- ACAAG CTTGC ATGCC TGC -3' (SEQ ID No. 118).

Example 14.2 Expression and transformation vectors of TALENs for truncated PMI target locus sequences

Artificial fusion nuclease DNA sequences were then assembled from library of fragments containing different RVD repeats, promoter and terminator to form TALEN expression cassettes directly after Type lis enzyme digestion and ligation as described (Cermak et al, 201 1, Nucleic Acid Research 39(12):e82; Zhang et al., 201 1, Nature Biotech 29:149-154). Several expression vectors (21438, 21792 and 21793) for TALENs against

truncated PMI target sequences were made. Vector 21438 comprises expression cassettes for TALENs cTNrPMIFwl-01 and cTNrPMIRvl-01. Vector 21792 comprises expression cassettes for TALENS cTNrPMIRvl-01 and cTNrPMIFwl-01. Vector 21793 comprises expression cassettes for TALENs cTNrPMIRvl-02 and cTNrPMIFwl-02. Initially, an existing targeting donor vector pNOV5045 (U.S. Patent No. 7,935,862) was used for testing targeted insertion. Later, additional targeting donor vectors 21779 and 22173 were also constructed and used for targeted insertion experiments (Table 9). Donor vectors pNOV5025, 21779, and 22713 contain the complementing 5'-region of the PMIintron cassette for restoring the PMI function and also other sequences of interest and regions of homology. Upon cleavage of the chromosomal target sequences by TALENs, donor vector sequence can be integrated into the target site via homologous recombination.

Example 14.3 Targeted insertion of transgenes into transgenic loci containing a nonfunctional truncated PMI gene mediated by TALEN

Selectable transgenic loci were generated from target vector pNOV5025 (described in U.S. Patent No. 7,935,862) using *Agrobacterium-mediated* transformation in maize line NP2222 as described using PPO as selectable marker. To test the effect of TALEN-mediated targeted insertion into these pNOV5025 loci, a donor vector (pNOV5045, 21779 or 22173) was co-delivered into immature maize embryo tissues along with a TALEN expression vector (21438, 21792 or 21793). After gene delivery and tissue recovery, transformed target tissues were placed on culture media containing mannose selection agent to recover events with targeted insertion, i.e. cells with reconstituted functional PMI gene as described (U.S. Patent No. 7,935,862). Targeted insertion events through homologous recombination should be resistant to mannose. To differentiate truly targeted events from selection escapes, tissues (callus or leaf) from putative mannose resistant events were first analyzed by PCR using primers spanning a targeted insertion junction. The presence of a positive PCR signal suggests TALEN-mediated targeted insertion into the pNOV5025 transgenic loci. Positive events are further analyzed by Southern blot analysis method to confirm that these events have truly targeted insertion as described (U.S. Patent No. 7,935,862). Table 9 shows the results of several targeted insertion experiments. The results demonstrate that useful trait genes can be inserted reproducibly into predetermined transgene loci by reconstituting a selectable marker gene at a useful frequency using different TALEN expression vectors and targeting donors. Both the full length and truncated version of TALENs can mediate targeted insertion at the transgenic loci.

Table 9. Targeted insertion experiments of pNOV5025 transgenic target loci with different donor vectors mediated by TALEN expression

Target locus	Nuclease vector	Donor	# Experi-ments	Total explants	Targeted events	Intact LC events
pNOV5025 transgenic loci with truncated PMI, F1 embryos	21438: FL TALEN	pNOV5045 : GUS + tPMI	8	6536	0	0
pNOV5025 transgenic lines with truncated PMI, F2 embryos	21438: FL TALEN	21779: tPMI	7	11521	4	4
pNOV5025 transgenic lines with truncated PMI, F2 embryos	21792: FL TALEN	21779: tPMI	3	8590	1	1
pNOV5025 transgenic lines with truncated PMI, F2 embryos	21793; dNC TALEN	21779: tPMI	5	10180	1	1

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 apparent that certain changes and modifications may be practiced with the scope of the present invention.

What is claimed is:

1. A method of integrating a transgene into a genomic nuclease cleavage site in a maize genome, comprising introducing into a maize cell:

a) a first nucleic acid molecule comprising at least 100 contiguous nucleotides, wherein said at least 100 contiguous nucleotides have at least 90% identity with a target site in a nucleotide sequence selected from the group comprising of SEQ ID NO:l, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:28, SEQ ID NO:66, and SEQ ID NO:67, and further comprising a transgene; and

b) a second nucleic acid molecule comprising a nucleotide sequence encoding a nuclease for site-directed cleavage at a genomic nuclease cleavage site adjacent to a nucleotide sequence selected from the group comprising of SEQ ID NO: 1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:28, SEQ ID NO:66, and SEQ ID NO:67, that corresponds to the at least 100 contiguous nucleotides of (a),

under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave the nucleotide sequence at the genomic nuclease cleavage site, whereby the transgene is integrated at the genomic nuclease target cleavage site in the maize genome.

2. The method of claim 1, wherein the genomic nuclease cleavage site is located within a chromosome interval on chromosome 1 defined by and including base pair (bp) position 38,860,000 to base pair (bp) position 39,015,000 as defined by Maize B73 RefGen_V2.

3. The method of any of claims 1-2, wherein the first nucleic acid molecule and the second nucleic acid molecule are introduced into the maize cell by biolistic nucleic acid delivery.

4. The method of any of claims 1-2, wherein the first nucleic acid molecule and the second nucleic acid molecule are introduced into the maize cell via an *Agrobacterium.*

5. The method of any of claims 1-2, wherein the first nucleic acid molecule and the second nucleic acid molecule are present on a single nucleic acid construct.

6. The method of any of claims 1-2, wherein the first nucleic acid molecule and the second nucleic acid molecule are present on separate nucleic acid constructs.

7. The method of any of claims 1-6, wherein the first nucleic acid molecule and/or the second nucleic acid molecule are transiently expressed in the maize cell.

8. A method of producing a maize plant or plant part, or progeny thereof, comprising a transgene integrated into a genomic nuclease cleavage site in the maize genome, comprising regenerating a maize plant from the maize cell produced by the method of claim 1.

9. A maize plant or plant part, or progeny thereof, comprising a transgene integrated into a genomic nuclease cleavage site in the maize genome, produced by the method of claim 8.

10. A method of enriching for a cell comprising a transgene inserted into a nuclease cleavage site in a genome of the cell, comprising:

a) introducing into a plurality of cells:

i) a first nucleic acid molecule comprising at least 100 contiguous nucleotides, wherein the at least 100 contiguous nucleotides have at least 90% identity with a target site in the genome of the cell, and further comprising a transgene; and

ii) a second nucleic acid molecule encoding a nuclease for site-directed cleavage at a nuclease cleavage site in the genome of the cell adjacent to the nucleotide sequence in the genome of the cell that corresponds to the at least 100 contiguous nucleotides of (i),

under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave at the nuclease cleavage site in the genome of the cell and integrate the transgene into the nuclease cleavage site in the genome of the cell;

b) culturing the cells of (a) to produce at least one cell line or tissue;

c) extracting a genomic DNA sample from each of the cell lines or tissues of (b);

d) performing real-time quantitative polymerase chain reaction (qPCR) assays T and G on the samples of (c), wherein the assays T and G respectively comprise the following probes:

i) a first probe comprising a nucleotide sequence that is complementary to a nucleotide sequence of the target site, at least five base pairs away from the nuclease cleavage site for carrying out assay T, and

ii) a second probe comprising a nucleotide sequence that is complementary to a nucleotide sequence of the transgene for carrying out assay G;

e) obtaining a DNA copy number of the target site from the results of assay T and a DNA copy number of the transgene from the results of assay G; and

f) enriching for a cell line or tissue that has reduced copy number in assay T relative to a reference and a copy number greater than zero for assay G,

thereby enriching for the cell comprising the transgene inserted into the nuclease cleavage site in the genome of the cell.

11. The method of claim 10, further comprising the step of discarding a cell line or tissue that has no change in the DNA copy number of assay T in comparison with a reference.

12. The method of claim 10, further comprising the step of discarding a cell line or tissue that has a copy number of zero for assay G.

13. A method of identifying a cell comprising a transgene inserted into a nuclease cleavage site in a genome of the cell, comprising:

a) introducing into a plurality of cells:

i) a first nucleic acid molecule comprising at least 100 contiguous nucleotides having at least 90% identity with a target site in the genome of the cell, and further comprising a transgene; and

ii) a second nucleic acid molecule encoding a nuclease for site-directed cleavage at a nuclease cleavage site in the genome of the cell adjacent to the nucleotide sequence in the genome corresponding to the at least 100 contiguous nucleotides of (i),

under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave at the nuclease cleavage site in the genome of the cell and integrate the transgene into the nuclease cleavage site in the genome of the cell;

b) culturing the cells of (a) to produce at least one cell line or tissue;

c) extracting a genomic DNA sample from each of the cell lines or tissues of (b);

d) performing real-time quantitative polymerase chain reaction (qPCR) assays T and G on the samples of (c), wherein the assays T and G respectively comprise the following probes:

i) a first probe comprising a nucleotide sequence that is complementary to a nucleotide sequence of the target site, at least five base pairs away from the nuclease cleavage site for carrying out assay T, and

ii) a second probe comprising a nucleotide sequence that is complementary to a nucleotide sequence of the transgene for carrying out assay G;

e) obtaining a DNA copy number of the target site from the results of assay T and a DNA copy number of the transgene from the results of assay G; and

f) identifying a cell line or tissue that has reduced copy number in assay T relative to a reference and a copy number greater than zero for assay G,

thereby identifying a cell comprising the transgene inserted into the nuclease cleavage site in the genome of the cell.

14. The method of claim 13, further comprising the step of discarding a cell line or tissue that has no change in the DNA copy number of assay T in comparison with a reference.

15. The method of claim 13, further comprising the step of discarding a cell line or tissue that has a copy number of zero for assay G.

16. The method of any of claims 10-15, wherein the assays are performed in the same mixture.

17. The method of any of claims 10-16, wherein the cell line or tissue is derived from a plant or a plant part.

18. The method of claim 17, wherein the plant is a maize plant.

19. The method of claim 18, wherein the nuclease cleavage site is a maize MTR604 transgene insertion site within a nucleotide sequence, wherein said nucleotide sequence has at least 90% identity to the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO:2.

20. The method of any of claims 10-19, wherein the nuclease is an engineered nuclease with programmable cleavage target specificity.

21. The method of claim 20, wherein the nuclease is a Cas9 nuclease comprising SEQ ID NO: 30.

22. A cell line or tissue that is enriched by the method of any of claims 10-12 or identified by the method of any of claims 13-21, wherein the cell line or tissue is from a plant.

23. A plant or plant part, or progeny thereof, derived from the cell line or tissue of claim 22.

24. A method of enriching for a cell comprising a mutation introduced into a nuclease cleavage site in a genome of the cell and lacking integration of a heterologous nucleotide sequence encoding a nuclease for site-directed cleavage of a nucleotide sequence at the nuclease cleavage site into the genome of the cell, comprising:

a) introducing a nucleic acid molecule comprising a heterologous sequence encoding a nuclease for site-directed cleavage of the nucleotide sequence at the nuclease cleavage site in the genome of the cell into a plurality of cells under conditions wherein expression of the nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave the nucleotide sequence at the nuclease cleavage site in the genome of the cell, thereby introducing a mutation at the nuclease cleavage site in the genome of the cell without integration of the heterologous nucleotide sequence encoding the nuclease into the genome of the cell;

b) culturing the plurality of cells of (a) to produce at least one cell line or tissue;

c) extracting a genomic DNA sample from each of the cell lines or tissues of (b);

d) performing real-time quantitative polymerase chain reaction (qPCR) assays 1 and 2 on the samples of (c), wherein the assays respectively comprise the following probes:

i) a first probe comprising a nucleotide sequence that is complementary to the nucleotide sequence comprising the nuclease cleavage site to carry out assay 1, and

ii) a second probe comprising a nucleotide sequence that is complementary to the heterologous nucleotide sequence encoding the nuclease to carry out assay 2;

e) obtaining a DNA copy number of the nuclease cleavage site from the results of assay 1 and a DNA copy number of the heterologous nucleotide sequence encoding the nuclease from the results of assay 2; and

f) enriching for a cell line or tissue that has a reduced copy number for assay 1 relative to a reference and a copy number equal to zero for assay 2,

thereby enriching for the cell comprising the mutation introduced into the nuclease cleavage site in the genome of the cell and lacking integration of the heterologous nucleotide sequence encoding the nuclease into the genome of the cell.

25. The method of claim 24, further comprising discarding a cell line or tissue that has no change in the DNA copy number of assay 1 relative to a reference.

26. The method of claim 24 further comprising discarding a cell line or tissue that has a copy number greater than zero for assay 2.

27. A method of identifying a cell comprising a mutation introduced into a nuclease cleavage site in a genome of the cell and lacking integration of a heterologous nucleotide sequence encoding a nuclease for site-directed cleavage of a nucleotide sequence at the nuclease cleavage site into the genome of the cell, comprising:

a) introducing a nucleic acid molecule comprising a heterologous sequence encoding a nuclease for site-directed cleavage of the nucleotide sequence at the nuclease cleavage site in the genome of the cell into a plurality of cells under conditions wherein expression of the nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave the nucleotide sequence at the nuclease cleavage site in the genome of the cell, thereby introducing a mutation

at the nuclease cleavage site in the genome of the cell without integration of the heterologous nucleotide sequence encoding the nuclease into the genome of the cell;

b) culturing the plurality of cells of (a) to produce at least one cell line or tissue;

c) extracting a genomic DNA sample from each of the cell line or tissue of (b);

d) performing real-time quantitative polymerase chain reaction (qPCR) assays 1 and 2 on the sample of (c), wherein the assays respectively comprise the following probes:

i) a first probe comprising a nucleotide sequence that is complementary to the nucleotide sequence comprising the nuclease cleavage site to carry out assay 1, and

ii) a second probe comprising a nucleotide sequence that is complementary to the heterologous nucleotide sequence encoding the nuclease to carry out assay 2;

e) obtaining a DNA copy number of the nuclease cleavage site from the results of assay 1 and a DNA copy number of the heterologous nucleotide sequence encoding the nuclease from the results of assay 2; and

f) identifying a cell line or tissue that has a reduced copy number for assay 1 relative to a reference and a copy number equal to zero for assay 2,

thereby identifying the cell comprising the mutation introduced into the nuclease cleavage site in the genome of the cell and lacking integration of the heterologous nucleotide sequence encoding the nuclease into the genome of the cell.

28. The method of claim 27, further comprising discarding a cell line or tissue that has no change in the DNA copy number of assay 1 as compared with a reference.

29. The method of claim 27 further comprising discarding a cell line or tissue that has a copy number greater than zero for assay 2.

30. The method of any of claims 24-29, wherein the assays are performed in the same mixture.

31. The method of any of claims 24-29, wherein the cell line or tissue is derived from a plant or plant part.

32. The method of claim 31, wherein the plant is a maize plant.

33. The method of claim 32, wherein the nuclease cleavage site is a maize MTR604 transgene insertion site within a nucleotide sequence, wherein said nucleotide sequence has at least 90% identity to the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO: 2.

34. The method of any of claims 24-33, wherein the nuclease is an engineered nuclease with programmable cleavage target specificity.

35. The method of claim 34, wherein the nuclease is a Cas9 nuclease comprising SEQ ID NO: 30.

36. A cell line or tissue that is enriched by the method of any of claims 24-26 or identified by the method of any of claims 27-35, wherein the cell line or tissue is from a plant.

37. A plant or plant part, or progeny thereof, derived from the cell line or tissue of claim 36.

38. A kit of reagents and instructions for carrying out the qPCR assays of any of claims 10-21 or 24-35.

39. A method of producing a plant, plant part, or progeny thereof comprising a mutation introduced at a nuclease cleavage site in a genome of a plant cell and lacking integration of a heterologous nucleotide sequence encoding a nuclease for site-directed cleavage of a nucleotide sequence at the nuclease cleavage site in the genome of the plant cell, comprising:

a) introducing into the plant cell a nucleic acid molecule comprising a heterologous nucleotide sequence encoding a nuclease for site-directed cleavage of the nucleotide sequence at the nuclease cleavage site in the genome of the plant cell under conditions wherein expression of the nucleic acid molecule occurs transiently to produce the nuclease and the nuclease can cleave

the nucleotide sequence at the nuclease cleavage site in the genome of the plant cell, thereby introducing a mutation at the nuclease cleavage site in the genome of the plant cell without integration of the heterologous nucleotide sequence encoding the nuclease into the genome of the plant cell; and

b) regenerating a plant, plant part, or progeny thereof from the plant cell of (a).

40. The method of claim 39, wherein the plant is a cereal plant.

41. The method of claim 39, wherein the plant is maize.

42. The method of claim 39, wherein the nuclease for site-directed cleavage is a Cas9 nuclease.

43. The method of claim 42, wherein the Cas9 nuclease comprises SEQ ID NO: 30.

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

a)introducing into the plant cell a first nucleic acid comprising at least 100 contiguous nucleotides, wherein the at least 100 contiguous nucleotides have at least 90% identity with a target site in the genome of the cell, and further comprising a transgene; and

b) a second nucleic acid molecule encoding nuclease for site-directed cleavage at a nuclease cleavage site in the genome of the cell adjacent to the nucleotide sequence in the genome of the cell that corresponds to the at least 100 contiguous nucleotides of (a), wherein the nuclease is a modified Cas9 nuclease comprising SEQ ID NO: 30,

under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave at the nuclease cleavage site in the genome of the cell and modify the target site in the genome of the plant cell.

45. The method of claim 44, wherein the plant cell is a maize cell.

46. The method of claim 45, wherein the maize cell is a transgenic maize cell.

47. The method of claim 46, wherein the transgenic maize cell is an event MIR604 transgenic maize cell.

48. A method of producing a maize plant, plant part, or progeny thereof comprising a modification at a target site in the genome of the plant cell, comprising:

a)introducing into the plant cell a first nucleic acid comprising at least 100 contiguous nucleotides, wherein the at least 100 contiguous nucleotides have at least 90% identity with a target site in the genome of the cell, and further comprising a transgene;

b) a second nucleic acid molecule encoding nuclease for site-directed cleavage at a nuclease cleavage site in the genome of the cell adjacent to the nucleotide sequence in the genome of the cell that corresponds to the at least 100 contiguous nucleotides of (a), wherein the nuclease is a modified Cas9 nuclease comprising SEQ ID NO: 30, under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave at the nuclease cleavage site in the genome of the cell and modify the target site in the genome of the plant cell; and

c) regenerating a plant, plant part, or progeny thereof from the plant cell of (a).

49. A maize plant, plant part, or progeny thereof produced by the method of claim 48.

50. The method of claim 44, wherein the first nucleic acid comprises a transgene.

51. A method of integrating a transgene into a genomic nuclease cleavage site in an event MIR604 transgenic maize genome, comprising introducing into an event MIR604 maize cell:

a) a first nucleic acid molecule comprising at least 100 contiguous nucleotides, wherein said at least 100 contiguous nucleotides have at least 90% identity with a target site in a nucleotide sequence selected from the group comprising SEQ ID NO: 133, SEQ ID NO: 134, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 137, SEQ ID NO: 138, and SEQ ID NO: 139, and further comprising a transgene; and

b) a second nucleic acid molecule comprising a nucleotide sequence encoding a nuclease for site-directed cleavage at a genomic nuclease cleavage site adjacent to a nucleotide sequence with at least 90% identity to a nucleotide sequence selected from the group comprising SEQ ID NO: 133, SEQ ID NO: 134, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 137, SEQ ID NO: 138, and SEQ ID NO: 139, that corresponds to the at least 100 contiguous nucleotides of (a),

under conditions wherein expression of the second nucleic acid molecule can occur to produce the nuclease and the nuclease can cleave the nucleotide sequence at the genomic nuclease cleavage site, whereby the transgene is integrated at the genomic nuclease target cleavage site in the event MIR604 maize genome.

52. A method of producing a maize plant, plant part, or progeny thereof comprising a transgene integrated into a genomic nuclease cleavage site in an event MIR604 maize genome, comprising regenerating a maize plant from the maize cell produced by the method of claim 51.

53. A maize plant, plant part, or progeny thereof comprising a transgene integrated into a genomic nuclease cleavage site in an event MIR604 maize genome, produced by the method of claim 52.

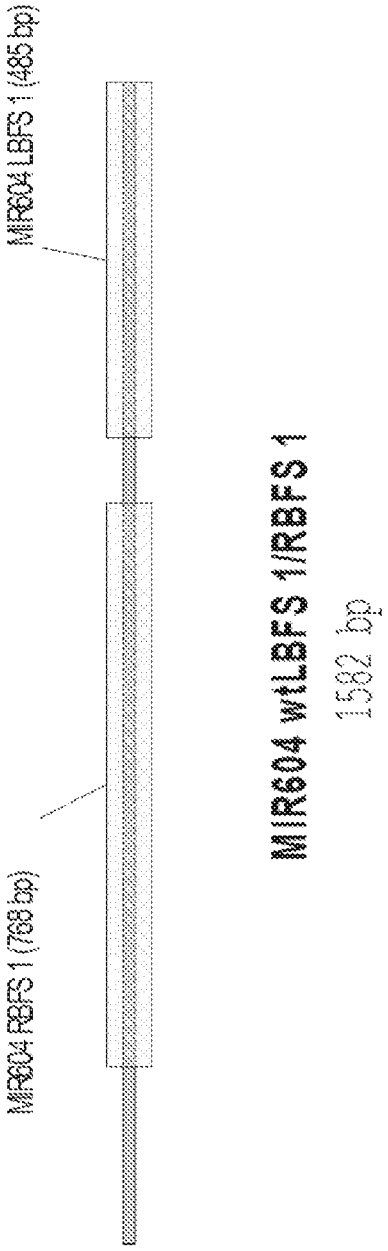

MIR604 LBFS 1 (485 bp)
MIR604 RBFS 1 (768 bp)
MIR604 wtLBFS 1/RBFS 1
1582 bp

FIG. 1

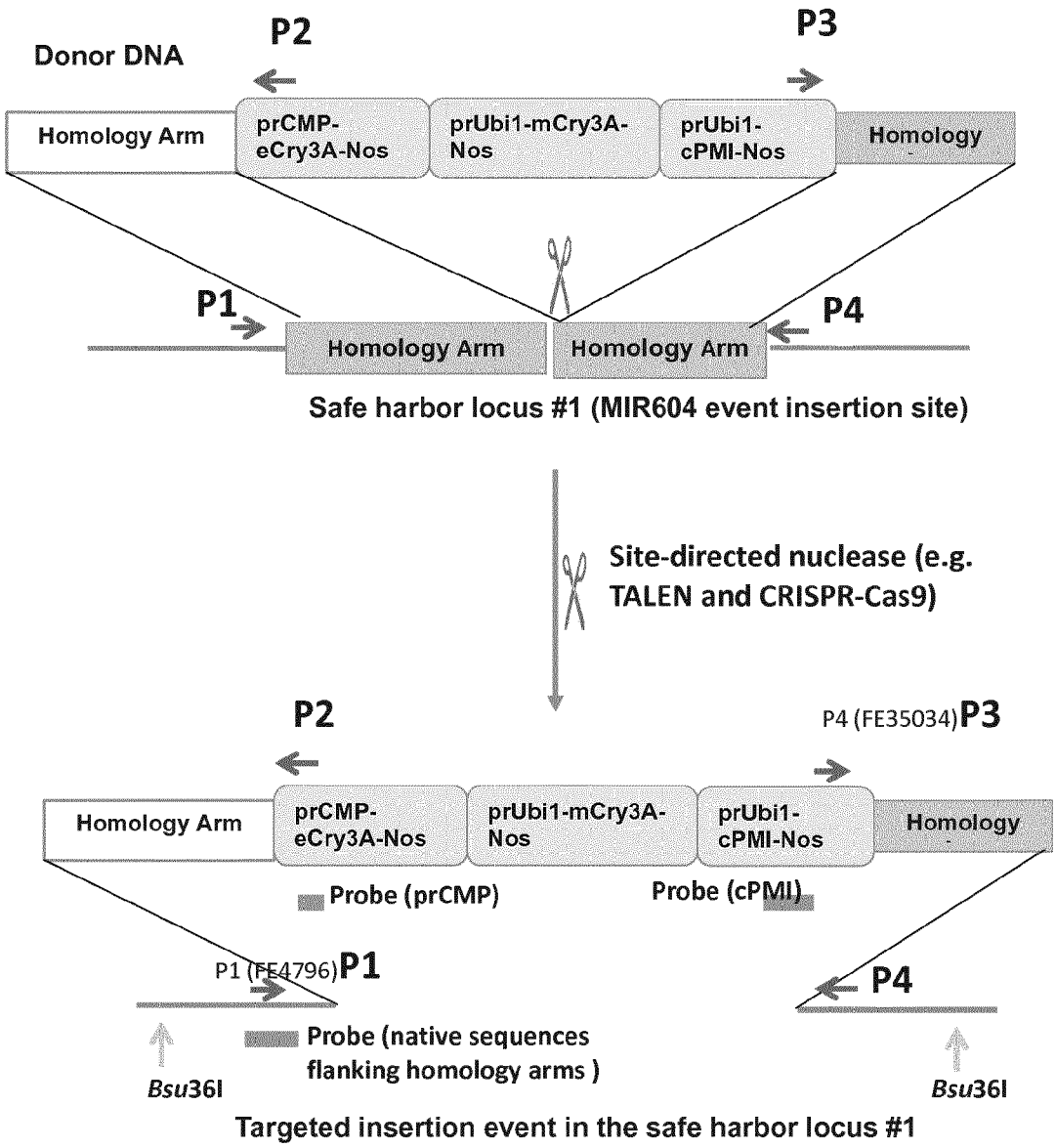

Donor DNA
P2
P3
Homology Arm
prCMP-eCry3A-Nos
prUbi1-mCry3A-Nos
prUbi1-cPMI-Nos
Homology
P1
P4
Homology Arm
Homology Arm
Safe harbor locus #1 (MIR604 event insertion site)
Site-directed nuclease (e.g. TALEN and CRISPR-Cas9)
P2
P4 (FE35034)P3
Homology Arm
prCMP-eCry3A-Nos
prUbi1-mCry3A-Nos
prUbi1-cPMI-Nos
Homology
Probe (prCMP)
Probe (cPMI)
P1 (FE4796)P1
P4
Probe (native sequences flanking homology arms)
Bsu36I
Bsu36I
Targeted insertion event in the safe harbor locus #1

FIG. 2

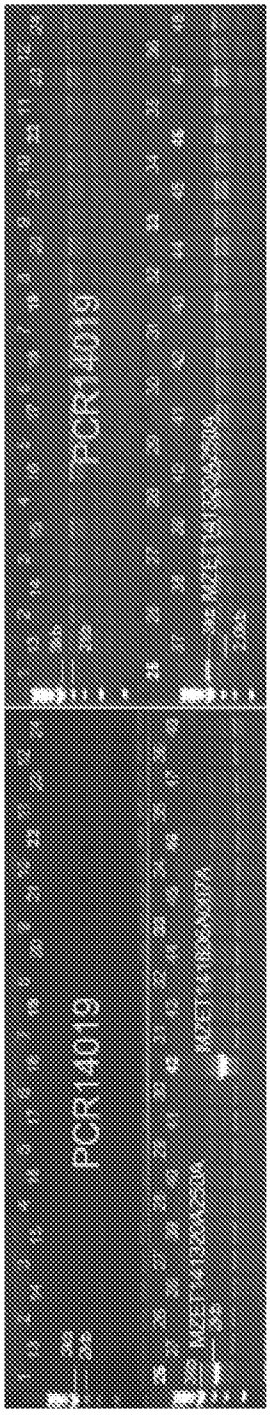


FIG. 3

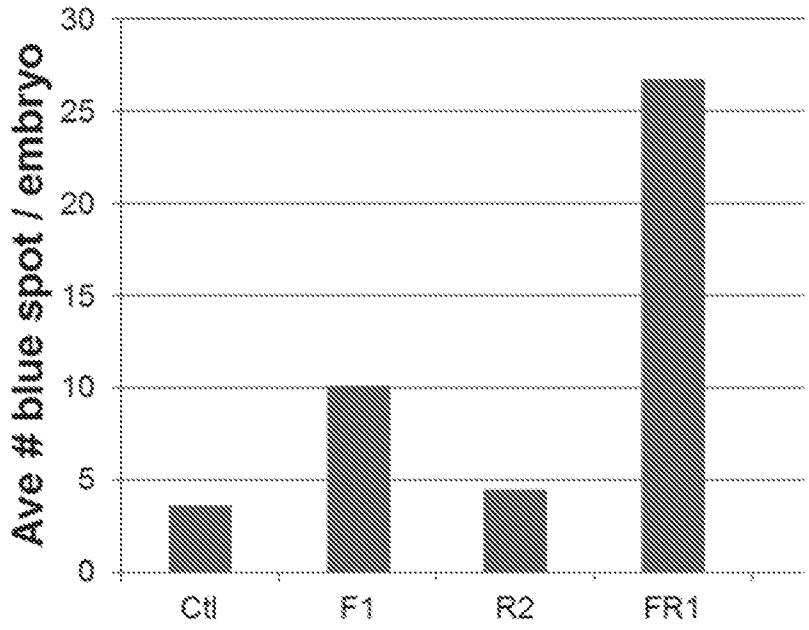


FIG. 4

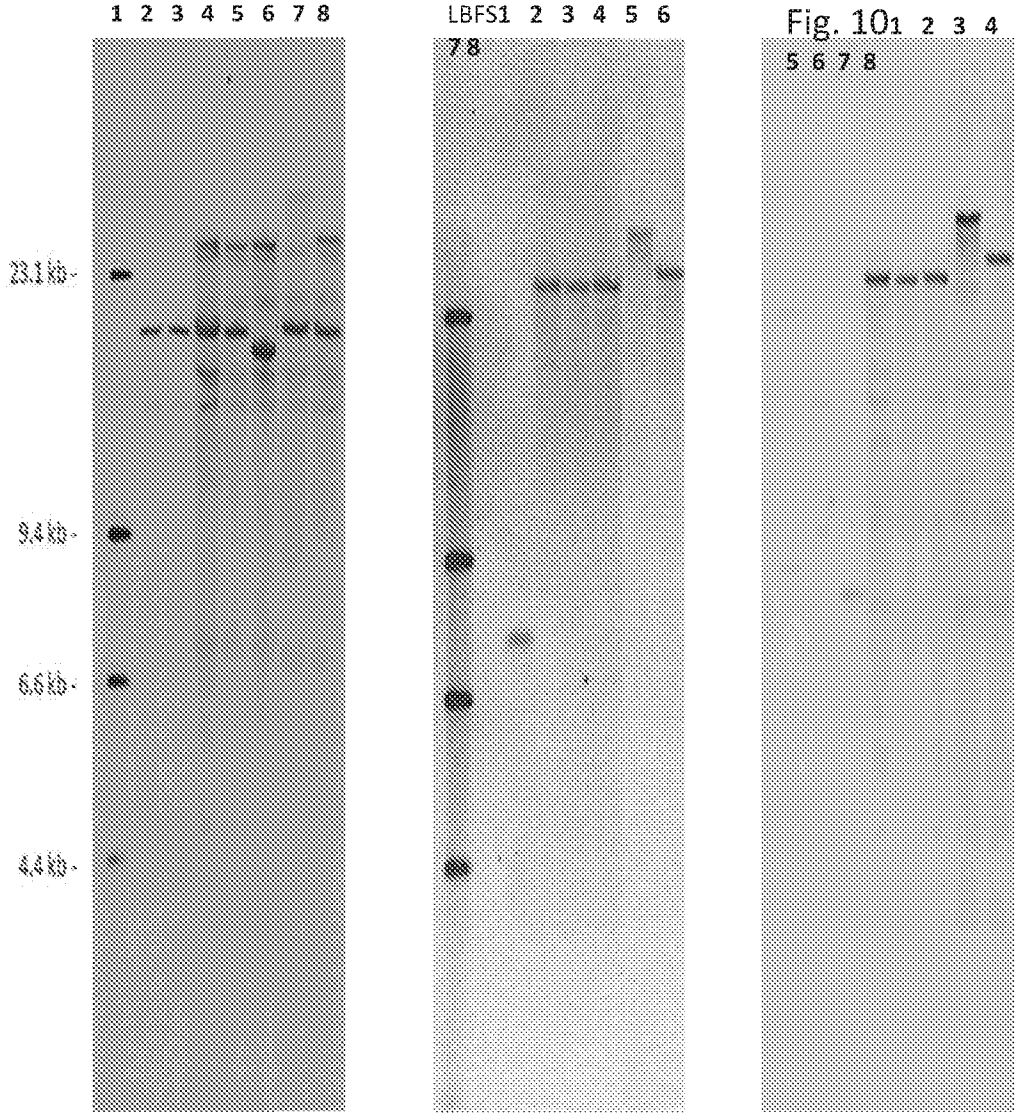
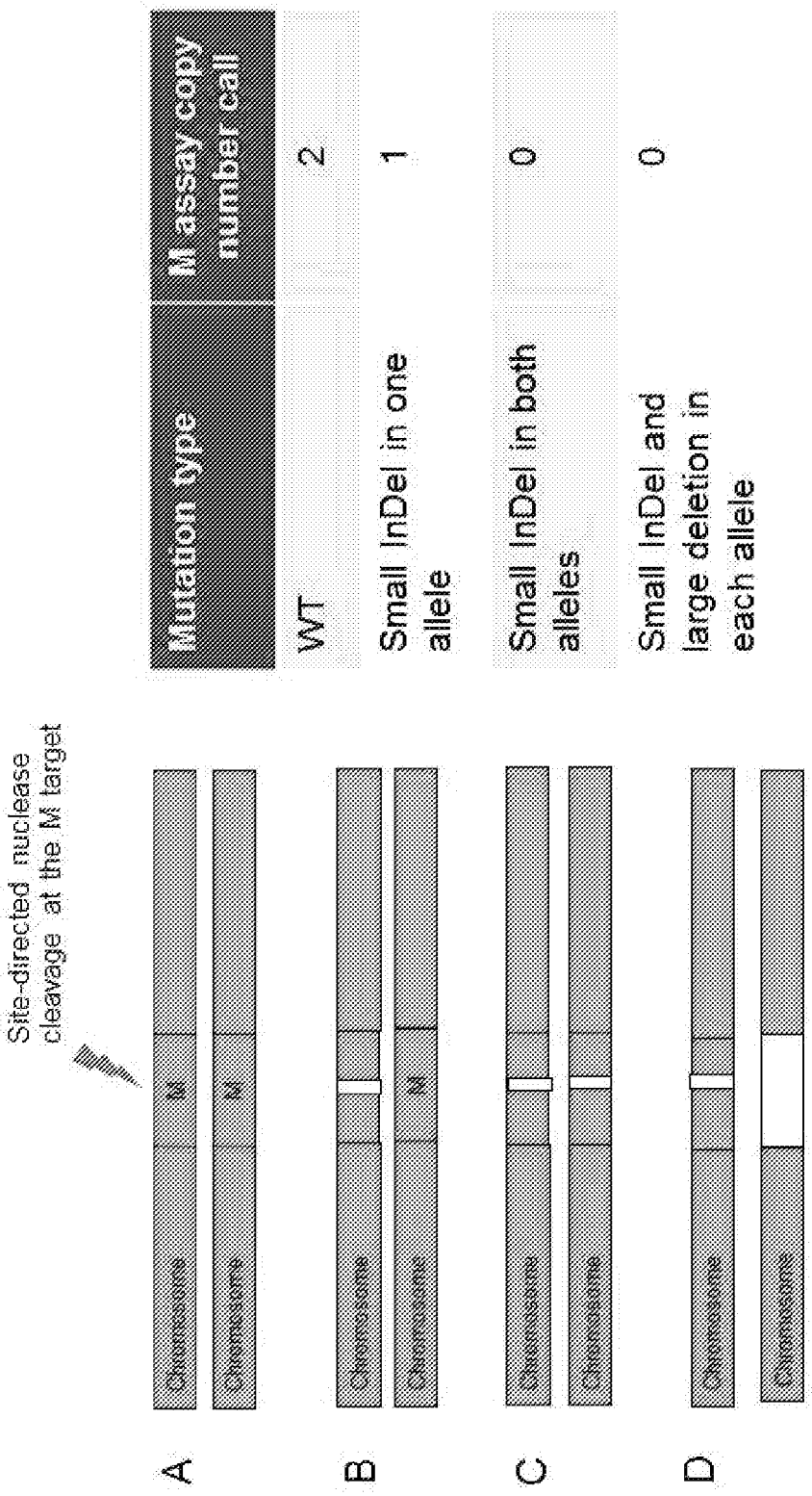

1 2 3 4 5 6 7 8
LBFS1 2 3 4 5 6 7 8
1 2 3 4 5 6 7 8
23.1kb
9.4kb
6.6kb
4.4kb
Probe 1
Probe 2
Probe 3

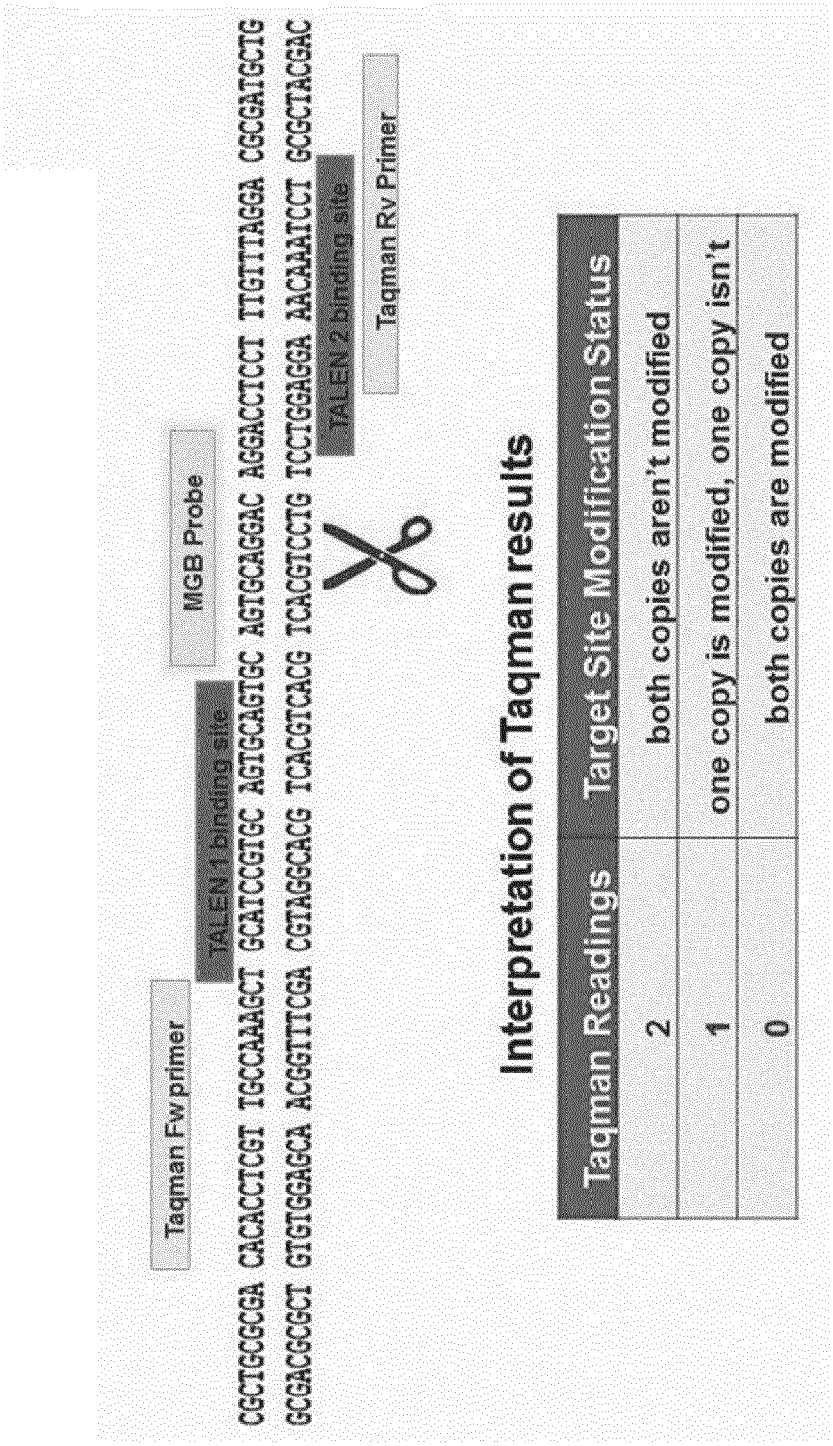
Fig. 10

FIG. 5



Mutation type	M assay copy number call
WT	2
Small InDel in one allele	1
Small InDel in both alleles	0
Small InDel and large deletion in each allele	0

FIG. 6



Taqman Readings	Target Site Modification Status
2	both copies aren't modified
1	one copy is modified, one copy isn't
0	both copies are modified

FIG. 7

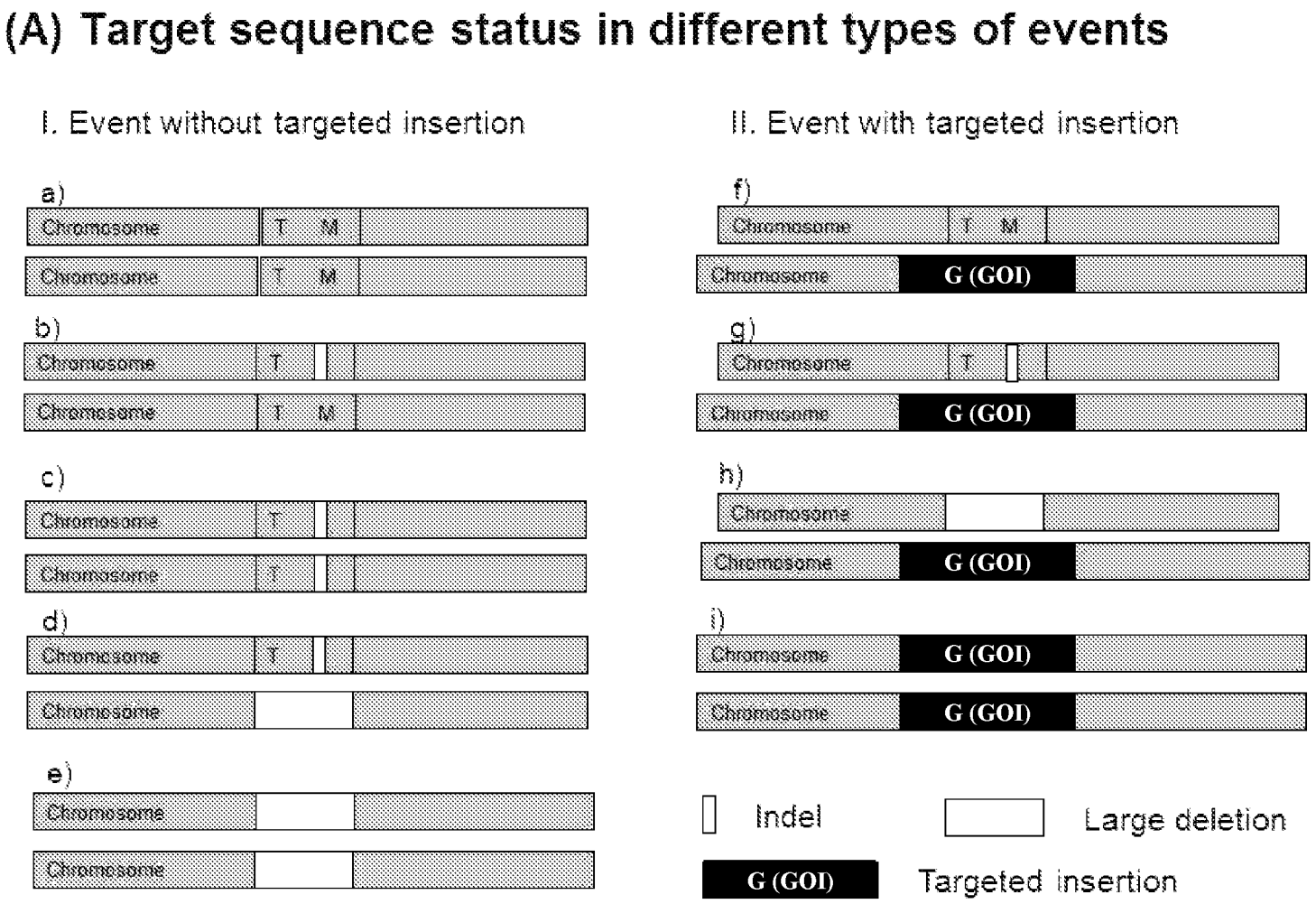

(A) Target sequence status in different types of events
I. Event without targeted insertion
II. Event with targeted insertion
a)
Chromosome
T M
Chromosome
T M
b)
Chromosome
T
Chromosome
T M
c)
Chromosome
T
Chromosome
T
d)
Chromosome
T
Chromosome
e)
Chromosome
Chromosome
f)
Chromosome
T M
Chromosome
G (GOI)
g)
Chromosome
T
Chromosome
G (GOI)
h)
Chromosome
Chromosome
G (GOI)
i)
Chromosome
G (GOI)
Chromosome
G (GOI)
Indel
Large deletion
G (GOI)
Targeted insertion

FIG. 8A

Event type	T assay copy number call	M assay copy number call	G assay copy number call
Event with no targeted insertion			
a) WT, no mutation in both alleles	2	2	≥ 0
b) Small indel in one allele	2	1	≥ 0
c) Small indel in both alleles	2	0	≥ 0
d) Small indel in one allele and large deletion in another allele	1	0	≥ 0
e) Large insertion in both alleles	0	0	≥ 0
Event with targeted insertion			
f) No mutation in one allele, targeted insertion in another allele	1	1	≥ 1
g) Small indel in each allele, targeted insertion in another allele	1	0	≥ 1
h) Large deletion in in each allele, targeted insertion in another allele	0	0	≥ 1
i) Targeted insertion in both alleles	0	0	≥ 2

FIG. 8B

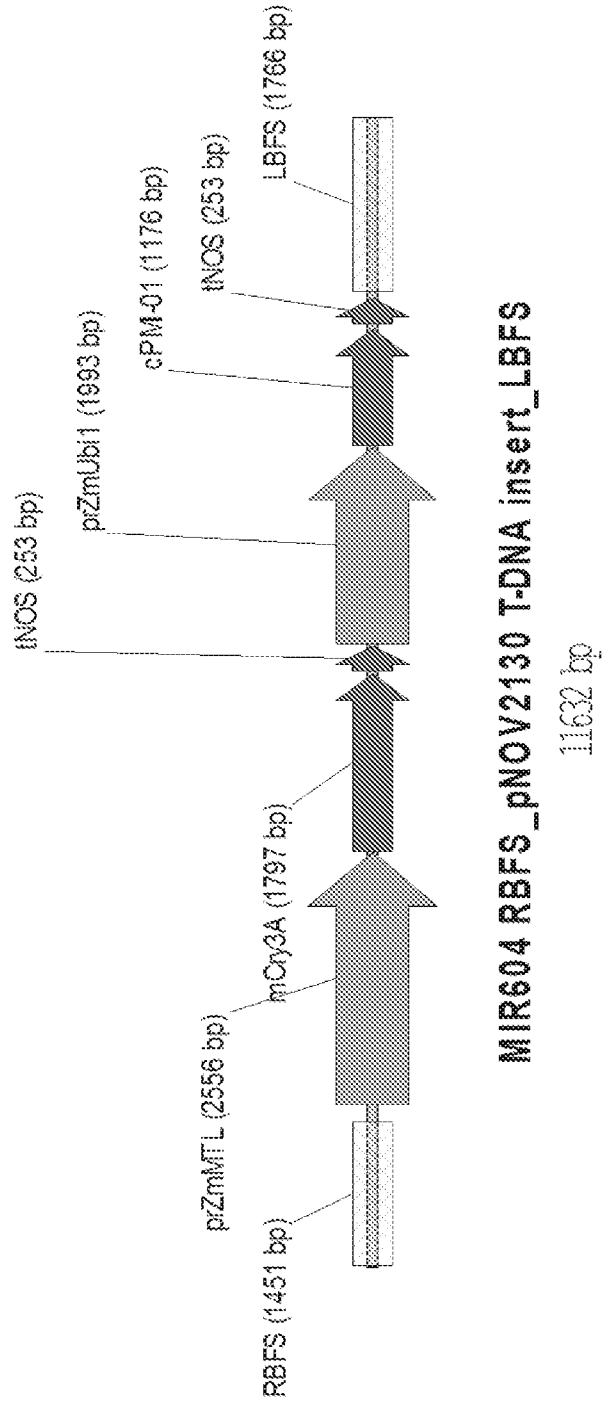

MIR604 RBFS_pNOV2130 T-DNA insert_LBFS
11632 bp
RBFS (1451 bp)
prZmMTL (2556 bp)
mCry3A (1797 bp)
tNOS (253 bp)
prZmUbi1 (1993 bp)
cPM-01 (1176 bp)
tNOS (253 bp)
LBFS (1766 bp)

FIG. 9

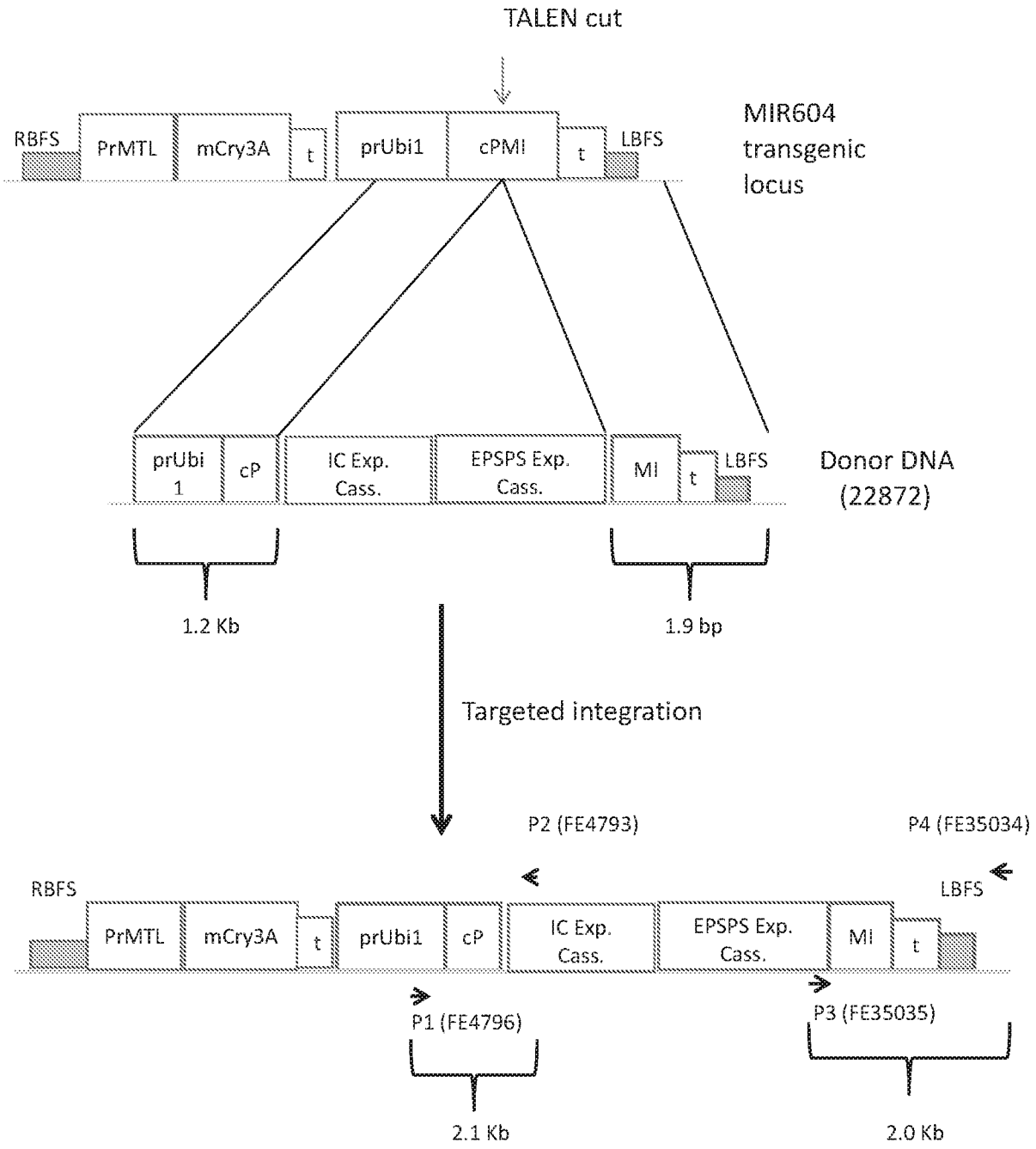

TALEN cut
RBFS
PrMTL
mCry3A
t
prUbi1
cPMI
t
LBFS
MIR604 transgenic locus
prUbi 1
cP
IC Exp. Cass.
EPSPS Exp. Cass.
MI
t
LBFS
Donor DNA (22872)
1.2 Kb
1.9 bp
Targeted integration
P2 (FE4793)
P4 (FE35034)
RBFS
PrMTL
mCry3A
t
prUbi1
cP
IC Exp. Cass.
EPSPS Exp. Cass.
MI
t
LBFS
P1 (FE4796)
P3 (FE35035)
2.1 Kb
2.0 Kb

FIG. 10

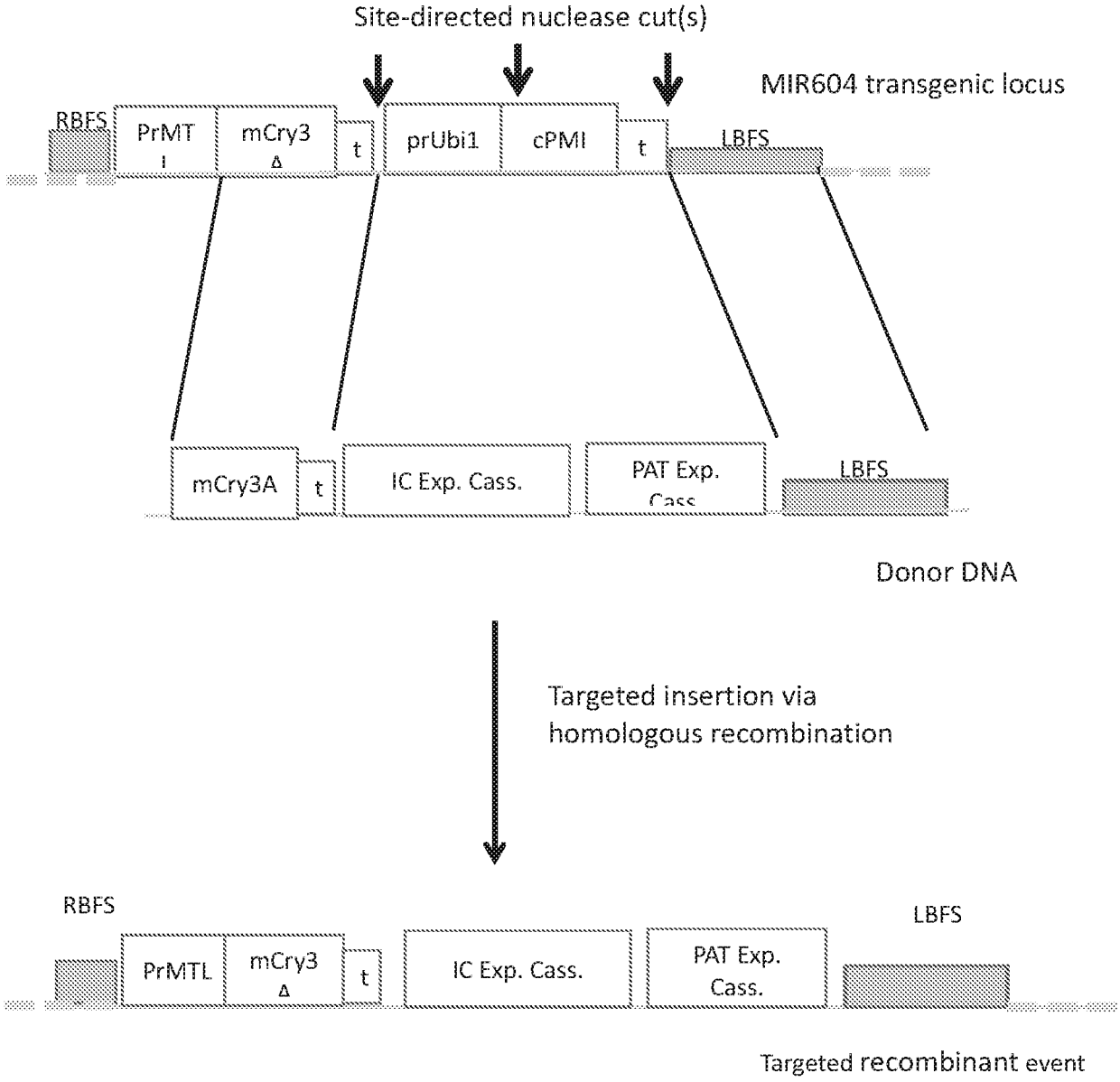

Site-directed nuclease cut(s)
MIR604 transgenic locus
RBFS
PrMT
I
mCry3
Δ
t
prUbi1
cPMI
t
LBFS
mCry3A
t
IC Exp. Cass.
PAT Exp.
Cass
LBFS
Donor DNA
Targeted insertion via
homologous recombination
RBFS
PrMTL
mCry3
Δ
t
IC Exp. Cass.
PAT Exp.
Cass.
LBFS
Targeted recombinant event

FIG. 11

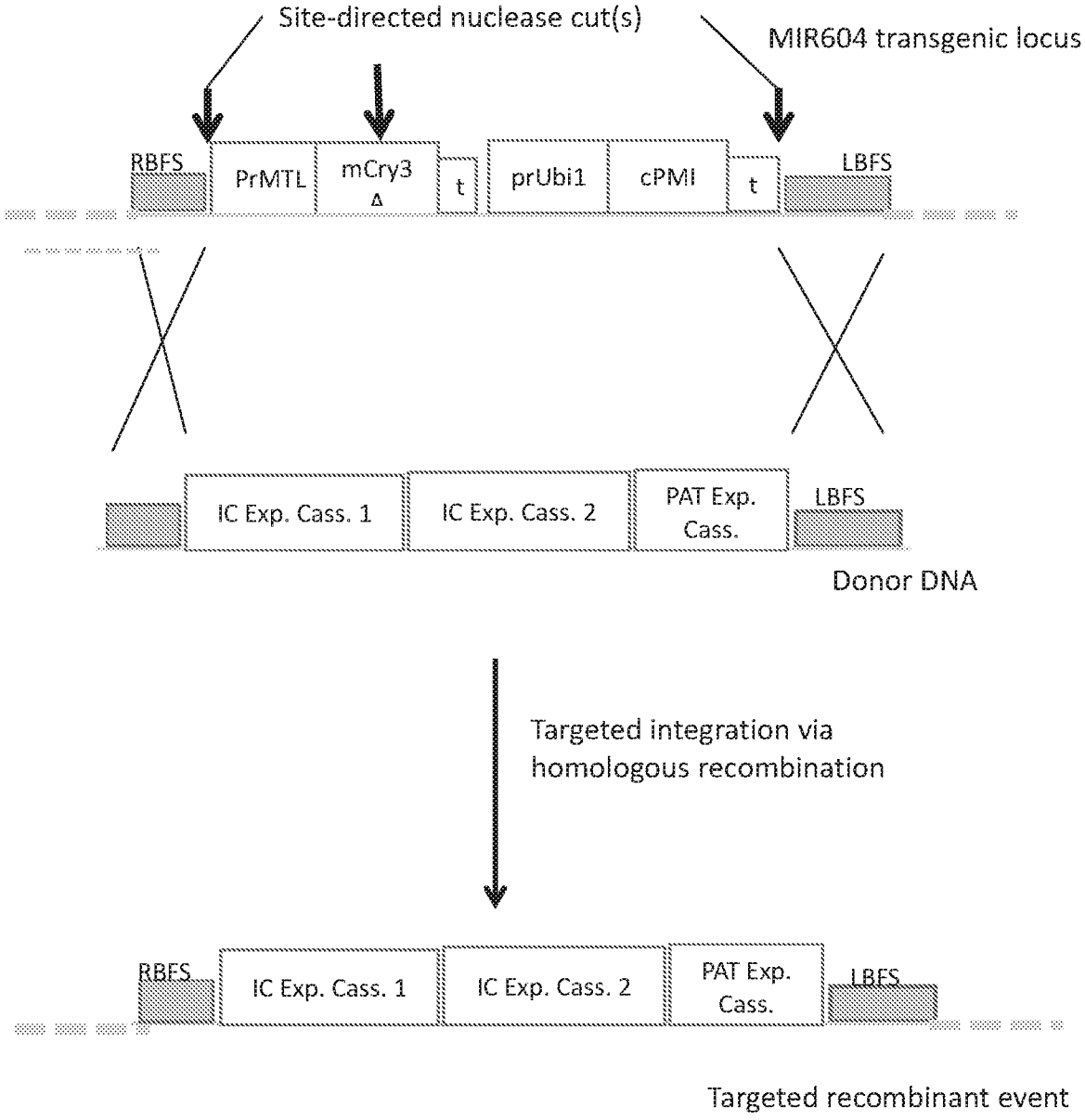

Site-directed nuclease cut(s)
MIR604 transgenic locus
RBFS
PrMTL
mCry3 Δ
t
prUbi1
cPMI
t
LBFS
IC Exp. Cass. 1
IC Exp. Cass. 2
PAT Exp. Cass.
LBFS
Donor DNA
Targeted integration via homologous recombination
RBFS
IC Exp. Cass. 1
IC Exp. Cass. 2
PAT Exp. Cass.
LBFS
Targeted recombinant event

FIG. 12

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/066619

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/82
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

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 , BIOSIS, Sequence Search , EMBASE, PAJ , WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2014/161821 Al (BAYER CROPSCI ENCE NV [BE]] 9 October 2014 (2014-10-09)	1-53
Y	abstract paragraph [0037] paragraph [0060] paragraph [0074] paragraph [0078] - paragraph [0084] paragraph [0093] - paragraph [0095] cl aims 1-22 paragraph [0031] ; figures 1-5 paragraph [0041] paragraph [0132]	1-53

	-/ - -	

[X] Further documents are listed in the continuation of Box C.

[X] 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) orwhich 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
11 March 2016	29/03/2016

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	Kel l er, Yves

2

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2005/216970 Al (STEINER HENRY-YORK [US] ET AL) 29 September 2005 (2005-09-29) abstract paragraph [0007] - paragraph [0008] paragraph [0014] - paragraph [0019] paragraph [0084] - paragraph [0086] paragraph [0091] paragraph [0092] exampl e 2 exampl e 4 exampl e 5 cl aims 1-43	1-53
X	Versi on ET AL: "For food testing purposes . FOR I N VITRO USE ONLY, foodproof SL GM0 MI R604 Mai ze Detection Ki t -5 'Nuclease - Order No. Z 720 06 Ki t for 50 reacti ons for a maximum of 48 sampl es Store the ki t at -15 t o -25 °C", , 1 June 2014 (2014-06-01) , XP55255618, Retri eved from the Internet: URL:http: //www. bc-diagnosti cs. com/download s/products/bpz/Z _720_06_20_foodproof _SL_GM 0_MI R604_Mai ze Detecti on Ki t_Vl-l .pdf [retri eved on 2016-03-07] " the whole document	38
X	QIUDENG QUE ET AL: "Mai ze transformation technology development for commercial event generati on", FRONTI ERS I N PLANT SCI ENCE, vol . 5, 5 August 2014 (2014-08-05) , XP055217826, D0I : 10. 3389/fpl s. 2014. 00379 abstract table 1 page 12, col umn 2 - page 15, col umn 1, paragraph 1	1-53
Y	W0 2007/142840 A2 (SYNGENTA PARTICI PATIONS AG [CH] ; LONG NYKOLL [US] ; PULLIAM DERRICK [US) 13 December 2007 (2007-12-13) paragraph [0005]	1-53

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/066619

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LIANG ZHEN ET AL: "Targeted Mutagenesi s inZea maysUsi ng TALENs and the CRISPR/Cas System" , JOURNAL OF GENETICS AND GENOMICS, ELSEVI ER BV, NL, vol . 41, no. 2 , 14 December 2013 (2013-12-14) , pages 63-68, XP028661345, ISSN : 1673-8527, DOI : 10. 1016/J .JGG.2013. 12 .001 the whole document	24-37 , 39-43
Y	LUISA BORTESI ET AL: "The CRISPR/Cas9 system for pl ant genome edi t i ng and beyond" , BIOTECHNOLOGY ADVANCES, vol . 33, no. 1 , 20 December 2014 (2014-12-20) , pages 41-52 , XP55217852 , ISSN : 0734-9750, DOI : 10. 1016/ j . bi otechadv .2014 . 12 . 006 the whole document	1-53
A	W0 2013/026740 A2 (BAYER CROPSCI ENCE NV [BE] ; BAYER CROPSCI ENC AG [DE] ; BAYER CROPSCI ENCE) 28 February 2013 (2013-02-28) table 1	1-53

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2015/066619

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2014161821 Al	09-10-2014	AU 2014247151 Al	08-10-2015
		CA 2908403 Al	09-10-2014
		CN 105121650 A	02-12-2015
		EP 2981614 Al	10-02-2016
		US 2016053274 Al	25-02-2016
		Wo 2014161821 Al	09-10-2014
US 2005216970 Al	29-09-2005	AR 048190 Al	05-04-2006
		AR 089141 A2	30-07-2014
		AU 2005235963 Al	03-11-2005
		BR PI0502582 A	16-05-2006
		CA 2559481 Al	03-11-2005
		CN 1933724 A	21-03-2007
		CN 102260668 A	30-11-2011
		EP 1737290 A2	03-01-2007
		EP 2281447 A2	09-02-2011
		EP 2289311 A2	02-03-2011
		ES 2539110 T3	26-06-2015
		JP 2007530036 A	01-11-2007
		KR 20070012382 A	25-01-2007
		KR 20130023409 A	07-03-2013
		KR 20130025447 A	11-03-2013
		MX 273859 B	04-02-2010
		NZ 549466 A	25-06-2010
		PT 1737290 E	02-07-2015
		RU 2351122 C2	10-04-2009
		UA 94893 C2	25-06-2011
		US 2005216970 Al	29-09-2005
		US 2008167456 Al	10-07-2008
		US 2011111420 Al	12-05-2011
		US 2012185973 Al	19-07-2012
		WO 2005103301 A2	03-11-2005
W0 2007142840 A2	13-12-2007	AP 2726 A	31-08-2013
		AR 061164 Al	06-08-2008
		AU 2007257230 Al	13-12-2007
		AU 2011201815 Al	19-05-2011
		BR PI0712392 A2	16-10-2012
		CA 2653992 Al	13-12-2007
		CA 2759093 Al	13-12-2007
		CL 2007001582 Al	18-04-2008
		CN 101548011 A	30-09-2009
		EA 200802384 Al	30-06-2009
		EA 201100469 Al	30-12-2011
		EP 2032700 A2	11-03-2009
		EP 2468902 Al	27-06-2012
		ES 2473602 T3	07-07-2014
		ES 2546255 T3	22-09-2015
		KR 20090023669 A	05-03-2009
		KR 20130019014 A	25-02-2013
		KR 20130020849 A	28-02-2013
		PH 12014502646 Al	14-12-2015
		PT 2032700 E	24-06-2014
		PT 2468902 E	12-10-2015
		UA 95637 C2	25-08-2011
		US 2009300784 Al	03-12-2009
		US 2012167260 Al	28-06-2012
		US 2012264130 Al	18-10-2012

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2015/0666 19

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
		US 2012272402 Al	25--10--2012
		us 2014162272 Al	12--06--2014
		Wo 2007142840 A2	13--12--2007
		ZA 200810169 A	26--08--2009
W0 2013026740 A2	28-02-2013	AR 088749 Al	02--07--2014
		AU 2012299691 Al	02--05--2013
		CN 103890181 A	25--06--2014
		CN 103981149 A	13--08--2014
		EP 2748323 A2	02--07--2014
		US 2014196169 Al	10--07--2014
		US 2015167009 Al	18--06--2015
		Wo 2013026740 A2	28--02--2013