



US 20250263732A1

(19) United States

(12) Patent Application Publication

Nuccio et al.

(10) Pub. No.: US 2025/0263732 A1

(43) Pub. Date: Aug. 21, 2025

## (54) INIR20 TRANSGENIC SOYBEAN WITH JUNCTION POLYNUCLEOTIDE DELETIONS

(71) Applicant: INARI AGRICULTURE TECHNOLOGY, INC., Cambridge, MA (US)

(72) Inventors: Michael Lee Nuccio, Salem, NH (US); Michael Andreas Kock, Rheinfelden (DE); Joshua L. Price, Cambridge, MA (US)

(21) Appl. No.: 19/197,017

(22) Filed: May 2, 2025

## Related U.S. Application Data

(63) Continuation of application No. 18/481,023, filed on Oct. 4, 2023, now Pat. No. 12,344,849, which is a continuation of application No. 18/058,156, filed on Nov. 22, 2022, now Pat. No. 11,814,632, which is a continuation of application No. PCT/US2021/043933, filed on Jul. 30, 2021.

(60) Provisional application No. 63/203,137, filed on Jul. 9, 2021, provisional application No. 63/202,569, filed on Jun. 16, 2021, provisional application No. 63/201,030, filed on Apr. 9, 2021, provisional application No. 63/201,029, filed on Apr. 9, 2021, provisional application No. 63/199,951, filed on Feb. 4, 2021, provisional application No. 63/199,949, filed on Feb. 4, 2021, provisional application No. 63/199,930, filed on Feb. 3, 2021, provisional application No. 63/059,813, filed on Jul. 31, 2020, provisional application No. 63/059,860, filed on Jul. 31, 2020, provisional

application No. 63/059,916, filed on Jul. 31, 2020, provisional application No. 63/059,963, filed on Jul. 31, 2020.

## Publication Classification

## (51) Int. Cl.

C12N 15/82	(2006.01)
A01H 1/02	(2006.01)
A01H 5/10	(2018.01)
A01H 6/46	(2018.01)
A01H 6/54	(2018.01)
C07K 14/415	(2006.01)
C12N 9/22	(2006.01)
C12N 15/11	(2006.01)
C12Q 1/6834	(2018.01)
C12Q 1/6895	(2018.01)

## (52) U.S. Cl.

CPC ..... C12N 15/8201 (2013.01); A01H 1/02 (2013.01); A01H 5/10 (2013.01); A01H 6/4684 (2018.05); A01H 6/542 (2018.05); C07K 14/415 (2013.01); C12N 9/22 (2013.01); C12N 15/11 (2013.01); C12N 15/8213 (2013.01); C12N 15/8286 (2013.01); C12Q 1/6834 (2013.01); C12Q 1/6895 (2013.01); C12N 2310/20 (2017.05); C12N 2800/80 (2013.01); C12Q 2600/13 (2013.01); C12Q 2600/156 (2013.01); C12Q 2600/158 (2013.01)

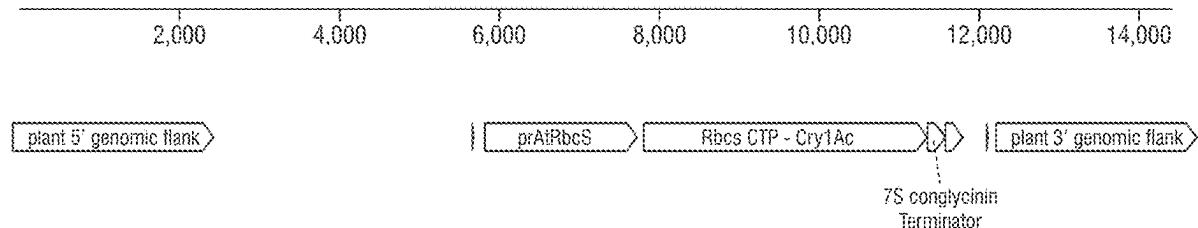
## (57)

## ABSTRACT

Transgenic INIR20 soybean plants comprising modifications of the MON87701 soybean locus which provide for facile excision of the modified MON87701 transgenic locus or portions thereof, methods of making such plants, and use of such plants to facilitate breeding are disclosed.

Specification includes a Sequence Listing.

## MON87701 insert (14416 bp)



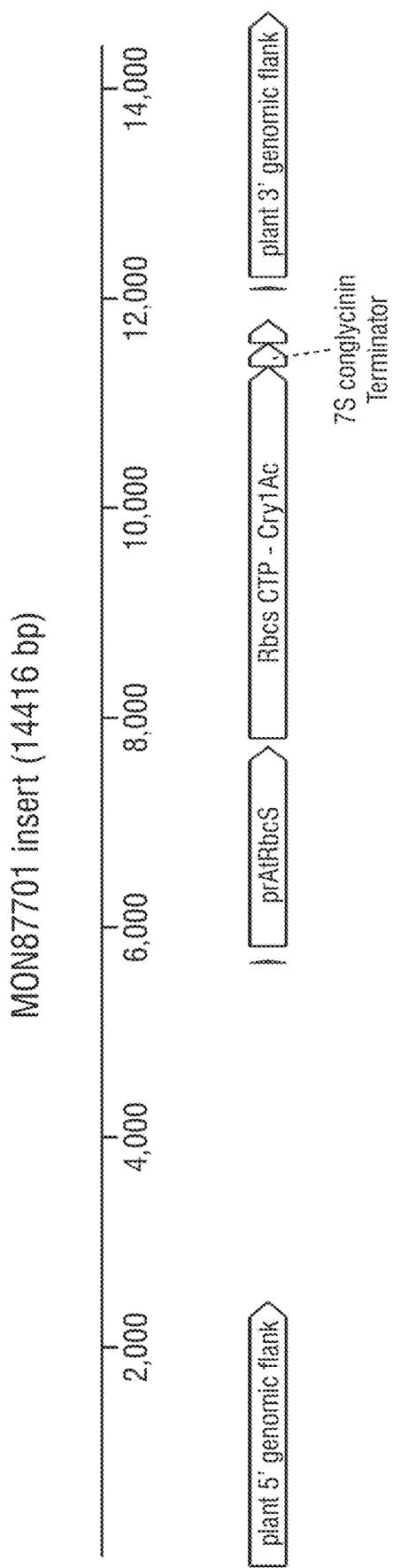


FIG. 1

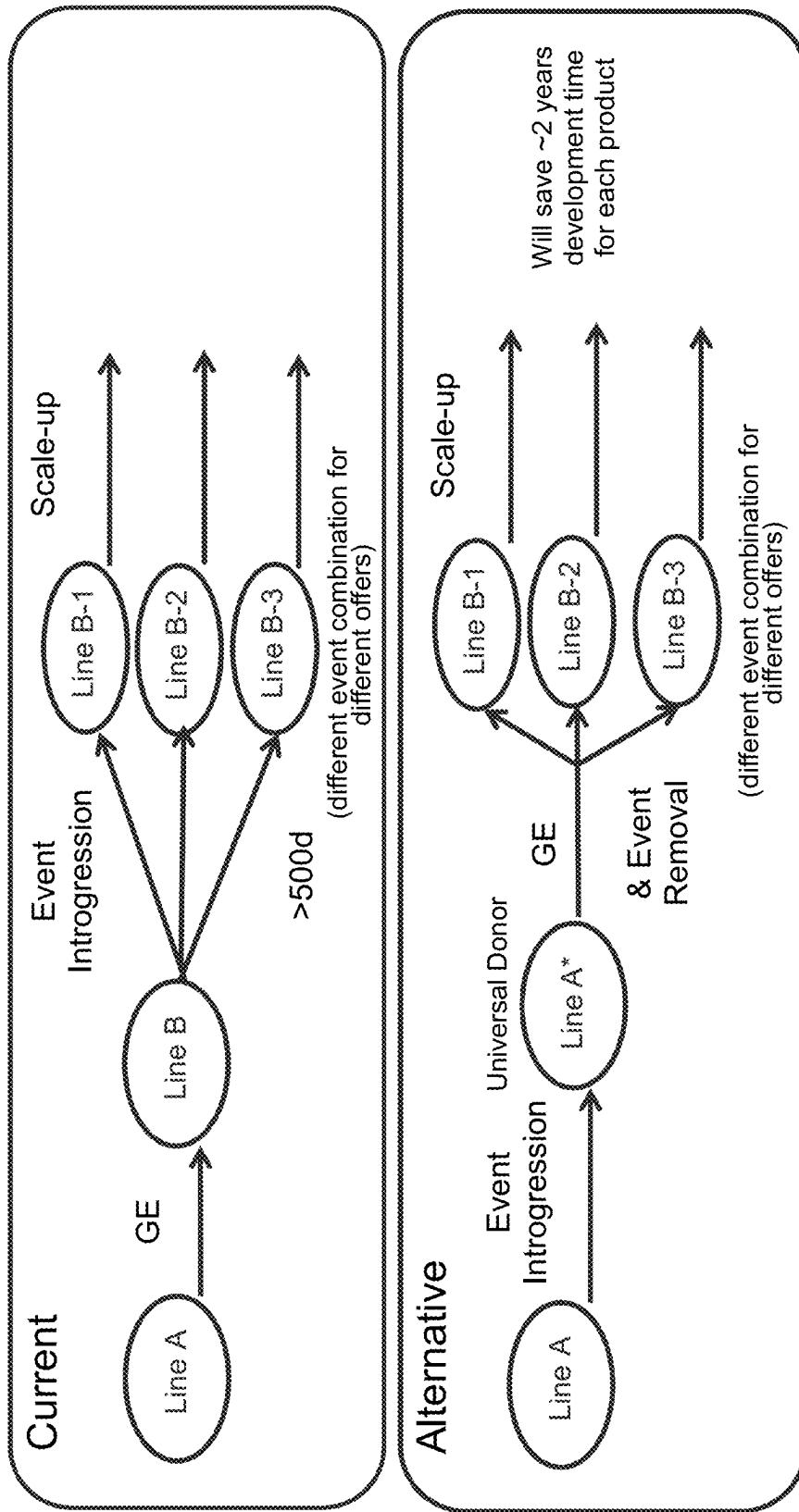


FIG. 2

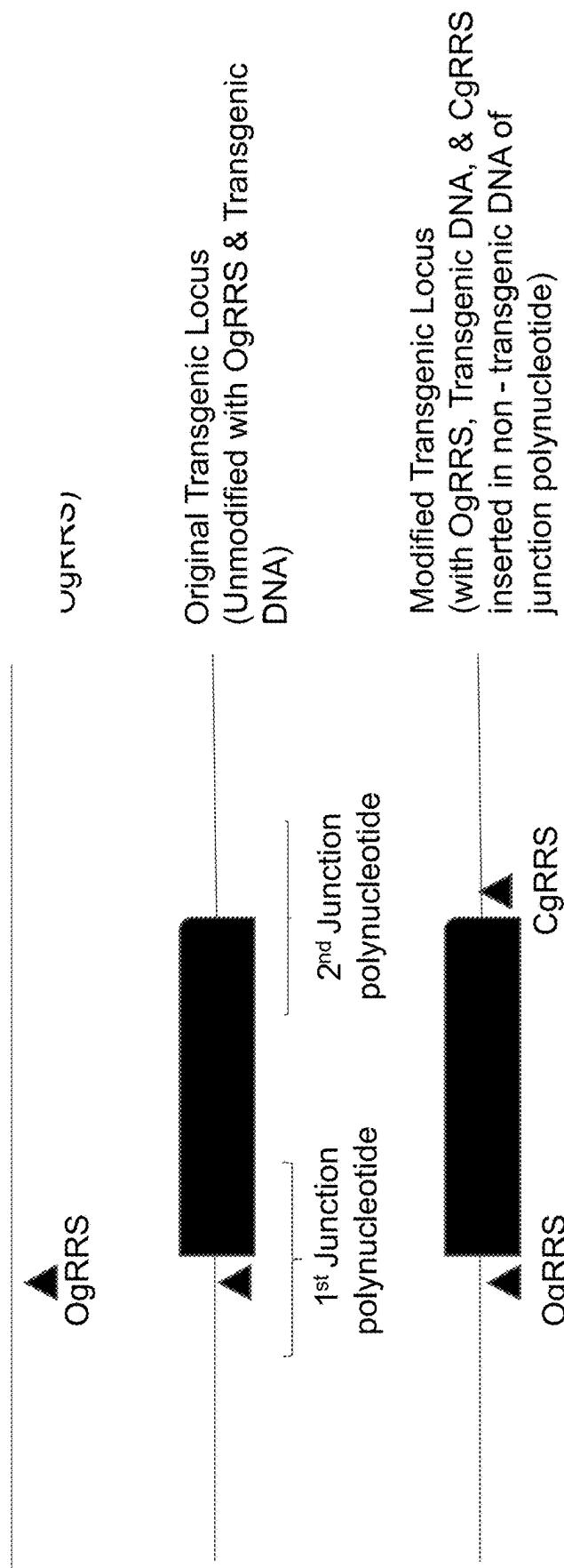


FIG. 3A

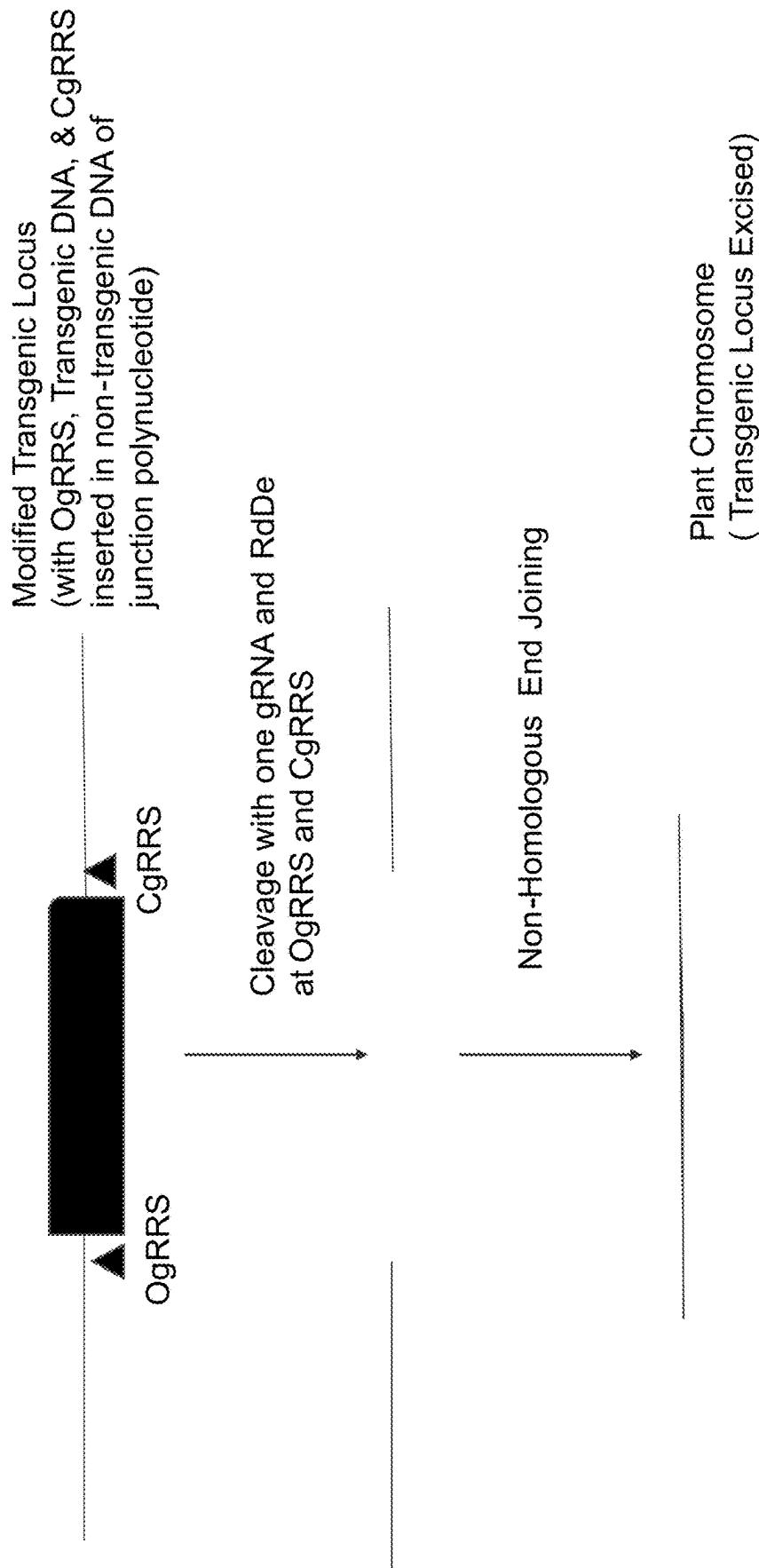


FIG. 3B

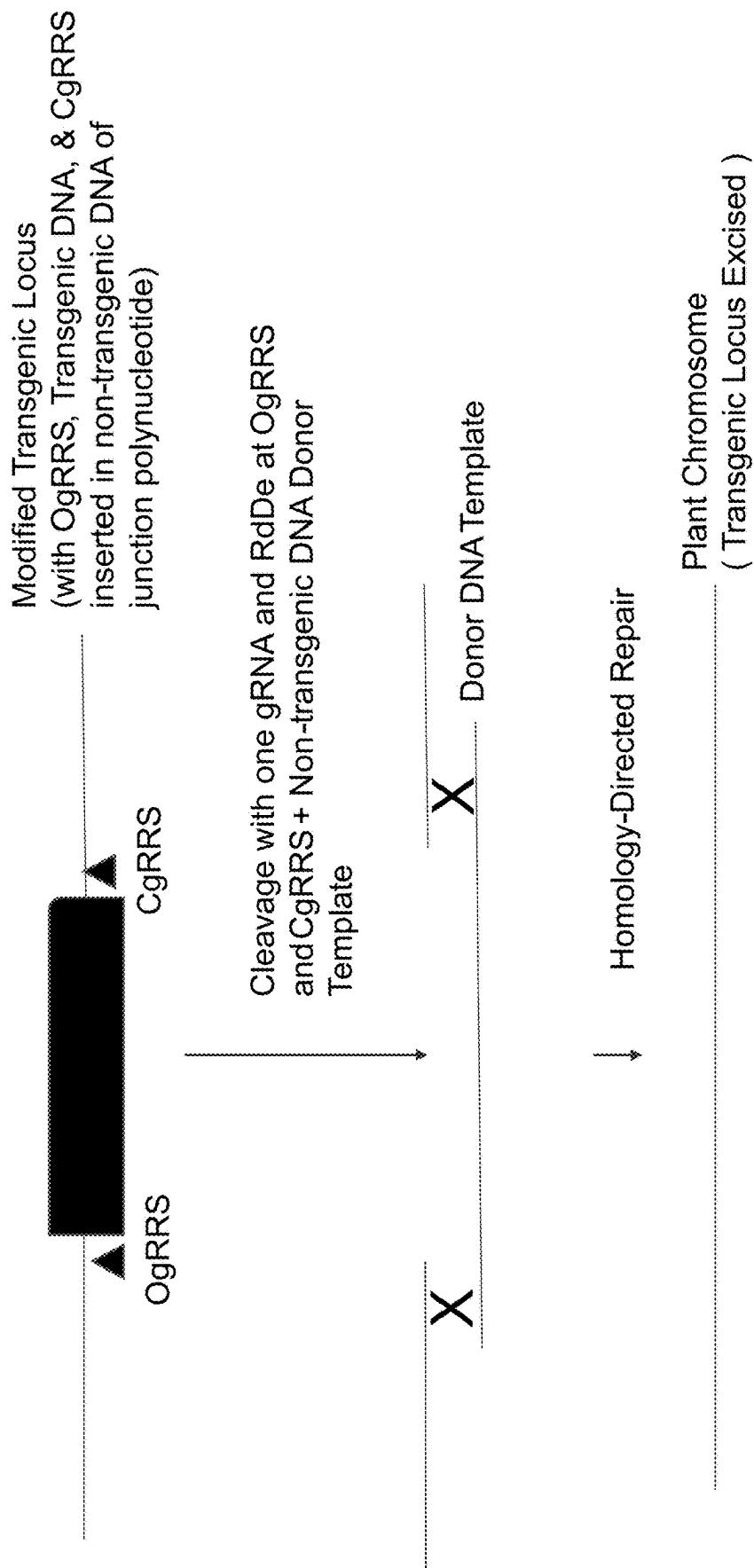


FIG. 3C

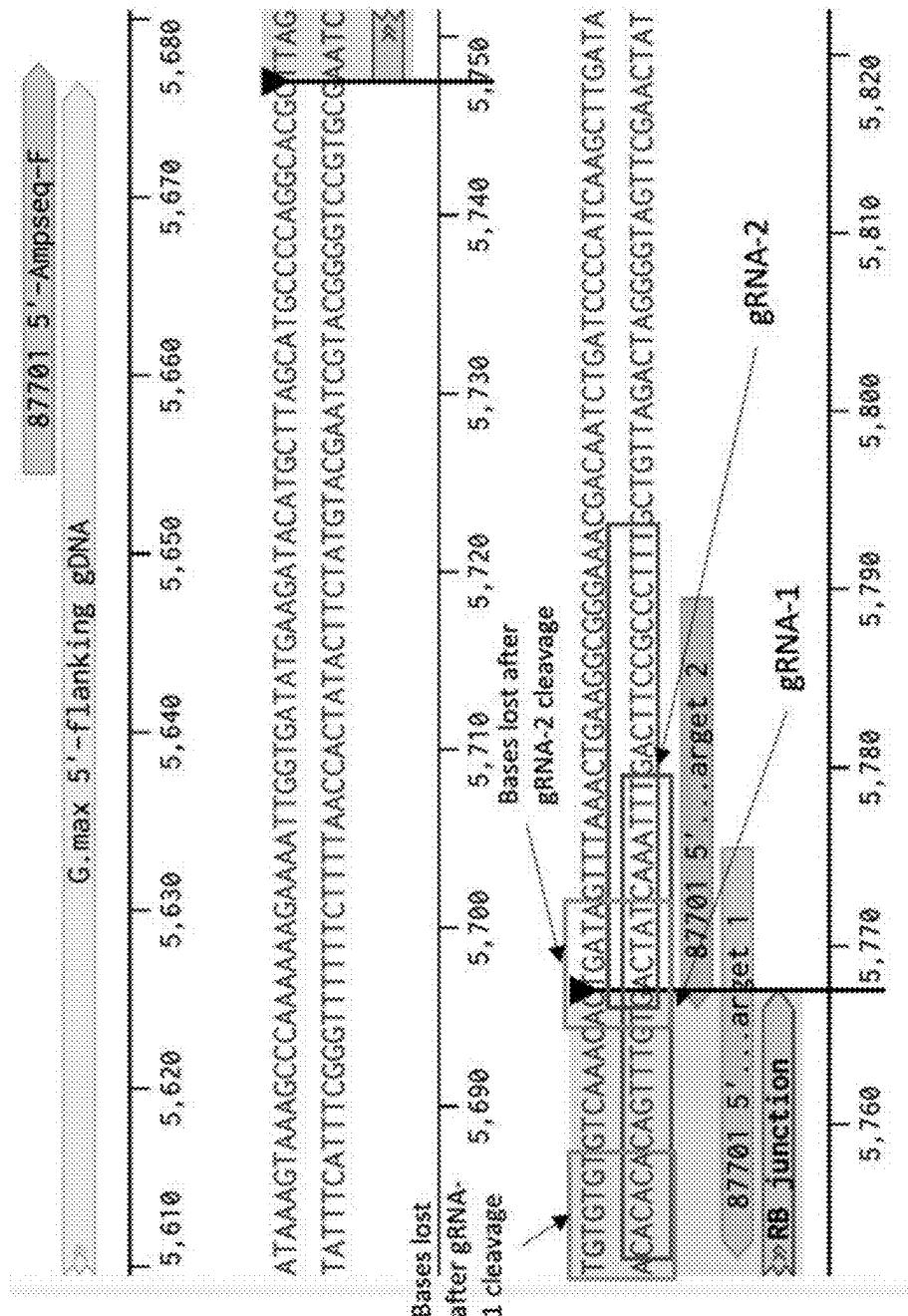


FIG. 4

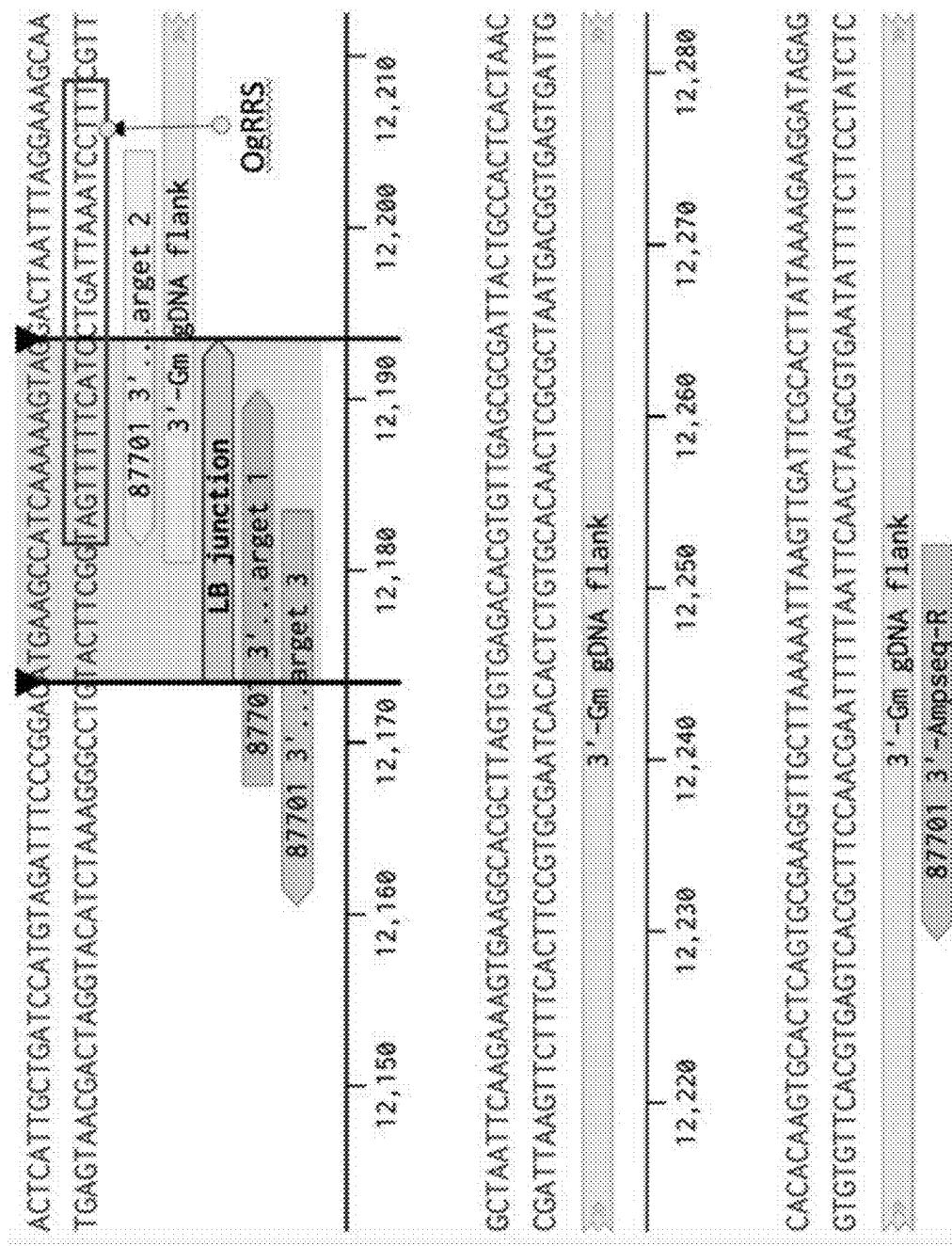


FIG. 5

**INIR20 TRANSGENIC SOYBEAN WITH JUNCTION POLYNUCLEOTIDE DELETIONS****CROSS-REFERENCE TO RELATED APPLICATIONS**

**[0001]** This is a continuation application of U.S. patent application Ser. No. 18/481,023, filed Oct. 4, 2023, which is a continuation of U.S. patent application Ser. No. 18/058,156, filed Nov. 22, 2022, now U.S. Pat. No. 11,814,632, which is a continuation of International Application No. PCT/US2021/043,933, filed on Jul. 30, 2021, which claims the benefit of priority to U.S. Provisional Patent Application No. 63/203,137, filed on Jul. 9, 2021, U.S. Provisional Patent Application No. 63/202,569, filed on Jun. 16, 2021, U.S. Provisional Patent Application Nos. 63/201,030 and 63/201,029, filed on Apr. 9, 2021, U.S. Provisional Patent Application Nos. 63/199,951 and 63/199,949, filed on Feb. 4, 2021, U.S. Provisional Patent Application No. 63/199,930, filed on Feb. 3, 2021, and U.S. Provisional Patent Application Nos. 63/059,813, 63/059,860, 63/059,916, and 63/059,963, filed on Jul. 31, 2020, which are each incorporated herein by reference in their entireties.

**REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY**

**[0002]** The instant application contains a Sequence Listing which has been submitted electronically in XML formatted and is herein incorporated by reference in its entirety. Said XML copy, created on May 1, 2025, is named "P13649US02.xml" and is 91,478 bytes in size.

**BACKGROUND**

**[0003]** Transgenes which are placed into different positions in the plant genome through non-site specific integration can exhibit different levels of expression (Weising et al., 1988, *Ann. Rev. Genet.* 22:421-477). Such transgene insertion sites can also contain various undesirable rearrangements of the foreign DNA elements that include deletions and/or duplications. Furthermore, many transgene insertion sites can also comprise selectable or scoreable marker genes which in some instances are no longer required once a transgenic plant event containing the linked transgenes which confer desirable traits are selected.

**[0004]** Commercial transgenic plants typically comprise one or more independent insertions of transgenes at specific locations in the host plant genome that have been selected for features that include expression of the transgene(s) of interest and the transgene-conferred trait(s), absence or minimization of rearrangements, and normal Mendelian transmission of the trait(s) to progeny. An example of a selected transgenic soybean event which confers resistance to lepidopteran insects is the MON87701 transgenic soybean event disclosed in U.S. Pat. No. 8,049,071. MON87701 transgenic soybean plants express a Cry1Ac protein which confers resistance to lepidopteran insects which include Velvetbean caterpillar (*Anticarsia gemmatalis*), Soybean looper (*Pseudaletia includens*), Soybean axil borer (*Episotia aporema*), Yellow Bear Moth (*Spilosoma virginica*), Corn earworm (*Helicoverpa zea*), Fall armyworm (*Spodoptera frugiperda*), and Sunflower looper (*Rachiplusia nu*).

**[0005]** Methods for removing selectable marker genes and/or duplicated transgenes in transgene insertion sites in plant genomes involving use of site-specific recombinase

systems (e.g., cre-lox) as well as for insertion of new genes into transgene insertion sites have been disclosed (Srivastava and Ow; *Methods Mol Biol.* 2015, 1287:95-103; Dale and Ow, 1991, *Proc. Natl Acad. Sci. USA* 88, 10558-10562; Srivastava and Thomson, *Plant Biotechnol J.*, 2016; 14 (2): 471-82). Such methods typically require incorporation of the recombination site sequences recognized by the recombinase at particular locations within the transgene.

**SUMMARY**

**[0006]** Transgenic soybean plant cells comprising an INIR20 transgenic locus comprising an originator guide RNA recognition site (OgRRS) in a first DNA junction polynucleotide of a MON87701 transgenic locus and a cognate guide RNA recognition site (CgRRS) in a second DNA junction polynucleotide of the MON87701 transgenic locus are provided. Transgenic soybean plant cells comprising an INIR20 transgenic locus comprising an insertion and/or substitution in a DNA junction polynucleotide of a MON87701 transgenic locus of DNA comprising a cognate guide RNA recognition site (CgRRS) are provided. In certain embodiments, the MON87701 transgenic locus is set forth in SEQ ID NO:1, is present in seed deposited at the ATCC under accession No. PTA-8194 is present in progeny thereof, is present in allelic variants thereof, or is present in other variants thereof. INIR20 transgenic soybean plant cells, transgenic soybean plant seeds, and transgenic soybean plants all comprising a transgenic locus set forth in SEQ ID NO: 2, 3, or 15 are provided. Transgenic soybean plant parts including seeds and transgenic soybean plants comprising the soybean plant cells are also provided.

**[0007]** Methods for obtaining a bulked population of inbred seed comprising selfing the aforementioned transgenic soybean plants and harvesting seed comprising the INIR20 transgenic locus from the selfed soybean plant are also provided.

**[0008]** Methods of obtaining hybrid soybean seed comprising crossing the aforementioned transgenic soybean plants to a second soybean plant which is genetically distinct from the first soy bean plant and harvesting seed comprising the INIR20 transgenic locus from the cross are provided. Methods for obtaining a bulked population of seed comprising selfing a transgenic soybean plant of comprising SEQ ID NO: 2, 3, or 15 and harvesting transgenic seed comprising the transgenic locus set forth in SEQ ID NO: 2, 3, or 15 are provided.

**[0009]** A DNA molecule comprising SEQ ID NO: 2, 3, 7, 8, 9, 14, 15, or 21 is provided. Processed transgenic soybean plant products and biological samples comprising the DNA molecules are provided. Nucleic acid molecules adapted for detection of genomic DNA comprising the DNA molecules, wherein said nucleic acid molecule optionally comprises a detectable label are provided. Methods of detecting a soybean plant cell comprising any aforementioned INIR20 transgenic locus, comprising the step of detecting a DNA molecule comprising SEQ ID NO: 2, 3, 7, 8, 9, 14, 15, or 21 are provided.

**[0010]** Methods of excising the INIR20 transgenic locus from the genome of the aforementioned soybean plant cells comprising the steps of: (a) contacting the edited transgenic plant genome of the plant cell with: (i) an RNA dependent DNA endonuclease (RdDe); and (ii) a guide RNA (gRNA) capable of hybridizing to the guide RNA hybridization site of the OgRRS and the CgRRS; wherein the RdDe recog-

nizes a OgRRS/gRNA and a CgRRS/gRNA hybridization complex; and, (b) selecting a transgenic plant cell, transgenic plant part, or transgenic plant wherein the INIR20 transgenic locus flanked by the OgRRS and the CORRS has been excised.

#### BRIEF DESCRIPTION OF THE DRAWINGS/FIGURES

[0011] FIG. 1 shows a schematic diagram of the MON87701 transgenic locus.

[0012] FIG. 2 shows a schematic diagram which compares current breeding strategies for introgression of transgenic events (i.e., transgenic loci) to alternative breeding strategies for introgression of transgenic events where the transgenic events (i.e., transgenic loci) can be removed following introgression to provide different combinations of transgenic traits. In FIG. 2, “GE” refers to genome editing (e.g., including introduction of targeted genetic changes with genome editing molecules) and “Event Removal” refers to excision of a transgenic locus (i.e., an “Event”) with genome editing molecules.

[0013] FIG. 3A, B, C. FIG. 3A shows a schematic diagram of a non-limiting example of: (i) an untransformed plant chromosome containing non-transgenic DNA which includes the originator guide RNA recognition site (OgRRS) (top); (ii) the original transgenic locus with the OgRRS in the non-transgenic DNA of the 1<sup>st</sup> junction polynucleotide (middle); and (iii) the modified transgenic locus with a cognate guide RNA inserted into the non-transgenic DNA of the 2<sup>nd</sup> junction polynucleotide (bottom). FIG. 3B shows a schematic diagram of a non-limiting example of a process where a modified transgenic locus with a cognate guide RNA inserted into the non-transgenic DNA of the 2<sup>nd</sup> junction polynucleotide (top) is subjected to cleavage at the OgRRS and CgRRS with one guide RNA (gRNA) that hybridizes to gRNA hybridization site in both the OgRRS and the CgRRS and an RNA dependent DNA endonuclease (RdDc) that recognizes and cleaves the gRNA/OgRRS and the gRNA/CgRRS complex followed by non-homologous end joining processes to provide a plant chromosome where the transgenic locus is excised. FIG. 3C shows a schematic diagram of a non-limiting example of a process where a modified transgenic locus with a cognate guide RNA inserted into the non-transgenic DNA of the 2<sup>nd</sup> junction polynucleotide (top) is subjected to cleavage at the OgRRS and CgRRS with one guide RNA (gRNA) that hybridizes to the gRNA hybridization site in both the OgRRS and the CgRRS and an RNA dependent DNA endonuclease (RdDc) that recognizes and cleaves the gRNA/OgRRS and the gRNA/CgRRS complex in the presence of a donor DNA template. In FIG. 3C, cleavage of the modified transgenic locus in the presence of the donor DNA template which has homology to non-transgenic DNA but lacks the OgRRS in the 1<sup>st</sup> and 2<sup>nd</sup> junction polynucleotides followed by homology-directed repair processes to provide a plant chromosome where the transgenic locus is excised and non-transgenic DNA present in the untransformed plant chromosome is at least partially restored.

[0014] FIG. 4 illustrates the locations of gRNA-1 (SEQ ID NO: 4) and gRNA-2 (SEQ ID NO: 5) recognition sites in the 5' junction polynucleotide of SEQ ID NO: 1. Sequences in the figure are the corresponding sequences of SEQ ID NO: 1 and their reverse complement.

[0015] FIG. 5 illustrates the location of the OgRRS of SEQ ID NO: 6 in the 3' junction polynucleotide of SEQ ID NO: 1. Sequences in the figure are the corresponding sequences of SEQ ID NO: 1 and their reverse complement.

#### DETAILED DESCRIPTION

[0016] Unless otherwise stated, nucleic acid sequences in the text of this specification are given, when read from left to right, in the 5' to 3' direction. Nucleic acid sequences may be provided as DNA or as RNA, as specified; disclosure of one necessarily defines the other, as well as necessarily defines the exact complements, as is known to one of ordinary skill in the art.

[0017] Where a term is provided in the singular, the inventors also contemplate embodiments described by the plural of that term.

[0018] The term “about” as used herein means a value or range of values which would be understood as an equivalent of a stated value and can be greater or lesser than the value or range of values stated by 10 percent. Each value or range of values preceded by the term “about” is also intended to encompass the embodiment of the stated absolute value or range of values.

[0019] The phrase “allelic variant” as used herein refers to a polynucleotide or polypeptide sequence variant that occurs in a different strain, variety, or isolate of a given organism.

[0020] The term “and/or” where used herein is to be taken as specific disclosure of each of the two specified features or components with or without the other. Thus, the term and/or” as used in a phrase such as “A and/or B” herein is intended to include “A and B,” “A or B,” “A” (alone), and “B” (alone). Likewise, the term “and/or” as used in a phrase such as “A, B, and/or C” is intended to encompass each of the following embodiments: A, B, and C; A, B, or C; A or C; A or B; B or C; A and C; A and B; B and C; A (alone); B (alone); and C (alone).

[0021] As used herein, the phrase “approved transgenic locus” is a genetically modified plant event which has been authorized, approved, and/or de-regulated for any one of field testing, cultivation, human consumption, animal consumption, and/or import by a governmental body. Illustrative and non-limiting examples of governmental bodies which provide such approvals include the Ministry of Agriculture of Argentina, Food Standards Australia New Zealand, National Biosafety Technical Committee (CTNBio) of Brazil, Canadian Food Inspection Agency, China Ministry of Agriculture Biosafety Network, European Food Safety Authority, US Department of Agriculture, US Department of Environmental Protection, and US Food and Drug Administration.

[0022] The term “backcross”, as used herein, refers to crossing an F1 plant or plants with one of the original parents. A backcross is used to maintain or establish the identity of one parent (species) and to incorporate a particular trait from a second parent (species). The term “backcross generation”, as used herein, refers to the offspring of a backcross.

[0023] As used herein, the phrase “biological sample” refers to either intact or non-intact (e.g., milled seed or plant tissue, chopped plant tissue, lyophilized tissue) plant tissue. It may also be an extract comprising intact or non-intact seed or plant tissue. The biological sample can comprise flour, meal, syrup, oil, starch, and cereals manufactured in whole or in part to contain crop plant by-products. In certain

embodiments, the biological sample is “non-regenerable” (i.e., incapable of being regenerated into a plant or plant part). In certain embodiments, the biological sample refers to a homogenate, an extract, or any fraction thereof containing genomic DNA of the organism from which the biological sample was obtained, wherein the biological sample does not comprise living cells.

[0024] As used herein, the terms “correspond,” “corresponding,” and the like, when used in the context of an nucleotide position, mutation, and/or substitution in any given polynucleotide (e.g., an allelic variant of SEQ ID NO: 1) with respect to the reference polynucleotide sequence (e.g., SEQ ID NO: 1) all refer to the position of the polynucleotide residue in the given sequence that has identity to the residue in the reference nucleotide sequence when the given polynucleotide is aligned to the reference polynucleotide sequence using a pairwise alignment algorithm (e.g., CLUSTAL O 1.2.4 with default parameters).

[0025] As used herein, the terms “Cpf1” and “Cas12a” are used interchangeably to refer to the same RNA dependent DNA endonuclease (RdDc). A Cas12a protein provided herein includes the protein of SEQ ID NO: 16.

[0026] The term “crossing” as used herein refers to the fertilization of female plants (or gametes) by male plants (or gametes). The term “gamete” refers to the haploid reproductive cell (egg or pollen) produced in plants by meiosis from a gametophyte and involved in sexual reproduction, during which two gametes of opposite sex fuse to form a diploid zygote. The term generally includes reference to a pollen (including the sperm cell) and an ovule (including the ovum). When referring to crossing in the context of achieving the introgression of a genomic region or segment, the skilled person will understand that in order to achieve the introgression of only a part of a chromosome of one plant into the chromosome of another plant, random portions of the genomes of both parental lines recombine during the cross due to the occurrence of crossing-over events in the production of the gametes in the parent lines. Therefore, the genomes of both parents must be combined in a single cell by a cross, where after the production of gametes from the cell and their fusion in fertilization will result in an introgression event.

[0027] As used herein, the phrases “DNA junction polynucleotide” and “junction polynucleotide” refers to a polynucleotide of about 18 to about 500 base pairs in length comprised of both endogenous chromosomal DNA of the plant genome and heterologous transgenic DNA which is inserted in the plant genome. A junction polynucleotide can thus comprise about 8, 10, 20, 50, 100, 200, 250, 500, or 1000 base pairs of endogenous chromosomal DNA of the plant genome and about 8, 10, 20, 50, 100, 200, 250, 500, or 1000 base pairs of heterologous transgenic DNA which span the one end of the transgene insertion site in the plant chromosomal DNA. Transgene insertion sites in chromosomes will typically contain both a 5' junction polynucleotide and a 3' junction polynucleotide. In embodiments set forth herein in SEQ ID NO: 1, the 5' junction polynucleotide is located at the 5' end of the sequence and the 3' junction polynucleotide is located at the 3' end of the sequence. In a non-limiting and illustrative example, a 5' junction polynucleotide of a transgenic locus is telomere proximal in a chromosome arm and the 3' junction polynucleotide of the transgenic locus is centromere proximal in the same chromosome arm. In another non-limiting and illustrative

example, a 5' junction polynucleotide of a transgenic locus is centromere proximal in a chromosome arm and the 3' junction polynucleotide of the transgenic locus is telomere proximal in the same chromosome arm. The junction polynucleotide which is telomere proximal and the junction polynucleotide which is centromere proximal can be determined by comparing non-transgenic genomic sequence of a sequenced non-transgenic plant genome to the non-transgenic DNA in the junction polynucleotides.

[0028] The term “donor,” as used herein in the context of a plant, refers to the plant or plant line from which the trait, transgenic event, or genomic segment originates, wherein the donor can have the trait, introgression, or genomic segment in either a heterozygous or homozygous state.

[0029] As used herein, the term “MON87701” is used to refer to any of a transgenic soybean locus, transgenic soybean plants and parts thereof including seed set forth in U.S. Pat. No. 8,049,071, which is incorporated herein by reference in its entirety. Representative MON87701 transgenic soybean seed have been deposited with American Type Culture Collection (ATCC, Manassas, Va. 20110-2209 USA) under Accession No. PTA-8194. MON87701 transgenic loci include loci having the sequence of SEQ ID NO: 1, the sequence of the MON87701 locus in the deposited seed of Accession No. PTA-8194 and any progeny thereof, as well as allelic variants and other variants of SEQ ID NO: 1.

[0030] As used herein, the terms “excise” and “delete,” when used in the context of a DNA molecule, are used interchangeably to refer to the removal of a given DNA segment or element (e.g., transgene element or transgenic locus or portion thereof) of the DNA molecule.

[0031] As used herein, the phrase “elite crop plant” refers to a plant which has undergone breeding to provide one or more trait improvements. Elite crop plant lines include plants which are an essentially homozygous, e.g., inbred or doubled haploid. Elite crop plants can include inbred lines used as is or used as pollen donors or pollen recipients in hybrid seed production (e.g., used to produce F1 plants). Elite crop plants can include inbred lines which are selfed to produce non-hybrid cultivars or varieties or to produce (e.g., bulk up) pollen donor or recipient lines for hybrid seed production. Elite crop plants can include hybrid F1 progeny of a cross between two distinct elite inbred or doubled haploid plant lines.

[0032] As used herein, an “event,” “a transgenic event,” “a transgenic locus” and related phrases refer to an insertion of one or more transgenes at a unique site in the genome of a plant as well as to DNA fragments, plant cells, plants, and plant parts (e.g., seeds) comprising genomic DNA containing the transgene insertion. Such events typically comprise both a 5' and a 3' junction polynucleotide and confer one or more useful traits including herbicide tolerance, insect resistance, male sterility, and the like.

[0033] As used herein, the phrases “endogenous sequence,” “endogenous gene,” “endogenous DNA,” “endogenous polynucleotide,” and the like refer to the native form of a polynucleotide, gene or polypeptide in its natural location in the organism or in the genome of an organism.

[0034] The terms “exogenous” and “heterologous” as are used synonymously herein to refer to any polynucleotide (e.g., DNA molecule) that has been inserted into a new location in the genome of a plant. Non-limiting examples of an exogenous or heterologous DNA molecule include a

synthetic DNA molecule, a non-naturally occurring DNA molecule, a DNA molecule found in another species, a DNA molecule found in a different location in the same species, and/or a DNA molecule found in the same strain or isolate of a species, where the DNA molecule has been inserted into a new location in the genome of a plant.

[0035] As used herein, the term “F1” refers to any offspring of a cross between two genetically unlike individuals.

[0036] The term “gene,” as used herein, refers to a hereditary unit consisting of a sequence of DNA that occupies a specific location on a chromosome and that contains the genetic instruction for a particular characteristics or trait in an organism. The term “gene” thus includes a nucleic acid (for example, DNA or RNA) sequence that comprises coding sequences necessary for the production of an RNA, or a polypeptide or its precursor. A functional polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence as long as the desired activity or functional properties (e.g., enzymatic activity, pesticidal activity, ligand binding, and/or signal transduction) of the RNA or polypeptide are retained.

[0037] The term “identifying,” as used herein with respect to a plant, refers to a process of establishing the identity or distinguishing character of a plant, including exhibiting a certain trait, containing one or more transgenes, and/or containing one or more molecular markers.

[0038] As used herein, the term “INIR20” is used to refer either individually collectively to items that include any or all of the MON87701 transgenic soybean loci which have been modified as disclosed herein, modified MON87701 transgenic soybean plants and parts thereof including seed, and DNA obtained therefrom.

[0039] The term “isolated” as used herein means having been removed from its natural environment.

[0040] As used herein, the terms “include,” “includes,” and “including” are to be construed as at least having the features to which they refer while not excluding any additional unspecified features.

[0041] As used herein, the phrase “introduced transgene” is a transgene not present in the original transgenic locus in the genome of an initial transgenic event or in the genome of a progeny line obtained from the initial transgenic event. Examples of introduced transgenes include exogenous transgenes which are inserted in a resident original transgenic locus.

[0042] As used herein, the terms “introgression”, “introgressed” and “introgressing” refer to both a natural and artificial process, and the resulting plants, whereby traits, genes or DNA sequences of one species, variety or cultivar are moved into the genome of another species, variety or cultivar, by crossing those species. The process may optionally be completed by backcrossing to the recurrent parent. Examples of introgression include entry or introduction of a gene, a transgene, a regulatory element, a marker, a trait, a trait locus, or a chromosomal segment from the genome of one plant into the genome of another plant.

[0043] The phrase “marker-assisted selection”, as used herein, refers to the diagnostic process of identifying, optionally followed by selecting a plant from a group of plants using the presence of a molecular marker as the diagnostic characteristic or selection criterion. The process usually involves detecting the presence of a certain nucleic acid sequence or polymorphism in the genome of a plant.

[0044] The phrase “molecular marker”, as used herein, refers to an indicator that is used in methods for visualizing differences in characteristics of nucleic acid sequences. Examples of such indicators are restriction fragment length polymorphism (RFLP) markers, amplified fragment length polymorphism (AFLP) markers, single nucleotide polymorphisms (SNPs), microsatellite markers (e.g. SSRs), sequence-characterized amplified region (SCAR) markers, Next Generation Sequencing (NGS) of a molecular marker, cleaved amplified polymorphic sequence (CAPS) markers or isozyme markers or combinations of the markers described herein which defines a specific genetic and chromosomal location.

[0045] As used herein the terms “native” or “natural” define a condition found in nature. A “native DNA sequence” is a DNA sequence present in nature that was produced by natural means or traditional breeding techniques but not generated by genetic engineering (e.g., using molecular biology/transformation techniques).

[0046] The term “offspring”, as used herein, refers to any progeny generation resulting from crossing, selfing, or other propagation technique.

[0047] The phrase “operably linked” refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter is operably linked to a coding sequence if the promoter affects its transcription or expression. When the phrase “operably linked” is used in the context of a PAM site and a guide RNA hybridization site, it refers to a PAM site which permits cleavage of at least one strand of DNA in a polynucleotide with an RNA dependent DNA endonuclease or RNA dependent DNA nuclease which recognize the PAM site when a guide RNA complementary to guide RNA hybridization site sequences adjacent to the PAM site is present. A OgrRS and its CgRRS are operably linked to junction polynucleotides when they can be recognized by a gRNA and an RdDe to provide for excision of the transgenic locus or portion thereof flanked by the junction polynucleotides.

[0048] As used herein, the term “plant” includes a whole plant and any descendant, cell, tissue, or part of a plant. The term “plant parts” include any part(s) of a plant, including, for example and without limitation: seed (including mature seed and immature seed); a plant cutting; a plant cell; a plant cell culture; or a plant organ (e.g., pollen, embryos, flowers, fruits, shoots, leaves, roots, stems, and explants). A plant tissue or plant organ may be a seed, protoplast, callus, or any other group of plant cells that is organized into a structural or functional unit. A plant cell or tissue culture may be capable of regenerating a plant having the physiological and morphological characteristics of the plant from which the cell or tissue was obtained, and of regenerating a plant having substantially the same genotype as the plant. Regenerable cells in a plant cell or tissue culture may be embryos, protoplasts, meristematic cells, callus, pollen, leaves, anthers, roots, root tips, silk, flowers, kernels, ears, cobs, husks, or stalks. In contrast, some plant cells are not capable of being regenerated to produce plants and are referred to herein as “non-regenerable” plant cells.

[0049] The term “purified,” as used herein defines an isolation of a molecule or compound in a form that is substantially free of contaminants normally associated with the molecule or compound in a native or natural environment and means having been increased in purity as a result

of being separated from other components of the original composition. The term “purified nucleic acid” is used herein to describe a nucleic acid sequence which has been separated from other compounds including, but not limited to polypeptides, lipids and carbohydrates.

[0050] The term “recipient”, as used herein, refers to the plant or plant line receiving the trait, transgenic event or genomic segment from a donor, and which recipient may or may not have the trait, transgenic event or genomic segment itself either in a heterozygous or homozygous state.

[0051] As used herein the term “recurrent parent” or “recurrent plant” describes an elite line that is the recipient plant line in a cross and which will be used as the parent line for successive backcrosses to produce the final desired line.

[0052] As used herein the term “recurrent parent percentage” relates to the percentage that a backcross progeny plant is identical to the recurrent parent plant used in the backcross. The percent identity to the recurrent parent can be determined experimentally by measuring genetic markers such as SNPs and/or RFLPs or can be calculated theoretically based on a mathematical formula.

[0053] The terms “selfed,” “selfing,” and “self,” as used herein, refer to any process used to obtain progeny from the same plant or plant line as well as to plants resulting from the process. As used herein, the terms thus include any fertilization process wherein both the ovule and pollen are from the same plant or plant line and plants resulting therefrom. Typically, the terms refer to self-pollination processes and progeny plants resulting from self-pollination.

[0054] The term “selecting”, as used herein, refers to a process of picking out a certain individual plant from a group of individuals, usually based on a certain identity, trait, characteristic, and/or molecular marker of that individual.

[0055] As used herein, the phrase “originator guide RNA recognition site” or the acronym “OgRRS” refers to an endogenous DNA polynucleotide comprising a protospacer adjacent motif (PAM) site operably linked to a guide RNA hybridization site (i.e., protospacer sequence). In certain embodiments, an OgRRS can be located in an untransformed plant chromosome or in non-transgenic DNA of a DNA junction polynucleotide of both an original transgenic locus and a modified transgenic locus. In certain embodiments, an OgRRS can be located in transgenic DNA of a DNA junction polynucleotide of both an original transgenic locus and a modified transgenic locus. In certain embodiments,

an OgRRS can be located in both transgenic DNA and non-transgenic DNA of a DNA junction polynucleotide of both an original transgenic locus and a modified transgenic locus (i.e., can span transgenic and non-transgenic DNA in a DNA junction polynucleotide).

[0056] As used herein the phrase “cognate guide RNA recognition site” or the acronym “CgRRS” refer to a DNA polynucleotide comprising a PAM site operably linked to a guide RNA hybridization site (i.e., protospacer sequence), where the CgRRS is absent from transgenic plant genomes comprising a first original transgenic locus that is unmodified and where the CORRS and its corresponding OgRRS can hybridize to a single gRNA. A CgRRS can be located in transgenic DNA of a DNA junction polynucleotide of a modified transgenic locus, in transgenic DNA of a DNA junction polynucleotide of a modified transgenic locus, or in both transgenic and non-transgenic DNA of a modified transgenic locus (i.e., can span transgenic and non-transgenic DNA in a DNA junction polynucleotide).

[0057] As used herein, the phrase “a transgenic locus excision site” refers to the DNA which remains in the genome of a plant or in a DNA molecule (e.g., an isolated or purified DNA molecule) wherein a segment comprising, consisting essentially of, or consisting of a transgenic locus has been deleted. In a non-limiting and illustrative example, a transgenic locus excision site can thus comprise a contiguous segment of DNA comprising at least 10 base pairs of DNA that is telomere proximal to the deleted transgenic locus or to the deleted segment of the transgenic locus and at least 10 base pairs of DNA that is centromere proximal to the deleted transgenic locus or to the deleted segment of the transgenic locus.

[0058] As used herein, the phrase “transgene element” refers to a segment of DNA comprising, consisting essentially of, or consisting of a promoter, a 5' UTR, an intron, a coding region, a 3'UTR, or a polyadenylation signal. Polyadenylation signals include transgene elements referred to as “terminators” (e.g., NOS, pinII, rbcs, Hsp17, TubA).

[0059] To the extent to which any of the preceding definitions is inconsistent with definitions provided in any patent or non-patent reference incorporated herein by reference, any patent or non-patent reference cited herein, or in any patent or non-patent reference found elsewhere, it is understood that the preceding definition will be used herein.

[0060] Various sequences set forth in the sequence listing are described in the following table.

TABLE 1

Description of sequences.		
SEQ ID NO	Description	
1	MON87701 transgenic locus. The MON87701 5' flank region comprises nucleotides 1-5757 of SEQ ID NO: 1. The MON87701 transgenic insert extends from nucleotides 5758-12183 of SEQ ID NO: 1. The MON87701 3' flanking DNA comprises nucleotides 12184-14416 of SEQ ID NO: 1.	
2	INIR20-1 (G1 Cut)	
3	INIR20-2 (Insertion of 27 bp CgRRS of SEQ ID NO: 8 in 5' junction polynucleotide of MON87701 with donor DNA template of SEQ ID NO: 21))	
4	gRNA-1 targeting 5' junction polynucleotide of MON87701	
5	gRNA-2 targeting 5' junction polynucleotide of MON87701	
6	OgRRS (located in 3' junction polynucleotide of SEQ ID NO: 1)	
7	CgRRS + Flanking DNA (located in 5' junction polynucleotide of INIR20-3 transgenic locus of SEQ ID NO: 15)	

TABLE 1-continued

		Description of sequences.
SEQ ID NO	Description	
8	CgRRS + Flanking DNA (located in 5' junction polynucleotide of INIR20-2 transgenic locus of SEQ ID NO: 3)	
9	MON87701 donor template sequence #1 containing SEQ ID NO: 7 CgRRS to yield INIR20-3 transgenic locus of SEQ ID NO: 15	
10	MON87701 5' target insertion site	
11	MON87701 -gRNA coding sequence for gRNAs targeting CgRRS and OgRRS of INIR20-2 (SEQ ID NO: 3) and INIR20-3 (SEQ ID NO: 15)	
12	MON87701 5' primer	
13	MON87701 3' primer	
14	MON87701 CgRRS and flank	
15	INIR20-3 resultant sequence (Insertion of CgRRS of SEQ ID NO: 7 in MON87701 5' junction polynucleotide with donor DNA template of SEQ ID NO: 9)	
16	(Cas12a Nuclease) (>sp U2UMQ6 CS12A_ACISB CRISPR-associated endonuclease Cas12a OS = <i>Acidaminococcus</i> sp. (strain BV3L6) OX = 1111120 GN = cas12a PE = 1 SV = 1)	
17	MON87701 5' Junction Polynucleotide	
18	MON87701 5' plant genomic flanking	
19	MON87701 3' Junction Polynucleotide	
20	MON87701 3' plant genomic flanking	
21	MON87701 donor template sequence #2 for SEQ ID NO: 8 CgRRS to yield INIR20-2 transgenic locus of SEQ ID NO: 3	

[0061] Genome editing molecules can permit introduction of targeted genetic change conferring desirable traits in a variety of crop plants (Zhang et al. *Genome Biol.* 2018; 19:210; Schindeler et al. *FEBS Lett.* 2018; 592 (12): 1954). Desirable traits introduced into crop plants such as soybean and soybean include herbicide tolerance, improved food and/or feed characteristics, male-sterility, and drought stress tolerance. Nonetheless, full realization of the potential of genome editing methods for crop improvement will entail efficient incorporation of the targeted genetic changes in germplasm of different elite crop plants adapted for distinct growing conditions. Such elite crop plants will also desirably comprise useful transgenic loci which confer various traits including herbicide tolerance, pest resistance (e.g.; insect, nematode, fungal disease, and bacterial disease resistance), conditional male sterility systems for hybrid seed production, abiotic stress tolerance (e.g., drought tolerance), improved food and/or feed quality, and improved industrial use (e.g., biofuel). Provided herein are methods whereby targeted genetic changes are efficiently combined with desired subsets of transgenic loci in elite progeny plant lines (e.g., elite inbreds used for hybrid seed production or for inbred varietal production). Also provided are plant genomes containing modified transgenic loci which can be selectively excised with a single gRNA molecule. Such modified transgenic loci comprise an originator guide RNA recognition site (OgRRS) which is identified in non-transgenic DNA of a first junction polynucleotide of the transgenic locus and cognate guide RNA recognition site (CgRRS) which is introduced (e.g., by genome editing methods) into a second junction polynucleotide of the transgenic locus and which can hybridize to the same gRNA as the OgRRS, thereby permitting excision of the modified transgenic locus with a single guide RNA. An originator guide RNA recognition site (OgRRS) comprises endogenous DNA found in untransformed plants and in endogenous non-transgenic DNA of junction polynucleotides of transgenic plants containing a modified or unmodified trans-

genic locus. The OgRRS located in non-transgenic DNA of a first DNA junction polynucleotide is used to design a related cognate guide RNA recognition site (CgRRS) which is introduced (e.g., by genome editing methods) into the second junction polynucleotide of the transgenic locus. A CgRRS is thus present in junction polynucleotides of modified transgenic loci provided herein and is absent from endogenous DNA found in untransformed plants and absent from endogenous non-transgenic DNA found in junction sequences of transgenic plants containing an unmodified transgenic locus. Also provided are unique transgenic locus excision sites created by excision of such modified transgenic loci, DNA molecules comprising the modified transgenic loci, unique transgenic locus excision sites and/or plants comprising the same, biological samples containing the DNA, nucleic acid markers adapted for detecting the DNA molecules, and related methods of identifying the elite crop plants comprising unique transgenic locus excision sites.

[0062] Also provided herein are methods whereby targeted genetic changes are efficiently combined with desired subsets of transgenic loci in elite progeny plant lines (e.g., elite inbreds used for hybrid seed production or for inbred varietal production). Examples of such methods include those illustrated in FIG. 2. In certain embodiments, INIR20 transgenic loci provided here are characterized by polynucleotide sequences that can facilitate as necessary the removal of the INIR20 transgenic loci from the genome. Useful applications of such INIR20 transgenic loci and related methods of making include targeted excision of a INIR20 transgenic locus or portion thereof in certain breeding lines to facilitate recovery of germplasm with subsets of transgenic traits tailored for specific geographic locations and/or grower preferences. Other useful applications of such INIR20 transgenic loci and related methods of making include removal of transgenic traits from certain breeding lines when it is desirable to replace the trait in the breeding line without disrupting other transgenic loci and/or non-

transgenic loci. In certain embodiments, soybean genomes containing INIR20 transgenic loci or portions thereof which can be selectively excised with one or more gRNA molecules and RdDe (RNA dependent DNA endonucleases) which form gRNA/target DNA complexes. Such selectively excisable INIR20 transgenic loci can comprise an originator guide RNA recognition site (OgRRS) which is identified in non-transgenic DNA, transgenic DNA, or a combination thereof in of a first junction polynucleotide of the transgenic locus and cognate guide RNA recognition site (CgRRS) which is introduced (e.g., by genome editing methods) into a second junction polynucleotide of the transgenic locus and which can hybridize to the same gRNA as the OgRRS, thereby permitting excision of the modified transgenic locus or portions thereof with a single guide RNA (e.g., as shown in FIGS. 3A and B). In certain embodiments, an originator guide RNA recognition site (OgRRS) comprises endogenous DNA found in untransformed plants and in endogenous non-transgenic DNA of junction polynucleotides of transgenic plants containing a modified or unmodified transgenic locus. In certain embodiments, an originator guide RNA recognition site (OgRRS) comprises exogenous transgenic DNA of junction polynucleotides of transgenic plants containing a modified or unmodified transgenic locus. The OgRRS located in non-transgenic DNA transgenic DNA, or a combination thereof in of a first DNA junction polynucleotide is used to design a related cognate guide RNA recognition site (CgRRS) which is introduced (e.g., by genome editing methods) into the second junction polynucleotide of the transgenic locus. A CgRRS is thus present in junction polynucleotides of modified transgenic loci provided herein and is absent from endogenous DNA found in untransformed plants and absent from junction sequences of transgenic plants containing an unmodified transgenic locus. A CgRRS is also absent from a combination of non-transgenic and transgenic DNA found in junction sequences of transgenic plants containing an unmodified transgenic locus. An example of OgRRS polynucleotide sequences in or near a 3' junction polynucleotide in an MON87701 transgenic locus include SEQ ID NO: 6. OgRRS polynucleotide sequences located in a first junction polynucleotide can be introduced into the second junction polynucleotide using donor DNA templates as elsewhere described herein. Donor DNA templates for introducing the SEQ ID NO: 6 OgRRS into the 5' junction polynucleotide of an MON87701 locus includes the donor DNA templates comprising SEQ ID NO: 9 and 21. Double stranded breaks in a 5' junction polynucleotide of SEQ ID NO: 1 can be introduced with the Guide-1 or Guide-2 gRNAs, which are respectively encoded by SEQ ID NO: 4 and 5, and a Cas12a nuclease. In certain embodiments, double stranded breaks in a 5' junction polynucleotide of SEQ ID NO: 1 can be introduced with the Guide-1 or 2 gRNAs and a Cas12a nuclease (e.g., a Cas12a nuclease of SEQ ID NO: 16). Integration of the SEQ ID NO: 9 or 21 donor DNA template comprising the CgRRS into the 5' junction polynucleotide of an MON87701 locus at the double stranded breaks introduced by the gRNAs encoded by SEQ ID NO: 4 or 5 and a Cas12a nuclease can provide an INIR20 locus comprising the CgRRS sequence set forth in SEQ ID NO: 3 or 15. Subsequences comprising the CORRS which is located in the 5" junction polynucleotide of the INIR20 transgenic locus are set forth in SEQ ID NO: 7, 8, and 14. Double stranded breaks in a 5' junction polynucleotide of SEQ ID NO: 1 can be introduced with

gRNAs encoded by SEQ ID NO: 4 or 5 and a Cas12a nuclease. A donor DNA template of SEQ ID NO: 9 or the equivalent thereof having longer or shorter homology arms can be used to obtain the CORRS insertion in the 5' junction polynucleotide that is set forth in SEQ ID NO: 7. An INIR20 transgenic locus containing the CgRRS insertion of SEQ ID NO: 7 is set forth in SEQ ID NO: 15. A donor DNA template of SEQ ID NO: 21 or the equivalent thereof having longer or shorter homology arms can be used to obtain the CgRRS insertion in the 5' junction polynucleotide of MON87701 that is set forth in SEQ ID NO: 8. An INIR20 transgenic locus containing the CgRRS insertion of SEQ ID NO: 8 is set forth in SEQ ID NO: 3.

**[0063]** Also provided herein are allelic variants of any of the INIR20 transgenic loci or DNA molecules provided herein. In certain embodiments, such allelic variants of INIR20 transgenic loci include sequences having at least 85%, 90%, 95%, 98%, or 99% sequence identity across the entire length or at least 20, 40, 100, 500, 1,000, 2,000, 4,000, 6,000, 8,000, 9,000, 10,000, 11,000, 12,000, 13,000, 14,000, or 14,416 nucleotides of SEQ ID NO: 2, 3, or 15. In certain embodiments, such allelic variants of INIR20 DNA molecules include sequences having at least 85%, 90%, 95%, 98%, or 99% sequence identity across the entire length of SEQ ID NO: 2, 3, 7, 8, 9, 14, 15, or 21.

**[0064]** Also provided are unique transgenic locus excision sites created by excision of INIR20 transgenic loci or selectively excisable INIR20 transgenic loci, DNA molecules comprising the INIR20 transgenic loci or unique fragments thereof (i.e., fragments of an INIR20 locus which are not found in an MON87701 transgenic locus), INIR20 plants comprising the same, biological samples containing the DNA, nucleic acid markers adapted for detecting the DNA molecules, and related methods of identifying soybean plants comprising unique INIR20 transgenic locus excision sites and unique fragments of a INIR20 transgenic locus. An example of such an excision site would include an excision site created by excising the INIR20 transgenic locus with a guide RNA encoded by SEQ ID NO: 4 or 5 and a suitable Cas RdDe (e.g., a Cas12a nuclease of SEQ ID NO: 16). DNA molecules comprising unique fragments of an INIR20 transgenic locus are diagnostic for the presence of an INIR20 transgenic locus or fragments thereof in a soybean plant, soybean cell, soybean seed, products obtained therefrom (e.g., seed meal or stover), and biological samples. DNA molecules comprising unique fragments of an INIR20 transgenic locus include DNA molecules comprising the CgRRS include SEQ ID NO: 7, 8, and 14.

**[0065]** Methods provided herein can be used to excise any transgenic locus where the first and second junction sequences comprising the endogenous non-transgenic genomic DNA and the heterologous transgenic DNA which are joined at the site of transgene insertion in the plant genome are known or have been determined. In certain embodiments provided herein, transgenic loci can be removed from crop plant lines to obtain crop plant lines with tailored combinations of transgenic loci and optionally targeted genetic changes. Such first and second junction sequences are readily identified in new transgenic events by inverse PCR techniques using primers which are complementary the inserted transgenic sequences. In certain embodiments, the first and second junction sequences of transgenic loci are published. An example of a transgenic locus which can be improved and used in the methods

provided herein is the soybean MON87701 transgenic locus. The soybean MON87701 transgenic locus is depicted in FIG. 1. Soybean plants comprising the MON87701 transgenic locus and seed thereof have been cultivated, been placed in commerce, and have been described in a variety of publications by various governmental bodies. Databases which have compiled descriptions of the MON87701 transgenic locus include the International Service for the Acquisition of Agri-biotech Applications (ISAAA) database (available on the world wide web internet site “[isaaa.org/gmapprovaldatabase/event](http://isaaa.org/gmapprovaldatabase/event)”), the GenBit LLC database (available on the world wide web internet site “[genbitgroup.com/en/gmo/gmodatabase](http://genbitgroup.com/en/gmo/gmodatabase)”), and the Biosafety Clearing-House (BCH) database (available on the <http://bch.cbd.int/database/organisms>”).

**[0066]** Sequences of the junction polynucleotides as well as the transgenic insert(s) of the MON87701 transgenic locus which can be improved by the methods provided herein are set forth or otherwise provided in SEQ ID NO: 1, U.S. Pat. No. 8,049,071, the sequence of the MON87701 locus in the deposited seed of ATCC accession No. PTA-8194, and elsewhere in this disclosure. In certain embodiments provided herein, the MON87701 transgenic locus set forth in SEQ ID NO: 1 or present in the deposited seed of ATCC accession No. PTA-8194 is referred to as an “original MON87701 transgenic locus.” Allelic or other variants of the sequence set forth SEQ ID NO: 1, the patent references set forth therein and incorporated herein by reference in their entireties, and elsewhere in this disclosure which may be present in certain variant MON87701 transgenic plant loci (e.g., progeny of deposited seed of accession No. PTA-8194 which contain allelic variants of SEQ ID NO: 1 or progeny originating from transgenic plant cells comprising the original MON87701 transgenic locus set forth in U.S. Pat. No. 8,049,071) can also be improved by identifying sequences in the variants that correspond to the SEQ ID NO: 1 by performing a pairwise alignment (e.g., using CLUSTAL O 1.2.4 with default parameters) and making corresponding changes in the allelic or other variant sequences. Such allelic or other variant sequences include sequences having at least 85%, 90%, 95%, 98%, or 99% sequence identity across the entire length or at least 20, 40, 100, 500, 1,000, 2,000, 4,000, 8,000, 10,000, 11,000, 12,000, 13,000 or 13659 nucleotides of SEQ ID NO: 1. Also provided are plants, plant parts including seeds, genomic DNA, and/or DNA obtained from INIR20 plants which comprise one or more modifications (e.g., via insertion of a CgRRS in a junction polynucleotide sequence) which provide for selective excision of the INIR20 transgenic locus or a portion thereof. Also provided herein are methods of detecting plants, genomic DNA, and/or DNA obtained from plants comprising a INIR20 transgenic locus which contains one or more of a CgRRS, deletions of selectable marker genes, deletions of non-essential DNA, and/or a transgenic locus excision site. A first junction polynucleotide of a MON87701 transgenic locus can comprise either one of the junction polynucleotides found at the 5' end or the 3' end of any one of the sequences set forth in SEQ ID NO: 1, allelic variants thereof, or other variants thereof. An OgRRS can be found within non-transgenic DNA, transgenic DNA, or a combination thereof in either one of the junction polynucleotides of any one of SEQ ID NO: 1, allelic variants thereof, or other variants thereof. A second junction polynucleotide of a transgenic locus can comprise either one of the junction

polynucleotides found at the 5' or 3' end of any one of the sequences set forth in SEQ ID NO: 1, allelic variants thereof, or other variants thereof. A CgRRS can be introduced within transgenic, non-transgenic DNA, or a combination thereof of either one of the junction polynucleotides of any one of SEQ ID NO: 1, allelic variants thereof, or other variants thereof to obtain an INIR20 transgenic locus. In certain embodiments, the OgRRS is found in non-transgenic DNA or transgenic DNA of the 5' junction polynucleotide of a transgenic locus of any one of SEQ ID NO: 1, allelic variants thereof, or other variants thereof and the corresponding CgRRS is introduced into the transgenic DNA, non-transgenic DNA, or a combination thereof in the 3' junction polynucleotide of the MON87701 transgenic locus of SEQ ID NO: 1, allelic variants thereof, or other variants thereof to obtain an INIR20 transgenic locus. In other embodiments, the OgRRS is found in non-transgenic DNA or transgenic DNA of the 3' junction polynucleotide of the MON87701 transgenic locus of any one of SEQ ID NO: 1, allelic variants thereof, or other variants thereof and the corresponding CgRRS is introduced into the transgenic DNA, non-transgenic DNA, or a combination thereof in the 5' junction polynucleotide of the transgenic locus of SEQ ID NO: 1, allelic variants thereof, or other variants thereof to obtain an INIR20 transgenic locus.

**[0067]** In certain embodiments, the CgRRS is comprised in whole or in part of an exogenous DNA molecule that is introduced into a DNA junction polynucleotide by genome editing. In certain embodiments, the guide RNA hybridization site of the CORRS is operably linked to a pre-existing PAM site in the transgenic DNA or non-transgenic DNA of the transgenic plant genome. In other embodiments, the guide RNA hybridization site of the CgRRS is operably linked to a new PAM site that is introduced in the DNA junction polynucleotide by genome editing. A CgRRS can be located in non-transgenic plant genomic DNA of a DNA junction polynucleotide of an INIR20 transgenic locus, in transgenic DNA of a DNA junction polynucleotide of an INIR20 transgenic locus or can span the junction of the transgenic and non-transgenic DNA of a DNA junction polynucleotide of an INIR20 transgenic locus. An OgRRS can likewise be located in non-transgenic plant genomic DNA of a DNA junction polynucleotide of an INIR20 transgenic locus, in transgenic DNA of a DNA junction polynucleotide of an INIR20 transgenic locus, or can span the junction of the transgenic and non-transgenic DNA of a DNA junction polynucleotide of an INIR20 transgenic locus.

**[0068]** Methods provided herein can be used in a variety of breeding schemes to obtain elite crop plants comprising subsets of desired modified transgenic loci comprising an OgRRS and a CgRRS operably linked to junction polynucleotide sequences and transgenic loci excision sites where undesired transgenic loci or portions thereof have been removed (e.g., by use of the OgRRS and a CgRRS). Such methods are useful at least insofar as they allow for production of distinct useful donor plant lines each having unique sets of modified transgenic loci and, in some instances, targeted genetic changes that are tailored for distinct geographies and/or product offerings. In an illustrative and non-limiting example, a different product lines comprising transgenic loci conferring only two of three types of herbicide tolerance (e.g., glyphosate, glufosinate, and dicamba) can be obtained from a single donor line comprising three distinct transgenic loci conferring resis-

tance to all three herbicides. In certain aspects, plants comprising the subsets of undesired transgenic loci and transgenic loci excision sites can further comprise targeted genetic changes. Such elite crop plants can be inbred plant lines or can be hybrid plant lines. In certain embodiments, at least two transgenic loci (e.g., transgenic loci including an INIR20 and another modified transgenic locus wherein an OgRRS and a CgRRS site is operably linked to a first and a second junction sequence and optionally a selectable marker gene and/or non-essential DNA are deleted) are introgressed into a desired donor line comprising elite crop plant germplasm and then subjected to genome editing molecules to recover plants comprising one of the two introgressed transgenic loci as well as a transgenic loci excision site introduced by excision of the other transgenic locus or portion thereof by the genome editing molecules. In certain embodiments, the genome editing molecules can be used to remove a transgenic locus and introduce targeted genetic changes in the crop plant genome. Introgression can be achieved by backcrossing plants comprising the transgenic loci to a recurrent parent comprising the desired elite germplasm and selecting progeny with the transgenic loci and recurrent parent germplasm. Such backcrosses can be repeated and/or supplemented by molecular assisted breeding techniques using SNP or other nucleic acid markers to select for recurrent parent germplasm until a desired recurrent parent percentage is obtained (e.g., at least about 95%, 96%, 97%, 98%, or 99% recurrent parent percentage). A non-limiting, illustrative depiction of a scheme for obtaining plants with both subsets of transgenic loci and the targeted genetic changes is shown in the FIG. 2 (bottom "Alternative" panel), where two or more of the transgenic loci ("Event" in FIG. 2) are provided in Line A and then moved into elite crop plant germplasm by introgression. In the non-limiting FIG. 2 illustration, introgression can be achieved by crossing a "Line A" comprising two or more of the modified transgenic loci to the elite germplasm and then backcrossing progeny of the cross comprising the transgenic loci to the elite germplasm as the recurrent parent) to obtain a "Universal Donor" (e.g., Line A+ in FIG. 2) comprising two or more of the modified transgenic loci. This elite germplasm containing the modified transgenic loci (e.g., "Universal Donor" of FIG. 2) can then be subjected to genome editing molecules which can excise at least one of the transgenic loci ("Event Removal" in FIG. 2) and introduce other targeted genetic changes ("GE" in FIG. 2) in the genomes of the elite crop plants containing one of the transgenic loci and a transgenic locus excision site corresponding to the removal site of one of the transgenic loci. Such selective excision of transgenic loci or portion thereof can be effected by contacting the genome of the plant comprising two transgenic loci with gene editing molecules (e.g., RdDe and gRNAs, TALENS, and/or ZFN) which recognize one transgenic loci but not another transgenic loci. Genome editing molecules that provide for selective excision of a first modified transgenic locus comprising an OgRRS and a CgRRS include a gRNA that hybridizes to the OgRRS and CgRRS of the first modified transgenic locus and an RdDe that recognizes the gRNA/OgRRS and gRNA/CgRRS complexes. Distinct plant lines with different subsets of transgenic loci and desired targeted genetic changes are thus recovered (e.g., "Line B-1," "Line B-2," and "Line B-3" in FIG. 2). In certain embodiments, it is also desirable to bulk up populations of inbred elite crop plants or their seed comprising

the subset of transgenic loci and a transgenic locus excision site by selfing. In certain embodiments, inbred progeny of the selfed soybean plants comprising the INIR20 transgenic loci can be used as a pollen donor or recipient for hybrid seed production. Such hybrid seed and the progeny grown therefrom can comprise a subset of desired transgenic loci and a transgenic loci excision site.

[0069] Hybrid plant lines comprising elite crop plant germplasm, at least one transgenic locus and at least one transgenic locus excision site, and in certain aspects, additional targeted genetic changes are also provided herein. Methods for production of such hybrid seed can comprise crossing elite crop plant lines where at least one of the pollen donor or recipient comprises at least the transgenic locus and a transgenic locus excision site and/or additional targeted genetic changes. In certain embodiments, the pollen donor and recipient will comprise germplasm of distinct heterotic groups and provide hybrid seed and plants exhibiting heterosis. In certain embodiments, the pollen donor and recipient can each comprise a distinct transgenic locus which confers either a distinct trait (e.g., herbicide tolerance or insect resistance), a different type of trait (e.g., tolerance to distinct herbicides or to distinct insects such as coleopteran insects), or a different mode-of-action for the same trait (e.g., resistance to lepidopteran insects by two distinct modes-of-action or resistance to lepidopteran insects by two distinct modes-of-action). In certain embodiments, the pollen recipient will be rendered male sterile or conditionally male sterile. Methods for inducing male sterility or conditional male sterility include emasculation (e.g., detasseling), cytoplasmic male sterility, chemical hybridizing agents or systems, a transgenes or transgene systems, and/or mutation(s) in one or more endogenous plant genes. Descriptions of various male sterility systems that can be adapted for use with the elite crop plants provided herein are described in Wan et al. Molecular Plant; 12, 3, (2019):321-342 as well as in U.S. Pat. No. 8,618,358; US20130031674; and US2003188347.

[0070] In certain embodiments, edited transgenic plant genomes, transgenic plant cells, parts, or plants containing those genomes, and DNA molecules obtained therefrom, can comprise a desired subset of transgenic loci and/or comprise at least one transgenic locus excision site. In certain embodiments, a segment comprising an INIR20 transgenic locus comprising an OgRRS in non-transgenic DNA of a 1<sup>st</sup> junction polynucleotide sequence and a CgRRS in a 2<sup>nd</sup> junction polynucleotide sequence is deleted with a gRNA and RdDe that recognize the OgRRS and the CgRRS to produce an INIR20 transgenic locus excision site. For example, an INIR20 transgenic locus set forth in SEQ ID NO: 3 or 15 can be deleted with a Cas12a RdDe (e.g. the Cas12a of SEQ ID NO:16) and a gRNA comprising an RNA encoded by SEQ ID NO: 11. In certain embodiments, the transgenic locus excision site can comprise a contiguous segment of DNA comprising at least 10 base pairs of DNA that is telomere proximal to the deleted segment of the transgenic locus and at least 10 base pairs of DNA that is centromere proximal to the deleted segment of the transgenic locus wherein the transgenic DNA (i.e., the heterologous DNA) that has been inserted into the crop plant genome has been deleted. In certain embodiments where a segment comprising a transgenic locus has been deleted, the transgenic locus excision site can comprise a contiguous segment of DNA comprising at least 10 base pairs DNA that is

telomere proximal to the deleted segment of the transgenic locus and at least 10 base pairs of DNA that is centromere proximal DNA to the deleted segment of the transgenic locus wherein the heterologous transgenic DNA and at least 1, 2, 5, 10, 20, 50, or more base pairs of endogenous DNA located in a 5' junction sequence and/or in a 3' junction sequence of the original transgenic locus that has been deleted. In such embodiments where DNA comprising the transgenic locus is deleted, a transgenic locus excision site can comprise at least 10 base pairs of DNA that is telomere proximal to the deleted segment of the transgenic locus and at least 10 base pairs of DNA that is centromere proximal to the deleted segment of the transgenic locus wherein all of the transgenic DNA is absent and either all or less than all of the endogenous DNA flanking the transgenic DNA sequences are present. In certain embodiments where a segment consisting essentially of an original transgenic locus has been deleted, the transgenic locus excision site can be a contiguous segment of at least 10 base pairs of DNA that is telomere proximal to the deleted segment of the transgenic locus and at least 10 base pairs of DNA that is centromere proximal to the deleted segment of the transgenic locus wherein less than all of the heterologous transgenic DNA that has been inserted into the crop plant genome is excised. In certain aforementioned embodiments where a segment consisting essentially of an original transgenic locus has been deleted, the transgenic locus excision site can thus contain at least 1 base pair of DNA or 1 to about 2 or 5, 8, 10, 20, or 50 base pairs of DNA comprising the telomere proximal and/or centromere proximal heterologous transgenic DNA that has been inserted into the crop plant genome. In certain embodiments where a segment consisting of an original transgenic locus has been deleted, the transgenic locus excision site can contain a contiguous segment of DNA comprising at least 10 base pairs of DNA that is telomere proximal to the deleted segment of the transgenic locus and at least 10 base pairs of DNA that is centromere proximal to the deleted segment of the transgenic locus wherein the heterologous transgenic DNA that has been inserted into the crop plant genome is deleted. In certain embodiments where DNA consisting of the transgenic locus is deleted, a transgenic locus excision site can comprise at least 10 base pairs of DNA that is telomere proximal to the deleted segment of the transgenic locus and at least 10 base pairs of DNA that is centromere proximal to the deleted segment of the transgenic locus wherein all of the heterologous transgenic DNA that has been inserted into the crop plant genome is deleted and all of the endogenous DNA flanking the heterologous sequences of the transgenic locus is present. In any of the aforementioned embodiments or in other embodiments, the continuous segment of DNA comprising the transgenic locus excision site can further comprise an insertion of 1 to about 2, 5, 10, 20, or more nucleotides between the DNA that is telomere proximal to the deleted segment of the transgenic locus and the DNA that is centromere proximal to the deleted segment of the transgenic locus. Such insertions can result either from endogenous DNA repair and/or recombination activities at the double stranded breaks introduced at the excision site and/or from deliberate insertion of an oligonucleotide. Plants, edited plant genomes, biological samples, and DNA molecules (e.g., including isolated or purified DNA molecules) comprising the INIR20 transgenic loci excision sites are provided herein.

[0071] In other embodiments, a segment comprising a INIR20 transgenic locus (e.g., a transgenic locus comprising an OgRRS in non-transgenic DNA of a 1<sup>st</sup> junction sequence and a CgRRS in a 2<sup>nd</sup> junction sequence) can be deleted with a gRNA and RdDe that recognize the OgRRS and the CORRS (e.g., the Cas12a RdDe of SEQ ID NO: 16 and a gRNA comprising an RNA encoded by SEQ ID NO: 11) and replaced with DNA comprising the endogenous non-transgenic plant genomic DNA present in the genome prior to transgene insertion. A non-limiting example of such replacements can be visualized in FIG. 3C, where the donor DNA template can comprise the endogenous non-transgenic plant genomic DNA present in the genome prior to transgene insertion along with sufficient homology to non-transgenic DNA on each side of the excision site to permit homology-directed repair. In certain embodiments, the endogenous non-transgenic plant genomic DNA present in the genome prior to transgene insertion can be at least partially restored. In certain embodiments, the endogenous non-transgenic plant genomic DNA present in the genome prior to transgene insertion can be essentially restored such that no more than about 5, 10, or 20 to about 50, 80, or 100 nucleotides are changed relative to the endogenous DNA at the essentially restored excision site.

[0072] In certain embodiments, edited transgenic plant genomes and transgenic plant cells, plant parts, or plants containing those edited genomes, comprising a modification of an original transgenic locus, where the modification comprises an OgRRS and a CgRRS which are operably linked to a 1<sup>st</sup> and a 2<sup>nd</sup> junction sequence, respectively or irrespectively, and optionally further comprise a deletion of a segment of the original transgenic locus. In certain embodiments, the modification comprises two or more separate deletions and/or there is a modification in two or more original transgenic plant loci. In certain embodiments, the deleted segment comprises, consists essentially of, or consists of a segment of non-essential DNA in the transgenic locus. Illustrative examples of non-essential DNA include but are not limited to synthetic cloning site sequences, duplications of transgene sequences; fragments of transgene sequences, and *Agrobacterium* right and/or left border sequences. In certain embodiments, the non-essential DNA is a duplication and/or fragment of a promoter sequence and/or is not the promoter sequence operably linked in the cassette to drive expression of a transgene. In certain embodiments, excision of the non-essential DNA improves a characteristic, functionality, and/or expression of a transgene of the transgenic locus or otherwise confers a recognized improvement in a transgenic plant comprising the edited transgenic plant genome. In certain embodiments, the non-essential DNA does not comprise DNA encoding a selectable marker gene. In certain embodiments of an edited transgenic plant genome, the modification comprises a deletion of the non-essential DNA and a deletion of a selectable marker gene. The modification producing the edited transgenic plant genome could occur by excising both the non-essential DNA and the selectable marker gene at the same time, e.g., in the same modification step, or the modification could occur step-wise. For example, an edited transgenic plant genome in which a selectable marker gene has previously been removed from the transgenic locus can comprise an original transgenic locus from which a non-essential DNA is further excised and vice versa. In certain embodiments, the modification comprising deletion of the non-

essential DNA and deletion of the selectable marker gene comprises excising a single segment of the original transgenic locus that comprises both the non-essential DNA and the selectable marker gene. Such modification would result in one excision site in the edited transgenic genome corresponding to the deletion of both the non-essential DNA and the selectable marker gene. In certain embodiments, the modification comprising deletion of the non-essential DNA and deletion of the selectable marker gene comprises excising two or more segments of the original transgenic locus to achieve deletion of both the non-essential DNA and the selectable marker gene. Such modification would result in at least two excision sites in the edited transgenic genome corresponding to the deletion of both the non-essential DNA and the selectable marker gene. In certain embodiments of an edited transgenic plant genome, prior to excision, the segment to be deleted is flanked by operably linked protospacer adjacent motif (PAM) sites in the original or unmodified transgenic locus and/or the segment to be deleted encompasses an operably linked PAM site in the original or unmodified transgenic locus. In certain embodiments, following excision of the segment, the resulting edited transgenic plant genome comprises PAM sites flanking the deletion site in the modified transgenic locus. In certain embodiments of an edited transgenic plant genome, the modification comprises a modification of a MON87701 transgenic locus.

**[0073]** In certain embodiments, improvements in a transgenic plant locus are obtained by introducing a new cognate guide RNA recognition site (CgRRS) which is operably linked to a DNA junction polynucleotide of the transgenic locus in the transgenic plant genome. Such CgRRS sites can be recognized by RdDe, and a single suitable guide RNA directed to the CgRRS and the originator gRNA Recognition Site (OgRRS) to provide for cleavage within the junction polynucleotides which flank an INIR20 transgenic locus. In certain embodiments, the CgRRS/gRNA and OgRRS/gRNA hybridization complexes are recognized by the same class of RdDe (e.g., Class 2 type II or Class 2 type V) or by the same RdDe (e.g., both the CgRRS/gRNA and OgRRS/gRNA hybridization complexes recognized by the same Cas9 or Cas12 RdDe). Such CgRRS and OgRRS can be recognized by RdDe and suitable guide RNAs containing crRNA sufficiently complementary to the guide RNA hybridization site DNA sequences adjacent to the PAM site of the CgRRS and the OgRRS to provide for cleavage within or near the two junction polynucleotides. Suitable guide RNAs can be in the form of a single gRNA comprising a crRNA or in the form of a crRNA/tracrRNA complex. In the case of the OgRRS site, the PAM and guide RNA hybridization site are endogenous DNA polynucleotide molecules found in the plant genome. In certain embodiments where the CgRRS is introduced into the plant genome by genome editing, gRNA hybridization site polynucleotides introduced at the CORRS are at least 17 or 18 nucleotides in length and are complementary to the crRNA of a guide RNA. In certain embodiments, the gRNA hybridization site sequence of the OgRRS and/or the CgRRS is about 17 or 18 to about 24 nucleotides in length. The gRNA hybridization site sequence of the OgRRS and the gRNA hybridization site of the CgRRS can be of different lengths or comprise different sequences so long as there is sufficient complementarity to permit hybridization by a single gRNA and recognition by a RdDe that recognizes and cleaves DNA at the gRNA/OgRRS and

gRNA/CgRRS complex. In certain embodiments, the guide RNA hybridization site of the CgRRS comprise about a 17 or 18 to about 24 nucleotide sequence which is identical to the guide RNA hybridization site of the OgRRS. In other embodiments, the guide RNA hybridization site of the CgRRS comprise about a 17 or 18 to about 24 nucleotide sequence which has one, two, three, four, or five nucleotide insertions, deletions or substitutions when compared to the guide RNA hybridization site of the OgRRS. Certain CgRRS comprising a gRNA hybridization site containing has one, two, three, four, or five nucleotide insertions, deletions or substitutions when compared to the guide RNA hybridization site of the OgRRS can undergo hybridization with a gRNA which is complementary to the OgRRS gRNA hybridization site and be cleaved by certain RdDe. Examples of mismatches between gRNAs and guide RNA hybridization sites which allow for RdDe recognition and cleavage include mismatches resulting from both nucleotide insertions and deletions in the DNA which is hybridized to the gRNA (e.g., Lin et al., doi: 10.1093/nar/gku402). In certain embodiments, an operably linked PAM site is co-introduced with the gRNA hybridization site polynucleotide at the CgRRS. In certain embodiments, the gRNA hybridization site polynucleotides are introduced at a position adjacent to a resident endogenous PAM sequence in the junction polynucleotide sequence to form a CgRRS where the gRNA hybridization site polynucleotides are operably linked to the endogenous PAM site. In certain embodiments, non-limiting features of the OgRRS, CgRRS, and/or the gRNA hybridization site polynucleotides thereof include: (i) absence of significant homology or sequence identity (e.g., less than 50% sequence identity across the entire length of the OgRRS, CgRRS, and/or the gRNA hybridization site sequence) to any other endogenous or transgenic sequences present in the transgenic plant genome or in other transgenic genomes of the soybean plant being transformed and edited; (ii) absence of significant homology or sequence identity (e.g., less than 50% sequence identity across the entire length of the sequence) of a sequence of a first OgRRS and a first CgRRS to a second OgRRS and a second CgRRS which are operably linked to junction polynucleotides of a distinct transgenic locus; (iii) the presence of some sequence identity (e.g., about 25%, 40%, or 50% to about 60%, 70%, or 80%) between the OgRRS sequence and endogenous sequences present at the site where the CORRS sequence is introduced; and/or (iv) optimization of the gRNA hybridization site polynucleotides for recognition by the RdDe and guide RNA when used in conjunction with a particular PAM sequence. In certain embodiments, the first and second OgRRS as well as the first and second CgRRS are recognized by the same class of RdDe (e.g., Class 2 type II or Class 2 type V) or by the same RdDe (e.g., Cas9 or Cas 12 RdDe). In certain embodiments, the first OgRRS site in a first junction polynucleotide and the CgRRS introduced in the second junction polynucleotide to permit excision of a first transgenic locus by a first single guide RNA and a single RdDe. Such nucleotide insertions or genome edits used to introduce CgRRS in a transgenic plant genome can be effected in the plant genome by using gene editing molecules (e.g., RdDe and guide RNAs, RNA dependent nickases and guide RNAs, Zinc Finger nucleases or nickases, or TALE nucleases or nickases) which introduce blunt double stranded breaks or staggered double stranded breaks in the DNA junction polynucleotides. In the case of DNA inser-

tions, the genome editing molecules can also in certain embodiments further comprise a donor DNA template or other DNA template which comprises the heterologous nucleotides for insertion to form the CgRRS. Guide RNAs can be directed to the junction polynucleotides by using a pre-existing PAM site located within or adjacent to a junction polynucleotide of the transgenic locus. Non-limiting examples of such pre-existing PAM sites present in junction polynucleotides, which can be used either in conjunction with an inserted heterologous sequence to form a CgRRS or which can be used to create a double stranded break to insert or create a CgRRS, include PAM sites recognized by a Cas12a enzyme. Non-limiting examples where a CgRRS is created in a DNA sequence are illustrated in Example 2.

[0074] Transgenic loci comprising OgRRS and CgRRS in a first and a second junction polynucleotides can be excised from the genomes of transgenic plants by contacting the transgenic loci with RdDe or RNA directed nickases, and a suitable guide RNA directed to the OgRRS and CgRRS (e.g., the Cas12a RdDe of SEQ ID NO: 16 and a gRNA comprising an RNA encoded by SEQ ID NO: 11). A non-limiting example where a modified transgenic locus is excised from a plant genome by use of a gRNA and an RdDe that recognizes an OgRRS/gRNA and a CgRRS/gRNA complex and introduces dsDNA breaks in both junction polynucleotides and repaired by NHEJ is depicted in FIG. 3B. In the depicted example set forth in FIG. 3B, the OgRRS site and the CORRS site are absent from the plant chromosome comprising the transgene excision site that results from the process. In other embodiments provided herein where a modified transgenic locus is excised from a plant genome by use of a gRNA and an RdDe that recognizes an OgRRS/gRNA and a CgRRS/gRNA complex and repaired by NHEJ or microhomology-mediated end joining (MMEJ), the OgRRS and/or other non-transgenic sequences that were originally present prior to transgene insertion are partially or essentially restored.

[0075] In certain embodiments, edited transgenic plant genomes provided herein can comprise additional new introduced transgenes (e.g., expression cassettes) inserted into the transgenic locus of a given event. Introduced transgenes inserted at the transgenic locus of an event subsequent to the event's original isolation can be obtained by inducing a double stranded break at a site within an original transgenic locus (e.g., with genome editing molecules including an RdDe and suitable guide RNA(s); a suitable engineered zinc-finger nuclease; a TALEN protein and the like) and providing an exogenous transgene in a donor DNA template which can be integrated at the site of the double stranded break (e.g. by homology-directed repair (HDR) or by non-homologous end-joining (NHEJ)). In certain embodiments, an OgRRS and a CgRRS located in a 1<sup>st</sup> junction polynucleotide and a 2<sup>nd</sup> junction polynucleotide, respectively, can be used to delete the transgenic locus and replace it with one or more new expression cassettes. In certain embodiments, such deletions and replacements are effected by introducing dsDNA breaks in both junction polynucleotides and providing the new expression cassettes on a donor DNA template (e.g., in FIG. 3C, the donor DNA template can comprise an expression cassette flanked by DNA homologous to non-transgenic DNA located telomere proximal and centromere proximal to the excision site). Suitable expression cassettes for insertion include DNA molecules comprising promoters which are operably linked to DNA encoding proteins and/or

RNA molecules which confer useful traits which are in turn operably linked to polyadenylation sites or terminator elements. In certain embodiments, such expression cassettes can also comprise 5' UTRs, 3' UTRs, and/or introns. Useful traits include biotic stress tolerance (e.g., insect resistance, nematode resistance, or disease resistance), abiotic stress tolerance (e.g., heat, cold, drought, and/or salt tolerance), herbicide tolerance, and quality traits (e.g., improved fatty acid compositions, protein content, starch content, and the like). Suitable expression cassettes for insertion include expression cassettes which confer insect resistance, herbicide tolerance, biofuel use, or male sterility traits contained in any of the transgenic events set forth in US Patent Application Public. Nos. 20090038026, 20130031674, 20150361446, 20170088904, 20150267221, 201662346688, and 20200190533 as well as in U.S. Pat. Nos. 6,342,660, 7,323,556, 6,040,497, 8,759,618, 7,157, 281, 6,852,915, 7,705,216, 10,316,330, 8,618,358, 8,450, 561, 8,212,113, 9,428,765, 9,540,655, 7,897,748, 8,273,959, 8,093,453, 8,901,378, 9,994,863, 7,928,296, and 8,466,346, each of which are incorporated herein by reference in their entireties.

[0076] In certain embodiments, INIR20 plants provided herein, including plants with one or more transgenic loci, modified transgenic loci, and/or comprising transgenic loci excision sites can further comprise one or more targeted genetic changes introduced by one or more of gene editing molecules or systems. Also provided are methods where the targeted genetic changes are introduced and one or more transgenic loci are removed from plants either in series or in parallel (e.g., as set forth in the non-limiting illustration in FIG. 2, bottom "Alternative" panel, where "GE" can represent targeted genetic changes induced by gene editing molecules and "Event Removal" represents excision of one or more transgenic loci with gene editing molecules). Such targeted genetic changes include those conferring traits such as improved yield, improved food and/or feed characteristics (e.g., improved oil, starch, protein, or amino acid quality or quantity), improved nitrogen use efficiency, improved biofuel use characteristics (e.g., improved ethanol production), male sterility/conditional male sterility systems (e.g., by targeting endogenous MS26, MS45 and MSCA1 genes), herbicide tolerance (e.g., by targeting endogenous ALS, EPSPS, HPPD, or other herbicide target genes), delayed flowering, non-flowering, increased biotic stress resistance (e.g., resistance to insect, nematode, bacterial, or fungal damage), increased abiotic stress resistance (e.g., resistance to drought, cold, heat, metal, or salt), enhanced lodging resistance, enhanced growth rate, enhanced biomass, enhanced tillering, enhanced branching, delayed flowering time, delayed senescence, increased flower number, improved architecture for high density planting, improved photosynthesis, increased root mass, increased cell number, improved seedling vigor, improved seedling size, increased rate of cell division, improved metabolic efficiency, and increased meristem size in comparison to a control plant lacking the targeted genetic change. Types of targeted genetic changes that can be introduced include insertions, deletions, and substitutions of one or more nucleotides in the crop plant genome. Sites in endogenous plant genes for the targeted genetic changes include promoter, coding, and non-coding regions (e.g., 5' UTRs, introns, splice donor and acceptor sites and 3' UTRs). In certain embodiments, the targeted genetic change comprises an insertion of a regula-

tory or other DNA sequence in an endogenous plant gene. Non-limiting examples of regulatory sequences which can be inserted into endogenous plant genes with gene editing molecules to effect targeted genetic changes which confer useful phenotypes include those set forth in US Patent Application Publication 20190352655, which is incorporated herein by reference in its entirety, such as: (a) auxin response element (AuxRE) sequence; (b) at least one D1-4 sequence (Ulmakov et al. (1997) *Plant Cell*, 9:1963-1971), (c) at least one DR5 sequence (Ulmakov et al. (1997) *Plant Cell*, 9:1963-1971); (d) at least one m5-DR5 sequence (Ulmakov et al. (1997) *Plant Cell*, 9:1963-1971); (e) at least one P3 sequence; (f) a small RNA recognition site sequence bound by a corresponding small RNA (e.g., an siRNA, a microRNA (miRNA), a trans-acting siRNA as described in U.S. Pat. No. 8,030,473, or a phased sRNA as described in U.S. Pat. No. 8,404,928; both of these cited patents are incorporated by reference herein); (g) a microRNA (miRNA) recognition site sequence; (h) the sequence recognizable by a specific binding agent includes a microRNA (miRNA) recognition sequence for an engineered miRNA wherein the specific binding agent is the corresponding engineered mature miRNA; (i) a transposon recognition sequence; (j) a sequence recognized by an ethylene-responsive element binding-factor-associated amphiphilic repression (EAR) motif; (k) a splice site sequence (e.g., a donor site, a branching site, or an acceptor site; see, for example, the splice sites and splicing signals set forth in the internet site [lemur\[dot\]amul\[dot\]edu\[dot\]pl/share/ERISdb/home.html](http://lemur[dot]amul[dot]edu[dot]pl/share/ERISdb/home.html)); (l) a recombinase recognition site sequence that is recognized by a site-specific recombinase; (m) a sequence encoding an RNA or amino acid aptamer or an RNA riboswitch, the specific binding agent is the corresponding ligand, and the change in expression is upregulation or downregulation; (n) a hormone responsive element recognized by a nuclear receptor or a hormone-binding domain thereof; (o) a transcription factor binding sequence; and (p) a polycomb response element (see Xiao et al. (2017) *Nature Genetics*, 49:1546-1552, doi: 10.1038/ng.3937). Non limiting examples of target soybean genes that can be subjected to targeted gene edits to confer useful traits include: (a) ZmIPK1 (herbicide tolerant and phytate reduced soybean; Shukla et al., *Nature*. 2009; 459:437-41); (b) ZmGL2 (reduced epicuticular wax in leaves; Char et al. *Plant Biotechnol J.* 2015; 13:1002); (c) ZmMTL (induction of haploid plants; Kelliher et al. *Nature*. 2017; 542:105); (d) Wx1 (high amylopectin content; US 20190032070; incorporated herein by reference in its entirety); (e) TMS5 (thermosensitive male sterile; Li et al. *J Genet Genomics*. 2017; 44:465-8); (f) ALS (herbicide tolerance; Svitashov et al.; *Plant Physiol*. 2015; 169:931-45); and (g) ARGOS8 (drought stress tolerance; Shi et al., *Plant Biotechnol J.* 2017; 15:207-16). Non-limiting examples of target genes in crop plants including soybean which can be subjected to targeted genetic changes which confer useful phenotypes include those set forth in U.S. patents application Nos. 20190352655, 20200199609, 20200157554, and 20200231982, which are each incorporated herein in their entireties; and Zhang et al. (*Genome Biol*. 2018; 19:210).

**[0077]** In certain embodiments, it will be desirable to use genome editing molecules to make modified transgenic loci by introducing a CgRRS into the transgenic loci, to excise modified transgenic loci comprising an OgRRS and a CgRRS, and/or to make targeted genetic changes in elite

crop plant or other germplasm. In certain embodiments, the genome edits can be effected in regenerable plant parts (e.g., plant embryos) of elite crop plants by transient provision of gene editing molecules or polynucleotides encoding the same and do not necessarily require incorporating a selectable marker gene into the plant genome (e.g., US 20160208271 and US20180273960, both incorporated herein by reference in their entireties; Svitashov et al. *Nat Commun*. 2016; 7:13274).

**[0078]** Gene editing molecules of use in methods provided herein include molecules capable of introducing a double-strand break ("DSB") or single-strand break ("SSB") in double-stranded DNA, such as in genomic DNA or in a target gene located within the genomic DNA as well as accompanying guide RNA or donor DNA template polynucleotides. Examples of such gene editing molecules include: (a) a nuclease comprising an RNA-guided nuclease, an RNA-guided DNA endonuclease or RNA directed DNA endonuclease (RdDn), a class 1 CRISPR type nuclease system, a type II Cas nuclease, a Cas9, a nCas9 nickase, a type V Cas nuclease, a Cas12a nuclease, a nCas12a nickase, a Cas12d (CasY), a Cas12e (CasX), a Cas12b (C2c1), a Cas12c (C2c3), a Cas12i, a Cas12j, a Cas14, an engineered nuclease, a codon-optimized nuclease, a zinc-finger nuclease (ZFN) or nickase, a transcription activator-like effector nuclease (TAL-effector nuclease or TALEN) or nickase (TALE-nickase), an Argonaute, and a meganuclease or engineered meganuclease; (b) a polynucleotide encoding one or more nucleases capable of effectuating site-specific alteration (including introduction of a DSB or SSB) of a target nucleotide sequence; (c) a guide RNA (gRNA) for an RNA-guided nuclease, or a DNA encoding a gRNA for an RNA-guided nuclease; (d) donor DNA template polynucleotides; and (e) other DNA templates (dsDNA, ssDNA, or combinations thereof) suitable for insertion at a break in genomic DNA (e.g., by non-homologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ)).

**[0079]** CRISPR-type genome editing can be adapted for use in the plant cells and methods provided herein in several ways. CRISPR elements, e.g., gene editing molecules comprising CRISPR endonucleases and CRISPR guide RNAs including single guide RNAs or guide RNAs in combination with tracrRNAs or scotRNA, or polynucleotides encoding the same, are useful in effectuating genome editing without remnants of the CRISPR elements or selective genetic markers occurring in progeny. In certain embodiments, the CRISPR elements are provided directly to the eukaryotic cell (e.g., plant cells), systems, methods, and compositions as isolated molecules, as isolated or semi-purified products of a cell free synthetic process (e.g., in vitro translation), or as isolated or semi-purified products of in a cell-based synthetic process (e.g., such as in a bacterial or other cell lysate). In certain embodiments, genome-inserted CRISPR elements are useful in plant lines adapted for use in the methods provide herein. In certain embodiments, plants or plant cells used in the systems, methods, and compositions provided herein can comprise a transgene that expresses a CRISPR endonuclease (e.g., a Cas9, a Cpf1-type or other CRISPR endonuclease). In certain embodiments, one or more CRISPR endonucleases with unique PAM recognition sites can be used. Guide RNAs (sgRNAs or crRNAs and a tracrRNA) used to form an RNA-guided endonuclease/guide RNA complex can specifically bind via hybridization to gRNA hybridization site sequences (i.e., protospacer

sequences) in the gDNA target site that are adjacent to a protospacer adjacent motif (PAM) sequence. The type of RNA-guided endonuclease typically informs the location of suitable PAM sites and design of crRNAs or sgRNAs. G-rich PAM sites, e.g., 5'-NGG are typically targeted for design of crRNAs or sgRNAs used with Cas9 proteins. Examples of PAM sequences include 5'-NGG (*Streptococcus pyogenes*), 5'-NNAGAA (*Streptococcus thermophilus* CRISPR1), 5'-NGGNG (*Streptococcus thermophilus* CRISPR3), 5'-NNGRRT or 5'-NNGRRR (*Staphylococcus aureus* Cas9, SaCas9), and 5'-NNNGATT (*Neisseria meningitidis*). T-rich PAM sites (e.g., 5'-TTN or 5'-TTTV, where "V" is A, C, or G) are typically targeted for design of crRNAs or sgRNAs used with Cas12a proteins (e.g., the Cas12a protein of SEQ ID NO: 16). In some instances, Cas12a can also recognize a 5'-CTA PAM motif. Other examples of potential Cas12a PAM sequences include TTN, CTN, TCN, CCN, TTIN, TCTN, TTCN, CTTN, ATTN, TCCN, TTGN, GTTN, CCCN, CCTN, TTAN, TCGN, CTCN, ACTN, GCTN, TCAN, GCCN, and CCGN (wherein N is defined as any nucleotide). Cpf1 (i.e., Cas12a) endonuclease and corresponding guide RNAs and PAM sites are disclosed in US Patent Application Publication 2016/0208243 A1, which is incorporated herein by reference for its disclosure of DNA encoding Cpf1 endonucleases and guide RNAs and PAM sites. Introduction of one or more of a wide variety of CRISPR guide RNAs that interact with CRISPR endonucleases integrated into a plant genome or otherwise provided to a plant is useful for genetic editing for providing desired phenotypes or traits, for trait screening, or for gene editing mediated trait introgression (e.g., for introducing a trait into a new genotype without backcrossing to a recurrent parent or with limited backcrossing to a recurrent parent). Multiple endonucleases can be provided in expression cassettes with the appropriate promoters to allow multiple genome site editing.

[0080] CRISPR technology for editing the genes of eukaryotes is disclosed in US Patent Application Publications 2016/0138008A1 and US2015/0344912A1, and in U.S. Pat. Nos. 8,697,359, 8,771,945, 8,945,839, 8,999,641, 8,993,233, 8,895,308, 8,865,406, 8,889,418, 8,871,445, 8,889,356, 8,932,814, 8,795,965, and 8,906,616. Cpf1 endonuclease and corresponding guide RNAs and PAM sites are disclosed in US Patent Application Publication 2016/0208243 A1. Other CRISPR nucleases useful for editing genomes include Cas12b and Cas12c (see Shmakov et al. (2015) Mol. Cell, 60:385-397; Harrington et al. (2020) Molecular Cell doi: 10.1016/j.molcel.2020.06.022) and CasX and CasY (see Burstein et al. (2016) Nature, doi: 10.1038/nature21059; Harrington et al. (2020) Molecular Cell doi: 10.1016/j.molcel.2020.06.022), or Cas12j (Pausch et al. (2020) Science 10.1126/science.abb1400). Plant RNA promoters for expressing CRISPR guide RNA and plant codon-optimized CRISPR Cas9 endonuclease are disclosed in International Patent Application PCT/US2015/018104 (published as WO 2015/131101 and claiming priority to U.S. Provisional Patent Application 61/945,700). Methods of using CRISPR technology for genome editing in plants are disclosed in US Patent Application Publications US 2015/0082478A1 and US 2015/0059010A1 and in International Patent Application PCT/US2015/038767 A1 (published as WO 2016/007347 and claiming priority to U.S. Provisional Patent Application 62/023,246). All of the patent publications referenced in this paragraph are incorporated

herein by reference in their entirety. In certain embodiments, an RNA-guided endonuclease that leaves a blunt end following cleavage of the target site is used. Blunt-end cutting RNA-guided endonucleases include Cas9, Cas12c, and Cas 12h (Yan et al., 2019). In certain embodiments, an RNA-guided endonuclease that leaves a staggered single stranded DNA overhanging end following cleavage of the target site following cleavage of the target site is used. Staggered-end cutting RNA-guided endonucleases include Cas12a, Cas12b, and Cas12e.

[0081] The methods can also use sequence-specific endonucleases or sequence-specific endonucleases and guide RNAs that cleave a single DNA strand in a dsDNA target site. Such cleavage of a single DNA strand in a dsDNA target site is also referred to herein and elsewhere as "nicking" and can be effected by various "nickases" or systems that provide for nicking. Nickases that can be used include nCas9 (Cas9 comprising a D10A amino acid substitution), nCas12a (e.g., Cas12a comprising an R1226A amino acid substitution; Yamano et al., 2016), Cas12i (Yan et al. 2019), a zinc finger nickase e.g., as disclosed in Kim et al., 2012), a TALE nickase (e.g., as disclosed in Wu et al., 2014), or a combination thereof. In certain embodiments, systems that provide for nicking can comprise a Cas nuclease (e.g., Cas9 and/or Cas12a) and guide RNA molecules that have at least one base mismatch to DNA sequences in the target editing site (Fu et al., 2019). In certain embodiments, genome modifications can be introduced into the target editing site by creating single stranded breaks (i.e., "nicks") in genomic locations separated by no more than about 10, 20, 30, 40, 50, 60, 80, 100, 150, or 200 base pairs of DNA. In certain illustrative and non-limiting embodiments, two nickases (i.e., a CAS nuclease which introduces a single stranded DNA break including nCas9, nCas12a, Cas12i, zinc finger nickases, TALE nickases, combinations thereof, and the like) or nickase systems can directed to make cuts to nearby sites separated by no more than about 10, 20, 30, 40, 50, 60, 80 or 100 base pairs of DNA. In instances where an RNA guided nickase and an RNA guide are used, the RNA guides are adjacent to PAM sequences that are sufficiently close (i.e., separated by no more than about 10, 20, 30, 40, 50, 60, 80, 100, 150, or 200 base pairs of DNA). For the purposes of gene editing, CRISPR arrays can be designed to contain one or multiple guide RNA sequences corresponding to a desired target DNA sequence; see, for example, Cong et al. (2013) *Science*, 339:819-823; Ran et al. (2013) *Nature Protocols*, 8:2281-2308. At least 16 or 17 nucleotides of gRNA sequence are required by Cas9 for DNA cleavage to occur; for Cpf1 at least 16 nucleotides of gRNA sequence are needed to achieve detectable DNA cleavage and at least 18 nucleotides of gRNA sequence were reported necessary for efficient DNA cleavage in vitro; see Zetsche et al. (2015) *Cell*, 163:759-771. In practice, guide RNA sequences are generally designed to have a length of 17-24 nucleotides (frequently 19, 20, or 21 nucleotides) and exact complementarity (i.e., perfect base-pairing) to the targeted gene or nucleic acid sequence; guide RNAs having less than 100% complementarity to the target sequence can be used (e.g., a gRNA with a length of 20 nucleotides and 1-4 mismatches to the target sequence) but can increase the potential for off-target effects. The design of effective guide RNAs for use in plant genome editing is disclosed in US Patent Application Publication 2015/0082478 A1, the entire specification of which is incorporated herein by reference.

More recently, efficient gene editing has been achieved using a chimeric “single guide RNA” (“sgRNA”), an engineered (synthetic) single RNA molecule that mimics a naturally occurring crRNA-tracrRNA complex and contains both a tracrRNA (for binding the nuclease) and at least one crRNA (to guide the nuclease to the sequence targeted for editing); see, for example, Cong et al. (2013) *Science*, 339:819-823; Xing et al. (2014) *BMC Plant Biol.*, 14:327-340. Chemically modified sgRNAs have been demonstrated to be effective in genome editing; see, for example, Hendel et al. (2015) *Nature Biotechnol.*, 985-991. The design of effective gRNAs for use in plant genome editing is disclosed in US Patent Application Publication 2015/0082478 A1, the entire specification of which is incorporated herein by reference.

[0082] Genomic DNA may also be modified via base editing. Both adenine base editors (ABE) which convert A/T base pairs to G/C base pairs in genomic DNA as well as cytosine base pair editors (CBE) which effect C to T substitutions can be used in certain embodiments of the methods provided herein. In certain embodiments, useful ABE and CBE can comprise genome site specific DNA binding elements (e.g., RNA-dependent DNA binding proteins including catalytically inactive Cas9 and Cas12 proteins or Cas9 and Cas12 nickases) operably linked to adenine or cytidine deaminases and used with guide RNAs which position the protein near the nucleotide targeted for substitution. Suitable ABE and CBE disclosed in the literature (Kim, Nat Plants, 2018 Mar.; 4 (3): 148-151) can be adapted for use in the methods set forth herein. In certain embodiments, a CBE can comprise a fusion between a catalytically inactive Cas9 (dCas9) RNA dependent DNA binding protein fused to a cytidine deaminase which converts cytosine (C) to uridine (U) and selected guide RNAs, thereby effecting a C to T substitution; see Komor et al. (2016) *Nature*, 533:420-424. In other embodiments, C to T substitutions are effected with Cas9 nuclease [Cas9n(D10A)] fused to an improved cytidine deaminase and optionally a bacteriophage Mu dsDNA (double-stranded DNA) end-binding protein Gam; see Komor et al., *Sci Adv*. 2017 August; 3 (8):eaao4774. In other embodiments, adenine base editors (ABEs) comprising an adenine deaminase fused to catalytically inactive Cas9 (dCas9) or a Cas9 D10A nuclease can be used to convert A/T base pairs to G/C base pairs in genomic DNA (Gaudelli et al., (2017) *Nature* 551 (7681): 464-471).

[0083] In certain embodiments, zinc finger nucleases or zinc finger nickases can also be used in the methods provided herein. Zinc-finger nucleases are site-specific endonucleases comprising two protein domains: a DNA-binding domain, comprising a plurality of individual zinc finger repeats that each recognize between 9 and 18 base pairs, and a DNA-cleavage domain that comprises a nuclease domain (typically FokI). The cleavage domain dimerizes in order to cleave DNA; therefore, a pair of ZFNs are required to target non-palindromic target polynucleotides. In certain embodiments, zinc finger nuclease and zinc finger nickase design methods which have been described (Urnov et al. (2010) *Nature Rev. Genet.*, 11:636-646; Mohanta et al. (2017) *Genes* vol. 8,12:399; Ramirez et al. *Nucleic Acids Res.* (2012); 40(12): 5560-5568; Liu et al. (2013) *Nature Communications*, 4:2565) can be adapted for use in the methods set forth herein. The zinc finger binding domains of the zinc finger nuclease or nickase provide specificity and can be engineered to specifically recognize any desired target DNA

sequence. The zinc finger DNA binding domains are derived from the DNA-binding domain of a large class of eukaryotic transcription factors called zinc finger proteins (ZFPs). The DNA-binding domain of ZFPs typically contains a tandem array of at least three zinc “fingers” each recognizing a specific triplet of DNA. A number of strategies can be used to design the binding specificity of the zinc finger binding domain. One approach, termed “modular assembly”, relies on the functional autonomy of individual zinc fingers with DNA. In this approach, a given sequence is targeted by identifying zinc fingers for each component triplet in the sequence and linking them into a multifinger peptide. Several alternative strategies for designing zinc finger DNA binding domains have also been developed. These methods are designed to accommodate the ability of zinc fingers to contact neighboring fingers as well as nucleotide bases outside their target triplet. Typically, the engineered zinc finger DNA binding domain has a novel binding specificity, compared to a naturally-occurring zinc finger protein. Engineering methods include, for example, rational design and various types of selection. Rational design includes, for example, the use of databases of triplet (or quadruplet) nucleotide sequences and individual zinc finger amino acid sequences, in which each triplet or quadruplet nucleotide sequence is associated with one or more amino acid sequences of zinc fingers which bind the particular triplet or quadruplet sequence. See, e.g., U.S. Pat. Nos. 6,453,242 and 6,534,261, both incorporated herein by reference in their entirety. Exemplary selection methods (e.g., phage display and yeast two-hybrid systems) can be adapted for use in the methods described herein. In addition, enhancement of binding specificity for zinc finger binding domains has been described in U.S. Pat. No. 6,794,136, incorporated herein by reference in its entirety. In addition, individual zinc finger domains may be linked together using any suitable linker sequences. Examples of linker sequences are publicly known, e.g., see U.S. Pat. Nos. 6,479,626; 6,903,185; and 7,153,949, incorporated herein by reference in their entirety. The nucleic acid cleavage domain is non-specific and is typically a restriction endonuclease, such as FokI. This endonuclease must dimerize to cleave DNA. Thus, cleavage by FokI as part of a ZFN requires two adjacent and independent binding events, which must occur in both the correct orientation and with appropriate spacing to permit dimer formation. The requirement for two DNA binding events enables more specific targeting of long and potentially unique recognition sites. FokI variants with enhanced activities have been described and can be adapted for use in the methods described herein; see, e.g., Guo et al. (2010) *J. Mol. Biol.*, 400:96-107.

[0084] Transcription activator like effectors (TALEs) are proteins secreted by certain *Xanthomonas* species to modulate gene expression in host plants and to facilitate the colonization by and survival of the bacterium. TALEs act as transcription factors and modulate expression of resistance genes in the plants. Recent studies of TALEs have revealed the code linking the repetitive region of TALEs with their target DNA-binding sites. TALEs comprise a highly conserved and repetitive region consisting of tandem repeats of mostly 33 or 34 amino acid segments. The repeat monomers differ from each other mainly at amino acid positions 12 and 13. A strong correlation between unique pairs of amino acids at positions 12 and 13 and the corresponding nucleotide in the TALE-binding site has been found. The simple relation-

ship between amino acid sequence and DNA recognition of the TALE binding domain allows for the design of DNA binding domains of any desired specificity. TALEs can be linked to a non-specific DNA cleavage domain to prepare genome editing proteins, referred to as TAL-effector nucleases or TALENs. As in the case of ZFNs, a restriction endonuclease, such as FokI, can be conveniently used. Methods for use of TALENs in plants have been described and can be adapted for use in the methods described herein, see Mahfouz et al. (2011) Proc. Natl. Acad. Sci. USA, 108:2623-2628; Mahfouz (2011) GM Crops, 2:99-103; and Mohanta et al. (2017) *Genes* vol. 8,12:399). TALE nickases have also been described and can be adapted for use in methods described herein (Wu et al.; Biochem Biophys Res Commun. (2014); 446 (1): 261-6; Luo et al; *Scientific Reports* 6, Article number: 20657 (2016).

[0085] Embodiments of the donor DNA template molecule having a sequence that is integrated at the site of at least one double-strand break (DSB) in a genome include double-stranded DNA, a single-stranded DNA, a single-stranded DNA/RNA hybrid, and a double-stranded DNA/RNA hybrid. In embodiments, a donor DNA template molecule that is a double-stranded (e.g., a dsDNA or dsDNA/RNA hybrid) molecule is provided directly to the plant protoplast or plant cell in the form of a double-stranded DNA or a double-stranded DNA/RNA hybrid, or as two single-stranded DNA (ssDNA) molecules that are capable of hybridizing to form dsDNA, or as a single-stranded DNA molecule and a single-stranded RNA (ssRNA) molecule that are capable of hybridizing to form a double-stranded DNA/RNA hybrid; that is to say, the double-stranded polynucleotide molecule is not provided indirectly, for example, by expression in the cell of a dsDNA encoded by a plasmid or other vector. In various non-limiting embodiments of the method, the donor DNA template molecule that is integrated (or that has a sequence that is integrated) at the site of at least one double-strand break (DSB) in a genome is double-stranded and blunt-ended; in other embodiments the donor DNA template molecule is double-stranded and has an overhang or "sticky end" consisting of unpaired nucleotides (e.g., 1, 2, 3, 4, 5, or 6 unpaired nucleotides) at one terminus or both termini. In an embodiment, the DSB in the genome has no unpaired nucleotides at the cleavage site, and the donor DNA template molecule that is integrated (or that has a sequence that is integrated) at the site of the DSB is a blunt-ended double-stranded DNA or blunt-ended double-stranded DNA/RNA hybrid molecule, or alternatively is a single-stranded DNA or a single-stranded DNA/RNA hybrid molecule. In another embodiment, the DSB in the genome has one or more unpaired nucleotides at one or both sides of the cleavage site, and the donor DNA template molecule that is integrated (or that has a sequence that is integrated) at the site of the DSB is a double-stranded DNA or double-stranded DNA/RNA hybrid molecule with an overhang or "sticky end" consisting of unpaired nucleotides at one or both termini, or alternatively is a single-stranded DNA or a single-stranded DNA/RNA hybrid molecule; in embodiments, the donor DNA template molecule DSB is a double-stranded DNA or double-stranded DNA/RNA hybrid molecule that includes an overhang at one or at both termini, wherein the overhang consists of the same number of unpaired nucleotides as the number of unpaired nucleotides created at the site of a DSB by a nuclease that cuts in an off-set fashion (e.g., where a Cas12 nuclease effects an

off-set DSB with 5-nucleotide overhangs in the genomic sequence, the donor DNA template molecule that is to be integrated (or that has a sequence that is to be integrated) at the site of the DSB is double-stranded and has 5 unpaired nucleotides at one or both termini). In certain embodiments, one or both termini of the donor DNA template molecule contain no regions of sequence homology (identity or complementarity) to genomic regions flanking the DSB; that is to say, one or both termini of the donor DNA template molecule contain no regions of sequence that is sufficiently complementary to permit hybridization to genomic regions immediately adjacent to the location of the DSB. In embodiments, the donor DNA template molecule contains no homology to the locus of the DSB, that is to say, the donor DNA template molecule contains no nucleotide sequence that is sufficiently complementary to permit hybridization to genomic regions immediately adjacent to the location of the DSB. In embodiments, the donor DNA template molecule is at least partially double-stranded and includes 2-20 base-pairs, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 base-pairs; in embodiments, the donor DNA template molecule is double-stranded and blunt-ended and consists of 2-20 base-pairs, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 base-pairs; in other embodiments, the donor DNA template molecule is double-stranded and includes 2-20 base-pairs, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 base-pairs and in addition has at least one overhang or "sticky end" consisting of at least one additional, unpaired nucleotide at one or at both termini. In an embodiment, the donor DNA template molecule that is integrated (or that has a sequence that is integrated) at the site of at least one double-strand break (DSB) in a genome is a blunt-ended double-stranded DNA or a blunt-ended double-stranded DNA/RNA hybrid molecule of about 18 to about 300 base-pairs, or about 20 to about 200 base-pairs, or about 30 to about 100 base-pairs, and having at least one phosphorothioate bond between adjacent nucleotides at a 5' end, 3' end, or both 5' and 3' ends. In embodiments, the donor DNA template molecule includes single strands of at least 11, at least 18, at least 20, at least 30, at least 40, at least 60, at least 80, at least 100, at least 120, at least 140, at least 160, at least 180, at least 200, at least 240, at about 280, or at least 320 nucleotides. In embodiments, the donor DNA template molecule has a length of at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, or at least 11 base-pairs if double-stranded (or nucleotides if single-stranded), or between about 2 to about 320 base-pairs if double-stranded (or nucleotides if single-stranded), or between about 2 to about 500 base-pairs if double-stranded (or nucleotides if single-stranded), or between about 5 to about 500 base-pairs if double-stranded (or nucleotides if single-stranded), or between about 5 to about 300 base-pairs if double-stranded (or nucleotides if single-stranded), or between about 11 to about 300 base-pairs if double-stranded (or nucleotides if single-stranded), or about 18 to about 300 base-pairs if double-stranded (or nucleotides if single-stranded), or between about 30 to about 100 base-pairs if double-stranded (or nucleotides if single-stranded). In embodiments, the donor DNA template molecule includes chemically modified nucleotides (see, e.g., the various modifications of internucleotide linkages, bases, and sugars described in Verma and Eckstein (1998) Annu. Rev. Biochem., 67:99-134); in embodiments, the naturally occur-

ring phosphodiester backbone of the donor DNA template molecule is partially or completely modified with phosphorothioate, phosphorodithioate, or methylphosphonate internucleotide linkage modifications, or the donor DNA template molecule includes modified nucleoside bases or modified sugars, or the donor DNA template molecule is labelled with a fluorescent moiety (e.g., fluorescein or rhodamine or a fluorescent nucleoside analogue) or other detectable label (e.g., biotin or an isotope). In another embodiment, the donor DNA template molecule contains secondary structure that provides stability or acts as an aptamer. Other related embodiments include double-stranded DNA/RNA hybrid molecules, single-stranded DNA/RNA hybrid donor molecules, and single-stranded donor DNA template molecules (including single-stranded, chemically modified donor DNA template molecules), which in analogous procedures are integrated (or have a sequence that is integrated) at the site of a double-strand break. Donor DNA templates provided herein include those comprising CgRRS sequences flanked by DNA with homology to a donor polynucleotide and include the donor DNA template set forth in SEQ ID NO: 9, 21, and equivalents thereof with longer or shorter homology arms. In certain embodiments, a donor DNA template can comprise an adapter molecule (e.g., a donor DNA template formed by annealing single stranded DNAs which do not overlap at their 5' and 3' terminal ends) with cohesive ends which can anneal to an overhanging cleavage site (e.g., introduced by a Cas12a nuclease and suitable gRNAs). In certain embodiments, integration of the donor DNA templates can be facilitated by use of a bacteriophage lambda exonuclease, a bacteriophage lambda beta SSAP protein, and an *E. coli* SSB essentially as set forth in US Patent Application Publication 20200407754, which is incorporated herein by reference in its entirety.

[0086] Donor DNA template molecules used in the methods provided herein include DNA molecules comprising, from 5' to 3', a first homology arm, a replacement DNA, and a second homology arm, wherein the homology arms containing sequences that are partially or completely homologous to genomic DNA (gDNA) sequences flanking a target site-specific endonuclease cleavage site in the gDNA. In certain embodiments, the replacement DNA can comprise an insertion, deletion, or substitution of 1 or more DNA base pairs relative to the target gDNA. In an embodiment, the donor DNA template molecule is double-stranded and perfectly base-paired through all or most of its length, with the possible exception of any unpaired nucleotides at either terminus or both termini. In another embodiment, the donor DNA template molecule is double-stranded and includes one or more non-terminal mismatches or non-terminal unpaired nucleotides within the otherwise double-stranded duplex. In an embodiment, the donor DNA template molecule that is integrated at the site of at least one double-strand break (DSB) includes between 2-20 nucleotides in one (if single-stranded) or in both strands (if double-stranded), e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides on one or on both strands, each of which can be base-paired to a nucleotide on the opposite strand (in the case of a perfectly base-paired double-stranded polynucleotide molecule). Such donor DNA templates can be integrated in genomic DNA containing blunt and/or staggered double stranded DNA breaks by homology-directed repair (HDR). In certain embodiments, a donor DNA template

homology arm can be about 20, 50, 100, 200, 400, or 600 to about 800, or 1000 base pairs in length. In certain embodiments, a donor DNA template molecule can be delivered to a plant cell) in a circular (e.g., a plasmid or a viral vector including a geminivirus vector) or a linear DNA molecule. In certain embodiments, a circular or linear DNA molecule that is used can comprise a modified donor DNA template molecule comprising, from 5' to 3', a first copy of the target sequence-specific endonuclease cleavage site sequence, the first homology arm, the replacement DNA, the second homology arm, and a second copy of the target sequence-specific endonuclease cleavage site sequence. Without seeking to be limited by theory, such modified donor DNA template molecules can be cleaved by the same sequence-specific endonuclease that is used to cleave the target site gDNA of the eukaryotic cell to release a donor DNA template molecule that can participate in HDR-mediated genome modification of the target editing site in the plant cell genome. In certain embodiments, the donor DNA template can comprise a linear DNA molecule comprising, from 5' to 3', a cleaved target sequence-specific endonuclease cleavage site sequence, the first homology arm, the replacement DNA, the second homology arm, and a cleaved target sequence-specific endonuclease cleavage site sequence. In certain embodiments, the cleaved target sequence-specific endonuclease sequence can comprise a blunt DNA end or a blunt DNA end that can optionally comprise a 5' phosphate group. In certain embodiments, the cleaved target sequence-specific endonuclease sequence comprises a DNA end having a single-stranded 5' or 3' DNA overhang. Such cleaved target sequence-specific endonuclease cleavage site sequences can be produced by either cleaving an intact target sequence-specific endonuclease cleavage site sequence or by synthesizing a copy of the cleaved target sequence-specific endonuclease cleavage site sequence. Donor DNA templates can be synthesized either chemically or enzymatically (e.g., in a polymerase chain reaction (PCR)). Donor DNA templates provided herein include those comprising CgRRS sequences flanked by DNA with homology to a donor polynucleotide. An example of a useful DNA donor template provided herein is a DNA molecule comprising SEQ ID NO: 9 or 21.

[0087] Various treatments are useful in delivery of gene editing molecules and/or other molecules to a MON87701 or INIR20 plant cell. In certain embodiments, one or more treatments is employed to deliver the gene editing or other molecules (e.g., comprising a polynucleotide, polypeptide or combination thereof) into a eukaryotic or plant cell, e.g., through barriers such as a cell wall, a plasma membrane, a nuclear envelope, and/or other lipid bilayer. In certain embodiments, a polynucleotide-, polypeptide-, or RNP-containing composition comprising the molecules are delivered directly, for example by direct contact of the composition with a plant cell. Aforementioned compositions can be provided in the form of a liquid, a solution, a suspension, an emulsion, a reverse emulsion, a colloid, a dispersion, a gel, liposomes, micelles, an injectable material, an aerosol, a solid, a powder, a particulate, a nanoparticle, or a combination thereof can be applied directly to a plant, plant part, plant cell, or plant explant (e.g., through abrasion or puncture or otherwise disruption of the cell wall or cell membrane, by spraying or dipping or soaking or otherwise directly contacting, by microinjection). For example, a plant cell or plant protoplast is soaked in a liquid genome editing

molecule-containing composition, whereby the agent is delivered to the plant cell. In certain embodiments, the agent-containing composition is delivered using negative or positive pressure, for example, using vacuum infiltration or application of hydrodynamic or fluid pressure. In certain embodiments, the agent-containing composition is introduced into a plant cell or plant protoplast, e.g., by microinjection or by disruption or deformation of the cell wall or cell membrane, for example by physical treatments such as by application of negative or positive pressure, shear forces, or treatment with a chemical or physical delivery agent such as surfactants, liposomes, or nanoparticles; see, e.g., delivery of materials to cells employing microfluidic flow through a cell-deforming constriction as described in US Published Patent Application 2014/0287509, incorporated by reference in its entirety herein. Other techniques useful for delivering the agent-containing composition to a eukaryotic cell, plant cell or plant protoplast include: ultrasound or sonication; vibration, friction, shear stress, vortexing, cavitation; centrifugation or application of mechanical force; mechanical cell wall or cell membrane deformation or breakage; enzymatic cell wall or cell membrane breakage or permeabilization; abrasion or mechanical scarification (e.g., abrasion with carborundum or other particulate abrasive or scarification with a file or sandpaper) or chemical scarification (e.g., treatment with an acid or caustic agent); and electroporation. In certain embodiments, the agent-containing composition is provided by bacterially mediated (e.g., *Agrobacterium* sp., *Rhizobium* sp., *Sinorhizobium* sp., *Mesorhizobium* sp., *Bradyrhizobium* sp., *Azobacter* sp., *Phyllobacterium* sp.) transfection of the plant cell or plant protoplast with a polynucleotide encoding the genome editing molecules (e.g., RNA dependent DNA endonuclease, RNA dependent DNA binding protein, RNA dependent nickase, ABE, or CBE, and/or guide RNA); see, e.g., Broothaerts et al. (2005) *Nature*, 433:629-633). Any of these techniques or a combination thereof are alternatively employed on the plant explant, plant part or tissue or intact plant (or seed) from which a plant cell is optionally subsequently obtained or isolated; in certain embodiments, the agent-containing composition is delivered in a separate step after the plant cell has been isolated.

**[0088]** In some embodiments, one or more polynucleotides or vectors driving expression of one or more genome editing molecules or trait-conferring genes (e.g., herbicide tolerance, insect resistance, and/or male sterility) are introduced into a MON87701 or INIR20 plant cell. In certain embodiments, a polynucleotide vector comprises a regulatory element such as a promoter operably linked to one or more polynucleotides encoding genome editing molecules and/or trait-conferring genes. In such embodiments, expression of these polynucleotides can be controlled by selection of the appropriate promoter, particularly promoters functional in a eukaryotic cell (e.g., plant cell); useful promoters include constitutive, conditional, inducible, and temporally or spatially specific promoters (e.g., a tissue specific promoter, a developmentally regulated promoter, or a cell cycle regulated promoter). Developmentally regulated promoters that can be used in plant cells include Phospholipid Transfer Protein (PLTP), fructose-1,6-bisphosphatase protein, NAD (P)-binding Rossmann-Fold protein, adipocyte plasma membrane-associated protein-like protein, Rieske [2Fe-2S] iron-sulfur domain protein, chlororespiratory reduction 6 protein, D-glycerate 3-kinase, chloroplastic-like protein,

chlorophyll a-b binding protein 7, chloroplastic-like protein, ultraviolet-B-repressible protein, Soul heme-binding family protein, Photosystem I reaction center subunit psi-N protein, and short-chain dehydrogenase/reductase protein that are disclosed in US Patent Application Publication No. 20170121722, which is incorporated herein by reference in its entirety and specifically with respect to such disclosure. In certain embodiments, the promoter is operably linked to nucleotide sequences encoding multiple guide RNAs, wherein the sequences encoding guide RNAs are separated by a cleavage site such as a nucleotide sequence encoding a microRNA recognition/cleavage site or a self-cleaving ribozyme (see, e.g., Ferré-D'Amaré and Scott (2014) *Cold Spring Harbor Perspectives Biol.*, 2:a003574). In certain embodiments, the promoter is an RNA polymerase III promoter operably linked to a nucleotide sequence encoding one or more guide RNAs. In certain embodiments, the RNA polymerase III promoter is a plant U6 spliceosomal RNA promoter, which can be native to the genome of the plant cell or from a different species, e.g., a U6 promoter from soybean, tomato, or soybean such as those disclosed U.S. Patent Application Publication 2017/0166912, or a homologue thereof; in an example, such a promoter is operably linked to DNA sequence encoding a first RNA molecule including a Cas12a gRNA followed by an operably linked and suitable 3' element such as a U6 poly-T terminator. In another embodiment, the RNA polymerase III promoter is a plant U3, 7SL (signal recognition particle RNA), U2, or U5 promoter, or chimerics thereof, e.g., as described in U.S. Patent Application Publication 20170166912. In certain embodiments, the promoter operably linked to one or more polynucleotides is a constitutive promoter that drives gene expression in eukaryotic cells (e.g., plant cells). In certain embodiments, the promoter drives gene expression in the nucleus or in an organelle such as a chloroplast or mitochondrion. Examples of constitutive promoters for use in plants include a CaMV 35S promoter as disclosed in U.S. Pat. Nos. 5,858,742 and 5,322,938, a rice actin promoter as disclosed in U.S. Pat. No. 5,641,876, a soybean chloroplast aldolase promoter as disclosed in U.S. Pat. No. 7,151,204, and the nopaline synthase (NOS) and octopine synthase (OCS) promoters from *Agrobacterium tumefaciens*. In certain embodiments, the promoter operably linked to one or more polynucleotides encoding elements of a genome-editing system is a promoter from figwort mosaic virus (FMV), a RUBISCO promoter, or a pyruvate phosphate dikinase (PPDK) promoter, which is active in photosynthetic tissues. Other contemplated promoters include cell-specific or tissue-specific or developmentally regulated promoters, for example, a promoter that limits the expression of the nucleic acid targeting system to germline or reproductive cells (e.g., promoters of genes encoding DNA ligases, recombinases, replicases, or other genes specifically expressed in germline or reproductive cells). In certain embodiments, the genome alteration is limited only to those cells from which DNA is inherited in subsequent generations, which is advantageous where it is desirable that expression of the genome-editing system be limited in order to avoid genotoxicity or other unwanted effects. All of the patent publications referenced in this paragraph are incorporated herein by reference in their entirety.

**[0089]** Expression vectors or polynucleotides provided herein may contain a DNA segment near the 3' end of an expression cassette that acts as a signal to terminate tran-

scription and directs polyadenylation of the resultant mRNA and may also support promoter activity. Such a 3' element is commonly referred to as a "3'-untranslated region" or "3'-UTR" or a "polyadenylation signal." In some cases, plant gene-based 3' elements (or terminators) consist of both the 3'-UTR and downstream non-transcribed sequence (Nuccio et al., 2015). Useful 3' elements include: *Agrobacterium tumefaciens* nos 3', tml 3', tmr 3', tms 3', ocs 3', and tr7 3' elements disclosed in U.S. Pat. No. 6,090,627, incorporated herein by reference, and 3' elements from plant genes such as the heat shock protein 17, ubiquitin, and fructose-1,6-biphosphatase genes from wheat (*Triticum aestivum*), and the glutelin, lactate dehydrogenase, and beta-tubulin genes from rice (*Oryza sativa*), disclosed in US Patent Application Publication 2002/0192813 A1. All of the patent publications referenced in this paragraph are incorporated herein by reference in their entireties.

[0090] In certain embodiments, the MON87701 or INIR20 plant cells used herein can comprise haploid, diploid, or polyploid plant cells or plant protoplasts, for example, those obtained from a haploid, diploid, or polyploid plant, plant part or tissue, or callus. In certain embodiments, plant cells in culture (or the regenerated plant, progeny seed, and progeny plant) are haploid or can be induced to become haploid; techniques for making and using haploid plants and plant cells are known in the art, see, e.g., methods for generating haploids in *Arabidopsis thaliana* by crossing of a wild-type strain to a haploid-inducing strain that expresses altered forms of the centromere-specific histone CENH3, as described by Maruthachalam and Chan in "How to make haploid *Arabidopsis thaliana*", protocol available at [www.openwetware.org/images/d/d3/Haploid\\_Arabidopsis\\_protocol.pdf](http://www.openwetware.org/images/d/d3/Haploid_Arabidopsis_protocol.pdf); (Ravi et al. (2014) *Nature Communications*, 5:5334, doi: 10.1038/ncomms6334). Haploids can also be obtained in a wide variety of monocot plants (e.g., soybean, wheat, rice, sorghum, barley) by crossing a plant comprising a mutated CENH3 gene with a wildtype diploid plant to generate haploid progeny as disclosed in U.S. Pat. No. 9,215,849, which is incorporated herein by reference in its entirety. Haploid-inducing soybean lines that can be used to obtain haploid soybean plants and/or cells include Stock 6, MHI (Moldovian Haplod Inducer), indeterminate gametophyte (ig) mutation, KEMS, RWK, ZEM, ZMS, KMS, and well as transgenic haploid inducer lines disclosed in U.S. Pat. No. 9,677,082, which is incorporated herein by reference in its entirety. Examples of haploid cells include but are not limited to plant cells obtained from haploid plants and plant cells obtained from reproductive tissues, e.g., from flowers, developing flowers or flower buds, ovaries, ovules, megasporangia, anthers, pollen, megagametophyte, and microspores. In certain embodiments where the plant cell or plant protoplast is haploid, the genetic complement can be doubled by chromosome doubling (e.g., by spontaneous chromosomal doubling by meiotic non-reduction, or by using a chromosome doubling agent such as colchicine, oryzalin, trifluralin, pronamide, nitrous oxide gas, anti-microtubule herbicides, anti-microtubule agents, and mitotic inhibitors) in the plant cell or plant protoplast to produce a doubled haploid plant cell or plant protoplast wherein the complement of genes or alleles is homozygous; yet other embodiments include regeneration of a doubled haploid plant from the doubled haploid plant cell or plant protoplast. Another embodiment is related to a hybrid plant having at least one parent plant that is a doubled haploid plant pro-

vided by this approach. Production of doubled haploid plants provides homozygosity in one generation, instead of requiring several generations of self-crossing to obtain homozygous plants. The use of doubled haploids is advantageous in any situation where there is a desire to establish genetic purity (i.e., homozygosity) in the least possible time. Doubled haploid production can be particularly advantageous in slow-growing plants or for producing hybrid plants that are offspring of at least one doubled-haploid plant.

[0091] In certain embodiments, the MON87701 or INIR20 plant cells used in the methods provided herein can include non-dividing cells. Such non-dividing cells can include plant cell protoplasts, plant cells subjected to one or more of a genetic and/or pharmaceutically-induced cell-cycle blockage, and the like.

[0092] In certain embodiments, the MON87701 or INIR20 plant cells used in the methods provided herein can include dividing cells. Dividing cells can include those cells found in various plant tissues including leaves, meristems, and embryos. These tissues include dividing cells from young soybean leaf, meristems and scutellar tissue from about 8 or 10 to about 12 or 14 days after pollination (DAP) embryos. The isolation of soybean embryos has been described in several publications (Brettschneider, Becker, and Lötz 1997; Leduc et al. 1996; Frame et al. 2011; K. Wang and Frame 2009). In certain embodiments, basal leaf tissues (e.g., leaf tissues located about 0 to 3 cm from the ligule of a soybean plant; Kirienko, Luo, and Sylvester 2012) are targeted for HDR-mediated gene editing. Methods for obtaining regenerable plant structures and regenerating plants from the NHEJ-, MMEJ-, or HDR-mediated gene editing of plant cells provided herein can be adapted from methods disclosed in US Patent Application Publication No. 20170121722, which is incorporated herein by reference in its entirety and specifically with respect to such disclosure. In certain embodiments, single plant cells subjected to the HDR-mediated gene editing will give rise to single regenerable plant structures. In certain embodiments, the single regenerable plant cell structure can form from a single cell on, or within, an explant that has been subjected to the NHEJ-, MMEJ-, or HDR-mediated gene editing.

[0093] In some embodiments, methods provided herein can include the additional step of growing or regenerating an INIR20 plant from a INIR20 plant cell that had been subjected to the gene editing or from a regenerable plant structure obtained from that INIR20 plant cell. In certain embodiments, the plant can further comprise an inserted transgene, a target gene edit, or genome edit as provided by the methods and compositions disclosed herein. In certain embodiments, callus is produced from the plant cell, and plantlets and plants produced from such callus. In other embodiments, whole seedlings or plants are grown directly from the plant cell without a callus stage. Thus, additional related aspects are directed to whole seedlings and plants grown or regenerated from the plant cell or plant protoplast having a target gene edit or genome edit, as well as the seeds of such plants. In certain embodiments wherein the plant cell or plant protoplast is subjected to genetic modification (for example, genome editing by means of, e.g., an RdDe), the grown or regenerated plant exhibits a phenotype associated with the genetic modification. In certain embodiments, the grown or regenerated plant includes in its genome two or more genetic or epigenetic modifications that in combination provide at least one phenotype of interest. In certain embodi-

ments, a heterogeneous population of plant cells having a target gene edit or genome edit, at least some of which include at least one genetic or epigenetic modification, is provided by the method; related aspects include a plant having a phenotype of interest associated with the genetic or epigenetic modification, provided by either regeneration of a plant having the phenotype of interest from a plant cell or plant protoplast selected from the heterogeneous population of plant cells having a target gene or genome edit, or by selection of a plant having the phenotype of interest from a heterogeneous population of plants grown or regenerated from the population of plant cells having a targeted genetic edit or genome edit. Examples of phenotypes of interest include herbicide resistance, improved tolerance of abiotic stress (e.g., tolerance of temperature extremes, drought, or salt) or biotic stress (e.g., resistance to nematode, bacterial, or fungal pathogens), improved utilization of nutrients or water, modified lipid, carbohydrate, or protein composition, improved flavor or appearance, improved storage characteristics (e.g., resistance to bruising, browning, or softening), increased yield, altered morphology (e.g., floral architecture or color, plant height, branching, root structure). In an embodiment, a heterogeneous population of plant cells having a target gene edit or genome edit (or seedlings or plants grown or regenerated therefrom) is exposed to conditions permitting expression of the phenotype of interest; e.g., selection for herbicide resistance can include exposing the population of plant cells having a target gene edit or genome edit (or seedlings or plants grown or regenerated therefrom) to an amount of herbicide or other substance that inhibits growth or is toxic, allowing identification and selection of those resistant plant cells (or seedlings or plants) that survive treatment. Methods for obtaining regenerable plant structures and regenerating plants from plant cells or regenerable plant structures can be adapted from published procedures (Roest and Gilissen, *Acta Bot. Neerl.*, 1989, 38 (1), 1-23; Bhaskaran and Smith, *Crop Sci.* 30 (6): 1328-1337; Ikeuchi et al., *Development*, 2016, 143:1442-1451). Methods for obtaining regenerable plant structures and regenerating plants from plant cells or regenerable plant structures can also be adapted from US Patent Application Publication No. 20170121722, which is incorporated herein by reference in its entirety and specifically with respect to such disclosure. Also provided are heterogeneous or homogeneous populations of such plants or parts thereof (e.g., seeds), succeeding generations or seeds of such plants grown or regenerated from the plant cells or plant protoplasts, having a target gene edit or genome edit. Additional related aspects include a hybrid plant provided by crossing a first plant grown or regenerated from a plant cell or plant protoplast having a target gene edit or genome edit and having at least one genetic or epigenetic modification, with a second plant, wherein the hybrid plant contains the genetic or epigenetic modification; also contemplated is seed produced by the hybrid plant. Also envisioned as related aspects are progeny seed and progeny plants, including hybrid seed and hybrid plants, having the regenerated plant as a parent or ancestor. The plant cells and derivative plants and seeds disclosed herein can be used for various purposes useful to the consumer or grower. In other embodiments, processed products are made from the INIR20 plant or its seeds, including: (a) soybean seed meal (defatted or non-defatted); (b) extracted proteins, oils, sugars, and starches; (c) fermentation products; (d) animal feed or human food products (e.g.,

feed and food comprising soybean seed meal (defatted or non-defatted) and other ingredients (e.g., other cereal grains, other seed meal, other protein meal, other oil, other starch, other sugar, a binder, a preservative, a humectant, a vitamin, and/or mineral; (e) a pharmaceutical; (f) raw or processed biomass (e.g., cellulosic and/or lignocellulosic material); and (g) various industrial products.

#### Embodiments

[0094] Various embodiments of the plants, genomes, methods, biological samples, and other compositions described herein are set forth in the following sets of numbered embodiments.

[0095] 1a. A transgenic soybean plant cell comprising an INIR20 transgenic locus comprising an originator guide RNA recognition site (OgRRS) in a first DNA junction polynucleotide of a MON87701 transgenic locus and a cognate guide RNA recognition site (CgRRS) in a second DNA junction polynucleotide of the MON87701 transgenic locus.

[0096] 1b. A transgenic soybean plant cell comprising an INIR20 transgenic locus comprising an insertion and/or substitution of DNA in a DNA junction polynucleotide of a MON87701 transgenic locus with DNA comprising a cognate guide RNA recognition site (CgRRS) or comprising a deletion in a 5' or 3' junction polynucleotide of a MON87701 transgenic locus.

[0097] 2. The transgenic soybean plant cell of embodiment 1a or 1b, wherein said CgRRS comprises the DNA molecule set forth in SEQ ID NO: 8 or 7; and/or wherein said MON87701 transgenic locus is set forth in SEQ ID NO:1, is present in seed deposited at the ATCC under accession No. PTA-8194, is present in progeny thereof, is present in allelic variants thereof, or is present in other variants thereof.

[0098] 3. The transgenic soybean plant cell of embodiments 1a, 1b, or 2, wherein said INIR20 transgenic locus comprises the DNA molecule set forth in SEQ ID NO: 2, 3, 15, or an allelic variant thereof.

[0099] 4. A transgenic soybean plant part comprising the soybean plant cell of any one of embodiments 1a, 1b, 2, or 3, wherein said soybean plant part is optionally a seed.

[0100] 5. A transgenic soybean plant comprising the soybean plant cell of any one of embodiments 1a, 1b, 2, or 3.

[0101] 6. A method for obtaining a bulked population of inbred seed comprising selfing the transgenic soybean plant of embodiment 5 and harvesting seed comprising the INIR20 transgenic locus from the selfed soybean plant.

[0102] 7. A method of obtaining hybrid soybean seed comprising crossing the transgenic soybean plant of embodiment 5 to a second soybean plant which is genetically distinct from the first soybean plant and harvesting seed comprising the INIR20 transgenic locus from the cross.

[0103] 8. A DNA molecule comprising SEQ ID NO: 2, 3, 7, 8, 9, 14, 15, or 21.

[0104] 9. A processed transgenic soybean plant product comprising the DNA molecule of embodiment 8.

[0105] 10. A biological sample containing the DNA molecule of embodiment 8.

[0106] 11. A nucleic acid molecule adapted for detection of genomic DNA comprising the DNA molecule of embodiment 8, wherein said nucleic acid molecule optionally comprises a detectable label.

[0107] 12. A method of detecting a soybean plant cell comprising the INIR20 transgenic locus of any one of embodiments 1a, 1b, 2, or 3, comprising the step of detecting DNA molecule comprising SEQ ID NO: 2, 3, 7, 8, 9, 14, 15, or 21.

[0108] 13. A method of excising the INIR20 transgenic locus from the genome of the soybean plant cell of any one of embodiments 1a, 1b, 2, or 3, comprising the steps of:

[0109] (a) contacting the plant cell or genome thereof with: (i) an RNA dependent DNA endonuclease (RdDe); and (ii) a guide RNA (gRNA) capable of hybridizing to the guide RNA hybridization site of the OgRRS and the CgRRS; wherein the RdDe recognizes a OgRRS/gRNA and a CgRRS/gRNA hybridization complex; and,

[0110] (b) selecting a transgenic plant cell, transgenic plant part, or transgenic plant wherein the INIR20 transgenic locus flanked by the OgRRS and the CORRS has been excised.

[0111] 14. The method of embodiment 13, wherein said INIR20 transgenic locus comprises the CgRRS in SEQ ID NO: 8 or 7 and the guide RNA comprises an RNA sequence encoded by SEQ ID NO: 11.

[0112] 15. The method of embodiment 14, wherein said INIR20 transgenic locus comprises the DNA molecule set forth in SEQ ID NO: 3, 15, or an allelic variant thereof.

#### EXAMPLES

**Example 1. Application of a Cas12a RNA Guided Endonuclease and Guide RNAs to Change or Excise the 5'-T-DNA Junction Sequence in the MON87701 Event**

[0113] The MON87701 5' junction polynucleotide sequence set forth in SEQ ID NO: 17 contains at least two Cas12a recognition sequences. The Guide-1 (gRNA-1) and Guide-2 (gRNA-2) sequence locations in the 5' junction polynucleotide are shown in FIG. 4. These guide RNAs can be used to modify some of the 5' junction polynucleotide sequence. In one embodiment, Guide-1 or Guide-2 are used alone to disrupt the MON87701 5'-junction sequence (e.g., by using a Cas12a endonuclease and 1 of Guide-1 or Guide-2 to cleave the 5' junction polynucleotide sequence and recovering genomic edits where the 5' DNA junction polynucleotide sequence of MON87701 is disrupted).

[0114] The Cas12a nuclease and the guide RNA is introduced into soybean plant cells containing the MON87701 event. In certain embodiments, the Cas12a nuclease and gRNA(s) are encoded and expressed from a T-DNA transformed into the MON87701 event via *Agrobacterium*-mediated transformation. Alternatively, the T-DNA can be transformed into any convenient soy line, and then crossed with the MON87701 event to combine the Cas12a ribonucleoprotein expressing T-DNA with the MON87701 event. The Cas12a nuclease and gRNAs can also be assembled in vitro then delivered to MON87701 explants as ribonucleoprotein complexes using a biolistic approach (Svitashov et al., Nat Commun. 2016; 7:13274; Zhang et al., 2021, Plant Commun. 2 (2): 100168). Also, a plasmid encoding a Cas12a nuclease, and the gRNA(s) can be delivered to MON87701 explants using a biolistic approach. This will produce plant cells that have a high likelihood of incurring mutations that disrupt the MON87701 5' junction polynucleotide sequence.

[0115] In the *Agrobacterium* approach, a binary vector that contains a strong constitutive expression cassette like the AtUbi10 promoter::AtUbi10 terminator driving Cas12a, a PolII or PolIII gene cassette driving the Cas12a gRNA(s) and a CaMV 35S:NPTII:NOS (e.g., for G418 or neomycin selection) or other suitable plant selectable marker (e.g., a phosphomannose isomerase (Reed et al. 2001, *In Vitro Cellular & Developmental Biology-Plant* 37:127-132) or hygromycin phosphotransferase (Itaya, et al. 2018, *In Vitro Cellular & Developmental Biology-Plant* 54:184-194)) is constructed and cells comprising the integrated T-DNA(s) are selected using an appropriate selection agent. An expression cassette driving a fluorescent protein like mScarlet may also be useful to monitor the plant transformation process.

[0116] The T-DNA-based expression cassettes are delivered from superbinary vectors in *Agrobacterium* strain LBA4404. Soy transformations are performed based on published methods (Zhang et al., 1999, *Plant Cell, Tissue and Organ Culture* 56 (1), 37-46). Briefly, cotyledonary explants are prepared from the 5-day-old soybean seedlings by making a horizontal slice through the hypocotyl region, approximately 3-5 mm below the cotyledon. A subsequent vertical slice is made between the cotyledons, and the embryonic axis is removed. This generates 2 cotyledonary node explants. Approximately 7-12 vertical slices are made on the adaxial surface of the explant about the area encompassing 3 mm above the cotyledon/hypocotyl junction and 1 mm below the cotyledon/hypocotyl junction. Explant manipulations are done with a No. 15 scalpel blade.

[0117] Explants are immersed in the *Agrobacterium* inoculum for 30 min and then co-cultured on 100×15 mm Petri plates containing the *Agrobacterium* resuspension medium solidified with 0.5% purified agar (BBL Cat #11853). The co-cultivation plates are overlaid with a piece of Whatman #1 filter paper (Mullins et al., 1990; Janssen and Gardner, 1993; Zhang et al., 1997). The explants (5 per plate) are cultured adaxial side down on the co-cultivation plates, that are overlaid with filter paper, for 3 days at 24° C., under an 18/6 hour light regime with an approximate light intensity of 80  $\mu\text{mol s}^{-1} \text{m}^{-2}$  (F17T8/750 cool white bulbs, Litetronics®). The co-cultivation plates are wrapped with Parafilm®.

[0118] Following the co-cultivation period explants are briefly washed in B5 medium supplemented with 1.67 mg  $1^{-1}$  BAP, 3% sucrose, 500 mg  $1^{-1}$  ticarcillin and 100 mg  $1^{-1}$  cefotaxime. The medium is buffered with 3 mM MES, pH 5.6. Growth regulator, vitamins and antibiotics are filter sterilized post autoclaving. Following the washing step, explants are cultured (5 per plate) in 100× 20 mm Petri plates, adaxial side up with the hypocotyl imbedded in the medium, containing the washing medium solidified with 0.8% purified agar (BBL Cat #11853) amended with either G418, neomycin, or kanamycin at concentrations permitting selection of transformants. This medium is referred to as shoot initiation medium (SI). Plates are wrapped with 3M pressure sensitive tape (Scotch™, 3M, USA) and cultured under the environmental conditions used during the seed germination step (at 24° C., 18/6 light regime, under a light intensity of approximately 150  $\mu\text{mol s}^{-1} \text{m}^{-2}$ ).

[0119] After 2 weeks of culture, the hypocotyl region is excised from each of the explants, and the remaining explant, cotyledon with differentiating node, is subsequently subcultured onto fresh SI medium. Following an additional 2 weeks of culture on SI medium, the cotyledons are

removed from the differentiating node. The differentiating node is subcultured to shoot elongation medium (SE) composed of Murashige and Skoog (MS) (1962) basal salts, B5 vitamins, 1 mg 1<sup>-1</sup> zeatin-riboside, 0.5 mg 1<sup>-1</sup> GA3 and 0.1 mg 1<sup>-1</sup> IAA, 50 mg 1<sup>-1</sup> glutamine, 50 mg 1<sup>-1</sup> asparagine, 3% sucrose and 3 mM MES, pH 5.6. The SE medium is amended with G418, neomycin, or kanamycin at concentrations permitting selection of transformants. The explants are subcultured biweekly to fresh SI medium until shoots reach a length greater than 3 cm. The elongated shoots are rooted on Murashige and Skoog salts with B5 vitamins, 1% sucrose, 0.5 mg 1<sup>-1</sup> NAA without further selection in Magenta boxes®.

[0120] When a sufficient amount of viable tissue is obtained, it can be screened for mutations at the MON87701 junction sequence, using a PCR-based approach. One way to screen is to design DNA oligonucleotide primers that flank and amplify the MON87701 junction plus surrounding sequence. For example, the primers of SEQ ID NO: 12 and SEQ ID NO: 13 will produce a product in a PCR reaction that can be analyzed for edits at the target site. The size of this product will vary based on the nature of the edit. Amplicons can be sequenced directly using an amplicon sequencing approach or ligated to a convenient plasmid vector for Sanger sequencing. Those plants in which the MON87701 5'-junction sequence is disrupted are selected and grown to maturity. The DNA encoding the Cas12a reagents can be segregated away from the modified junction sequence in a subsequent generation.

Example 2. Insertion of a CgRRS Element in the 5'-Junction of the MON87701 Event

[0121] Two plant gene expression vectors are prepared. Plant expression cassettes for expressing a bacteriophage lambda exonuclease, a bacteriophage lambda beta SSAP protein, and an *E. coli* SSB are constructed essentially as set forth in US Patent Application Publication 20200407754, which is incorporated herein by reference in its entirety. A DNA sequence encoding a tobacco c2 nuclear localization signal (NLS) is fused in-frame to the DNA sequences encoding the exonuclease, the bacteriophage lambda beta SSAP protein, and the *E. coli* SSB to provide a DNA sequence encoding the c2 NLS-Exo, c2 NLS lambda beta SSAP, and c2 NLS-SSB fusion proteins that are set forth in SEQ ID NO: 135, SEQ ID NO: 134, and SEQ ID NO: 133 of US Patent Application Publication 20200407754, respectively, and incorporated herein by reference in their entireties. DNA sequences encoding the c2 NLS-Exo, c2 NLS lambda beta SSAP, and c2NLS-SSB fusion proteins are operably linked to suitable promoter(s) (e.g., AtUbi10, CaMV35S, and/or SIUbi10 promoter) and suitable polyadenylation site(s) (e.g., nos 3', PeaE9 3', tmr 3', tms 3', AtUbi10 3', and tr7 3' elements), to provide the exonuclease, SSAP, and SSB plant expression cassettes.

[0122] A DNA donor template sequence (SEQ ID NO: 9 or 21) that targets the 5'-T-DNA junction polynucleotide of the MON87701 event (SEQ ID NO:1) for HDR-mediated insertion of a 27 base pair OgRRS sequence (SEQ ID NO: 6) that is identical to a Cas12a recognition site (i.e., OgRRS) at the 3'-junction polynucleotide of the MON87701 T-DNA insert is constructed. The location of the OgRRS in the 3' junction polynucleotide of SEQ ID NO: 1 is depicted in FIG. 5. The DNA donor sequence includes a replacement template with desired insertion region (27 base pairs long)

flanked on both sides by homology arms about 500-635 bp in length. The homology arms match (i.e., are homologous to) gDNA (genomic DNA) regions flanking the target genomic DNA insertion site (SEQ ID NO: 10) in the MON87701 transgenic locus (SEQ ID NO: 1). The replacement template region comprising the donor DNA is flanked at each end by DNA sequences identical to the MON87701 5' junction polynucleotide sequence and contains a CgRRS element recognized by the same Cas12a RNA-guided nuclease and a gRNA (e.g., comprising an RNA encoded by SEQ ID NO: 11) that recognize the OgRRS located in the 3' junction polynucleotide.

[0123] A plant expression cassette that provides for expression of the RNA-guided sequence-specific Cas12a endonuclease is constructed. A plant expression cassette that provides for expression of a guide RNA (e.g., encoded by SEQ ID NO: 4 or 5) complementary to sequences adjacent to the insertion site is constructed. An *Agrobacterium* superbinary plasmid transformation vector containing a cassette that provides for the expression of a suitable plant selectable marker (e.g., a neomycin phosphotransferase (nptII) or hygromycin phosphotransferase (hpt)) is constructed. Once the cassettes, donor sequence and *Agrobacterium* superbinary plasmid transformation vector are constructed, they are combined to generate two soybean transformation plasmids. In other embodiments, other gRNAs (Guide-1 or Guide-2) can be used to introduce double stranded breaks in the MON87701 5' junction polynucleotide for insertion of a CgRRS using similar donor DNA templates and the aforementioned Cas12a, SSAP, SSB, and EXO reagents.

[0124] A soybean transformation plasmid is constructed with a neomycin phosphotransferase (nptII) or hygromycin phosphotransferase (hpt) cassette, the RNA-guided sequence-specific endonuclease cassette, the guide RNA cassette, and the MON87701 5'-T\_DNA junction sequence DNA donor sequence into the *Agrobacterium* superbinary plasmid transformation vector (the control vector).

[0125] A soybean transformation plasmid is constructed with a neomycin phosphotransferase (nptII) or hygromycin phosphotransferase (hpt) cassette, the RNA-guided sequence-specific endonuclease cassette, the guide RNA cassette, the SSB cassette, the lambda beta SSAP cassette, the Exo cassette, and the MON87701 5'-T\_DNA junction sequence donor DNA template sequence (SEQ ID NO: 9 or 21) into the *Agrobacterium* superbinary plasmid transformation vector (the lambda red vector).

[0126] All constructs are transformed into *Agrobacterium* strain LBA4404.

[0127] Soybean transformations are performed as described in Example 1 or based on published methods (Ishida et. al, Nature Protocols 2007; 2, 1614-1621). Briefly, immature embryos from inbred line GIBE0104, approximately 1.8-2.2 mm in size, are isolated from surface sterilized ears 10-14 days after pollination. Embryos are placed in an *Agrobacterium* suspension made with infection medium at a concentration of OD 600=1.0. Acetosyringone (200 µM) is added to the infection medium at the time of use. Embryos and *Agrobacterium* are placed on a rocker shaker at slow speed for 15 minutes. Embryos are then poured onto the surface of a plate of co-culture medium. Excess liquid media is removed by tilting the plate and drawing off all liquid with a pipette. Embryos are flipped as necessary to maintain a scutellum up orientation. Co-culture plates are placed in a box with a lid and cultured in the dark at 22° C. for 3 days.

Embryos are then transferred to resting medium, maintaining the scutellum up orientation. Embryos remain on resting medium for 7 days at 27-28° C. Embryos that produced callus are transferred to Selection 1 medium with G418 or hygromycin at concentrations permitting selection of transformants when a nptII or hpt selectable marker, respectively, is used and cultured for an additional 7 days. Callused embryos are placed on Selection 2 medium with suitable concentrations of the selection agent for the selectable marker and cultured for 14 days at 27-28° C. Growing calli resistant to the selection agent are transferred to Pre-Regeneration media with suitable concentrations of the selection agent for the selectable marker to initiate shoot development. Calli remains on Pre-Regeneration media for 7 days. Calli beginning to initiate shoots are transferred to Regeneration medium with G418 or hygromycin at concentrations permitting selection of transformants when a nptII or hpt selectable marker is used in Phytatrays and cultured in light at 27-28° C. Shoots that reached the top of the Phytatray with intact roots are isolated into Shoot Elongation medium prior to transplant into soil and gradual acclimatization to greenhouse conditions.

[0128] When a sufficient amount of viable tissue is obtained, it can be screened for insertion at the MON87701 junction sequence, using a PCR-based approach. The PCR

primer on the 5'-end can be SEQ ID NO: 12. The PCR primer on the 3'-end can be SEQ ID NO: 13). The above primers that flank donor DNA homology arms are used to amplify the MON87701 5'-junction polynucleotide sequence. The correct donor sequence insertion will produce a PCR product which can be distinguished from PCR products obtained from unedited MON87701 loci. Unique DNA fragments comprising a CgRRS in the MON87701 5' junction polynucleotide are set forth in SEQ ID NO: 7, 8, 9, 14. Amplicons can be sequenced directly using an amplicon sequencing approach or ligated to a convenient plasmid vector for Sanger sequencing. Those plants in which the MON87701 junction sequence now contains the intended Cas12a recognition sequence (e.g., a CgRRS of SEQ ID NO: 7, 8, or 14) are selected and grown to maturity. The T-DNA encoding the Cas12a reagents can be segregated away from the modified junction sequence in a subsequent generation. The resultant INIR20 transgenic locus (SEQ ID NO: 2 or 15) comprising the CgRRS and OgRRS (e.g., which each comprise SEQ ID NO: 6) can be excised using Cas12a and a suitable gRNA which hybridizes to DNA comprising SEQ ID NO: 11 at both the OgRRS and the CORRS.

[0129] The breadth and scope of the present disclosure should not be limited by any of the above-described embodiments.

#### SEQUENCE LISTING

```

Sequence total quantity: 21
SEQ ID NO: 1      moltype = DNA  length = 14416
FEATURE          Location/Qualifiers
source           1..14416
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 1
ggggaaatcc ctcttcata ttaagaacat aaaaatcaac agggaaaaata agttcaccaa 60
cccgaaaccat cacatcctt atgaaacctg cggggtaagc agcacttcta ttttccaaat 120
gaatcaccac atctgtatgc tgcaagggtc caagagataa agaatggaaa atggacagag 180
gggtgacact aactgtatgc cctagatcta gcattgtatt atcaatatta ctgttccaa 240
taatgtcaagg tatacagaca gtacctgtt ccttacatct ctcaggaatg taaggaaacaa 300
atttacatat caatgtcgac acatttctgc ccatgtcaat cctttcattt ccttgcgt 360
tccttttgc ggtgcacaaat tcctttagaa acttgacaca tccttgcatt tgcttgatgg 420
catctagcag aggtatgtt acctctactt tcctgaagg ctccaagatc tcctttctg 480
cttcttcattt tttttttcat ttggaaatgtt caagggttgc atggaaaggd gataaggaggc 540
tgcggtaagt cagaattact agaagaaggd ccacctgtcat gaaaatttt gttaggaage 600
tttcttttgc ttgtcaactat ctcatccctt ttctcaggatc tagaaatgttgc ctggacaggd 660
tcaggtgcgg gtgtgtctac ttgtggaggd acttgaaattt ggttgcaga cctcaaggd 720
atgacactca catttttcgg attttgcaca gtttgcaga gcaattttgtc agaattttgg 780
gaatgagctt ggttcaactt agtaggcattt cgcggccat tattttgcag actctgtatgt 840
aagggtcttg tctcttgcgtt aatttgcata ttcttgcattt tcattttgcct cactaactct 900
tctaaggaaag gtttgcggg aggcttgcgtt gtttgcattt ttgttgcattt tcattttgcgt 960
ttgtgtgtctt gtttgcattt ttgttgcattt ttgttgcattt ttgttgcattt tcattttgcgt 1020
ggaggggacac acttgttgcgtt ttgttgcattt ttgttgcattt ttgttgcattt tcattttgcgt 1080
caacctggat tttttttttt ttgttgcattt ttgttgcattt ttgttgcattt ttgttgcattt tcattttgcgt 1140
ttgttgcattt ttgttgcattt ttgttgcattt ttgttgcattt ttgttgcattt ttgttgcattt tcattttgcgt 1200
gattactgcata aagaaggaca aagatctgtt ttgttgcattt ttgttgcattt ttgttgcattt tcattttgcgt 1260
actcttgcata cttttttttt ttgttgcattt ttgttgcattt ttgttgcattt ttgttgcattt tcattttgcgt 1320
gcatacaattt tttttttttt ttgttgcattt ttgttgcattt ttgttgcattt ttgttgcattt tcattttgcgt 1380
tcatcgactt cttttttttt ttgttgcattt ttgttgcattt ttgttgcattt ttgttgcattt tcattttgcgt 1440
gcctttttttt caatccatca cttttttttt ttgttgcattt ttgttgcattt ttgttgcattt ttgttgcattt tcattttgcgt 1500
ctggcagcat caatccatca cttttttttt ttgttgcattt ttgttgcattt ttgttgcattt ttgttgcattt tcattttgcgt 1560
agggttgcgtt cttttttttt ttgttgcattt ttgttgcattt ttgttgcattt ttgttgcattt tcattttgcgt 1620
cagttactcat acaatgcattt tttttttttt ttgttgcattt ttgttgcattt ttgttgcattt ttgttgcattt tcattttgcgt 1680
gcagtggccat tagatgcagg gaagaatttcc tccaaagaca ccctttaag gtcattttttt 1740
ctgaaaatgg acctggggagc aaggtagtagt agccaaatctt ttgttgcattt ttgttgcattt ttgttgcattt tcattttgcgt 1800
tgaggaaatgg cttttttttt ttgttgcattt ttgttgcattt ttgttgcattt ttgttgcattt tcattttgcgt 1860
caaacaatat ggaactcattt aagatgtttt tgaggatctt cttttttttt ttgttgcattt ttgttgcattt tcattttgcgt 1920
ttgggttagca aatgtttagt tccaggatctt agaacaatgtt gaacaccctt atcaggatctt 1980
tgaatgcaca agttttccata agtgaaatca ggtgcaccca tctcccttaag agtccctctca 2040
cgagggtggag gtttgcattt ttgttgcattt ttgttgcattt ttgttgcattt ttgttgcattt tcattttgcgt 2100
aatatttcag aatcaccaga aacaaaatatac tcagaatgtt caaaatgtc aaaaatgcaca 2160

```

-continued

---

taatgattag	gatgcacact	atgcctaact	aatcttatgaa	aggttctatc	tatccagga	2220	
tgcgagggtt	ataatccac	tagatgcc	ctagtcatgc	actatatgt	gcaataatg	2280	
tgttctcaa	aaacccaa	caagcacca	gggagggtt	aaaactacaac	tatgtcaaa	tgatatccaa	2340
atgagttgaa	attttgtgag	cagcaccct	aaatcatgaa	aagatagcac	aaaaatttc	2400	
aaacgaaaat	tcaaaagtcta	actatgaaa	ctacttaaga	aaagtttga	aaaataggac	2460	
aataatactt	gaaaaataaa	aaaaacata	gtaaacacgt	gattttgcg	gtttgggaga	2520	
ctccaacccg	ctaaaaacgg	ttgcacaat	atgagaaat	tttttctacc	ccaaatgcca	2580	
caatatgaga	aagtttgc	aaaatctgt	tcccaaaat	tttgctctc	tcaattccaa	2640	
ccacaccaag	tgctcttagt	attttcaca	aaaaatata	gccaaaata	acaactctaa	2700	
ctatcaaac	aaaatcagct	aattttatg	caaataatgt	cgctaattcc	tagtcactaa	2760	
tcactgttca	cagcaaaaca	ccaaactgt	cagtcgct	acagtcgt	aaaggagac	2820	
gcaactgaaa	tgcaaaacag	aatgtacac	aaaacaaac	aactaaacac	tattatgaa	2880	
cttggccca	ctgtccccg	acaacggcgc	caaatttgat	cgaggtcg	cccgaatcaa	2940	
ataaaccata	aaaatgcgt	atctggaa	tgatcttgc	tcatctcca	acgagcaatg	3000	
gtcaaccaat	gttcaata	gatagtata	aaacaataa	gaattgggg	gggggggtat	3060	
ttgttttgt	aattttaca	acaacaaat	ttaattaga	aaatacaga	attaaacat	3120	
gttatttccc	cttgattcat	aagcaagtct	cttatectag	gttaggagga	tttatccct	3180	
accaggctca	ccacttata	caaccctaa	ttaaattat	aagcggaaat	taacataagg	3240	
ttgtctttat	atgattaagc	aacacata	ccaattatc	atgaacaaa	tgcgatcatta	3300	
agcatcaaca	taaattaaagc	gcaaaatata	ttaatcaac	actaagcatg	catggattag	3360	
tagcaacaaa	tacagtagaa	ttggtgaga	tgaaaaactg	atcaatattc	aatagtaata	3420	
acaaaacctc	aaagagagg	gtgttgat	ctcaagagaa	acaacgcgt	gagacttagc	3480	
cttccattaa	tcgttagaaa	acgaaattgt	agaaaacgaa	ttttattct	tgtgacaat	3540	
gtgcgtacac	agtaataaaa	actggatgt	caaaaccta	aaattatct	tctctccaaa	3600	
aaaactccct	aaactaaaac	cctggtgct	ttatataat	cctcagoccc	aaagttaca	3660	
aatctatccc	cagtccaaac	ccataaaacga	aataaaataa	aatctggaca	agataaagata	3720	
agattggat	aaataaaatc	tggacgaaat	aaaatctgg	taagataaga	tttgataaaa	3780	
taaaatttgc	tgcttttc	aagtccago	ccaaattccg	attcaagccc	attttttat	3840	
aattcttctg	aaattttat	aaaaatacga	aattatgtca	gtaggccaa	atgataaaa	3900	
tgcatataat	atttgacaat	taaggctaa	cgtatattaa	aatatgtaca	aaaagggtt	3960	
agaatatgg	gataatgtac	acatcaccac	tatggggagc	aattttaaa	tgcatttgag	4020	
tttttttacc	tgagacacag	tgcaatgg	ttcccaagga	ttcatttg	ctttttttt	4080	
atatgatgg	gtcactacat	tggcttgc	aaagaaactg	aatttgggg	attaaagaaa	4140	
cacaaaataa	aaacaaatga	aactatgtaa	tagaaatgt	gcctattgt	tcttggaaaa	4200	
agtccaaacc	tttgcgtatt	ggataaaaat	catattacc	actttagtgc	tgtcaatca	4260	
aacactatgg	ttggataaaa	tctcaactct	agatataatc	caaggatata	tatgaccaac	4320	
attagtcat	tttagaaatgt	aaagtggaca	aatttggat	ttcattccct	aatgacatta	4380	
taaacatgt	ttttttccat	gaccctttt	caatgtaaat	acaattttc	ccttagtta	4440	
gataactctat	atatgcatt	taatgtatgt	atgaaaatgt	acctaagtgt	ttgtgtatgg	4500	
ttaagtttgc	gactactct	gatataac	tcctcatctc	caatctata	caaaagata	4560	
ttgtcaatgc	gtacctgaa	tttgcgtat	tgcaatgtgt	agtttcttct	gaaggccac	4620	
gcttgtatag	taaccagaag	ccaggaggga	gtctctaa	gtctcaactc	gtatccct	4680	
tggaaatgt	ttttttctt	taaagaaaaa	agagataatgt	taccaatgt	aatatccct	4740	
tagccaaata	gggacatcat	agaaaaacaa	acttttctt	atgatatttta	atgcaactac	4800	
atattttagg	tcgcgttgt	tcgtataaaa	ataagggtct	agacaacaca	aaaatattt	4860	
tccaacgtt	gattttaaa	atggctgaga	gacaataaca	aataaagaat	gatgactgg	4920	
acaaaacctt	aaaatctgtt	aactctactg	atctcatata	actttttttt	cagtgtctaa	4980	
aaaatggat	aaatacaat	tattttat	ttttactttt	attatctac	acctttttc	5040	
tactcattt	tttcaattc	ccttccat	gggcaccc	ttttgtccgc	gagagtgcgt	5100	
tggactttt	ttgtttctt	tttgcatt	atttctttt	tttcattgtt	attttattca	5160	
aatgttccca	tcatcatct	actctttt	gttatgttt	tttttttttgc	ccaaactccaa	5220	
cgaggccgt	ccgcgcac	catcatc	accttgc	ggcctcagc	cgcaaggccc	5280	
tgaccccaat	ggcgcacagg	gtcatgc	tttcttaa	gtgtcttct	tttgcgtat	5340	
cgtcaatgt	ttgtatgtt	accttgcatt	tttcaatgt	tttttttttgc	tgtgggttag	5400	
tctagggtgc	tctggccgtt	tccaaacc	ggccaaaaaa	aatgaaatgt	ggtagggaa	5460	
ggggccctag	tttgcattaa	aataaataat	gtcaatgtat	tgataactgc	tatgtatagg	5520	
tatattttgg	gattttat	taatgtat	attttttttt	tcctctttt	tcttccct	5580	
tgttcaata	atttgcattt	taatcatcatt	taatgtttt	tgtatggaaat	attaaaatgtt	5640	
gatgtatgtt	tgatacttag	tgtatgtt	gatgtatgtt	ttttccgttgc	ttttccgttgc	5700	
gatgtatgtt	tgatgtatgtt	gatgtatgtt	ttttccgttgc	ttttccgttgc	ttttccgttgc	5760	
caaaaactgt	tgatgtatgtt	gatgtatgtt	ttttccgttgc	ttttccgttgc	ttttccgttgc	5820	
tatcgtat	tcgcgtatgtt	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	5880	
tcgcacggatc	cccggttacc	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	5940	
tcgcacggatc	cccggttacc	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	5940	
tgcgtatgtt	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	6000	
tactttaaga	aaatcttac	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	6060	
tctagcgtat	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	6120	
atataatata	attttttt	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	6180	
tatgtgtgt	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	6240	
tgcacaaat	aaatcttac	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	6300	
gatgtatgtt	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	6360	
gtatgtatgtt	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	6420	
cgttaggttt	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	6480	
tgaatataatcc	tctataat	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	6540	
tatgtatgtt	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	6600	
gacccaaatcc	atattccat	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	6660	
tcaatataat	tgcaacggaa	ttttccgttgc	ttttccgttgc	ttttccgttgc	ttttccgttgc	6720	

-continued

---

taggaggttg ggaagacagg cccagaaaga gatttatctg acttgtttt tgtatagttt	6780
tcaatgttca taaaggaga tggagacttg agaaagtttt ttggactt gttagctt	6840
gttggcgtt tttttttt atcaataact ttgttggct tatgttggt aatatttcg	6900
tggactctt agtttttaa gacgtgctaa ctttgggg ctatgactt gtgtacat	6960
attgttaacag atgacttgat gtgcgactaa tctttacaca taaaacatag ttctgtttt	7020
tgaaggttct tattttcatt ttatattgaa tgttatataat ttatataatctt ttatataatctt	7080
agtaaaaggc aaattttgtct ttaaatggaa aaaatataat attccacatgt ttccacat	7140
ctttagcatt tagcgttaca aattcaaaaa ttcccattt ttatcatga atcataccat	7200
tatataattaa ctaaatccaa ggtaaaaaaa aggtatgaaa gctctatgt aagtaaaaata	7260
taaattcccc ataaggaaag ggccaagttcc accaggcaag taaaatggc aagcaccaat	7320
ccaccatcac acaattttcac tcatagataa cgataagat catggataa ttctccacgt	7380
ggcatttttc cagcggttca agccgataag ggttcaaca cctctctt ggccttgg	7440
gccgttacca agtaaaatta acctcacaca tatccacact caaaatccaa cgggttagat	7500
cctagttccac ttgaatctca tgatctctt accctccgt cactccaaag ctgttctca	7560
ttgttggttat cattatataat agatggcaa agcactagac caaacctcg tcacacaaag	7620
agtaaaaggc aacaatggct ttcttcatgt ttcttcccg tactatgtt gccttccgg	7680
ctcaggccac tatggtcgtt ctttcaacg gacttaatgc ctccgttgc ttccagcca	7740
cccgcaaggc taacaacgc atttccca tcacaagca cggcggaa gttactgca	7800
tgcaggtgtt gcttccgatt ggaaagaaga agttttagac tctcttcttccctgttgc	7860
ttaccgttcc cgggtgtcgc gtcaatgtca tgccggccat ggacaaacaa ccaaacatca	7920
acgaatgtat tccatacaac tgcttggata acccagaatgt tgaagtactt ggtggagaa	7980
gcattgaac ccgttactactt cccatcgaca ttcccttgc ttgcacacag ttctgtctca	8040
gcgagttcgcc gccagggtgtt gggttcttgc ttggactatgt tgacatcatc ttgggtatct	8100
tttggccatc tcaatggat gcatcccttgc tgcaatgttgc gcaatgttgc aaccagagga	8160
tgcaggttgcggttccatc tcaatggat gcatcccttgc tgcaatgttgc gcaatgttgc aaccagagga	8220
aaatctatgc agagagatcc agagatgtggg aagccgtatcc tactaaccat gcttcccg	8280
agggaaatgcg tattttatcc aacgcacatgtt gaccacatgtt atcccatatgtt	8340
tcgcagttca gaaatccaa gttccctctt tgccgttgc ctttgcgttgc gctaatttttcc	8400
acctcagcgtt gttccggatc gtttgcgttgc ttggccaaatgttgc gatgtgttgc	8460
ccatcaatag ccgttacaaac gacccatgttgc ggttgcgttgc aaacttacaccat gaccacgttgc	8520
tttgcgttgcggttccatc tcaatggat gcatcccttgc tgccgttgc ttgggttgc ttgttgc	8580
gatataaccca gtttggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	8640
cgaatcatgttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	8700
atactaaatccaa gtttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc ttgttgc ttgttgc	8760
aaggctccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	8820
atgttcacatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	8880
tcagcggccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	8940
aaatgtatccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	9000
gaagacccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	9060
tcgcctatgg ttttttttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	9120
attcccttgcgatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	9180
acagggttgcgttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	9240
tcagatgttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	9300
ccgatagttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	9360
tttcaggaccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	9420
tttcagaatag ttttttttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	9480
gatgttgcgttgcgttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	9540
catccatcatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	9600
ggatgttgcgttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	9660
gtgtttagaaa ttttttttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	9720
ttactgcacatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	9780
tctttacatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	9840
aagtgtccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	9900
tctccggatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	9960
ccaaatccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	10020
ccatccaaatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	10080
acgatgttgcgttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	10140
ccaggtatcatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	10200
ggatgttgcgttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	10260
ctggccatcatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	10320
ggatgttgcgttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	10380
acttcttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	10440
tcatcttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc ttgttgc	10500
aagagaaatccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	10560
gggacaaatccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	10620
ccgttgcgttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	10680
ccatgtatccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	10740
tgtccgtatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	10800
ccgcatttcatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	10860
tcagctgttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	10920
tcttgggttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	10980
gaggcttccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	11040
tccacgatcatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	11100
tctatccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	11160
gtgcctatccatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	11220
cctccgtatc tcaatggatggaa gatgttgcgttgc ttgcgttgc ttgttgc ttgttgc	11280

---

-continued

---

```

acagaggta cagggactac acaccactc cagttggcta tggatccaag gagcttgagt 11340
actttccgta gaccgacaaa gtgtggatcg agatcggtga aaccgggaa accttcatcg 11400
tggacagcgt ggagcttc ttgtgggg aataatgaga tcccgccct tgcgttcaat 11460
tttgaggcct ttttactgaa taqtatgtt gtactaaat gtatgttga atagctata 11520
gtgagcggg  aaagtatcg  gctatttaac tatgacttga gtcctatct  tgaataaata 11580
aatcagcata tgatgtttt gtttggta ctccaactgt ctgcgttagt aatttgat 11640
ggttggcact tggcacgtat aataatgtt aagtaattt ctgcgttagt aaattaacta 11700
gattagatgtt gttgttata tacaaaaggc attaaatcg atacatctt gacaattgt 11760
cacggcttac cagaaaagaa attgcattt ttttgggtc ttcagactg acaagatcg 11820
tctgaqgtct aacaacttctt aagggatctt atgttgcata gtcctgcac aatattgtt 11880
tgacactcg  ccggggggc  cgcatcgatc gtggaaatgtt tcattctaa  ccccatgg  11940
acgtgaatgtt agacacgtcg  aaataaaggat ttcggattha gaataattt tttattgtt 12000
tcgcctataa atacgacgg  tcgtatattt tgcgttattt aaaaatgttact ttcattttt 12060
aataaacgtt cggacatctt catttttgc  ttggaaaaaaa atttggtaattt actctttttt 12120
tttctccata ttgaccatca tactcttgc  tgatccatgtt agatttccc  gacatgaa  12180
catcaaaag taggactaat ttggaaagc  aagctaattt aagaaatgtt aggacacgtt 12240
agtgtgagac acgtgttgc  cgccgttactt gcccactactt aaccacacaa gtgcacttc 12300
tgcgaaatgtt gttttttttt taatgttgc  cgccgttataa aagaaggat agagatgaa 12360
gaaaaaaacac  aaaaatataca  ttggatccat  aagacaaa  gtggaaaggc  gcaaaacgca 12420
acatttagaa  tcatttcttc  cctcaatttcc  ctttttcaat  ttccccctt  actaaatatt 12480
ctcctcttgc aattataaag cctcttatga caatgacaa  cttttttttt  12540
aactttagatcg tcaactgtctt ttaatataat ttctcttctt atcttattatg aatattcaat 12600
acaagaaataa tggccatattt ccagggtttt ccagggttgc  cttttttttt  12660
cccaggactt aagccaaagg  aacccttgc  aaaaatgttactt tttttttttt  12720
tacatttata caggggtttt tccctcgaa  aaatacaca  gtcctggcaa  aaaaaagagc 12780
ggggaaatgaa tttaaaaaca gcatgttgc  ttccacacacg  caaacaacacg  ggtatgcct 12840
cgttttctgtt aaggtgtacg  gaattttcc  aataatgttactt aacgcacatgac  catgcacttc 12900
aaaaaggtgtt gggcccgaga  ctgttgcacgg  ttggatccat  ttggccaaattt 12960
ctggcaaaaaa ggaatccctt ctttcttgc  tacaccgttc  tgcttataat  gtcgttacta 13020
ggagggttgc  ctttgcattt  ttgttttttgc  gggggcattt  cctgttgc  ttccctgggt 13080
tttttttacac tataatagcc  aacccgttgc  ttatctcttgc  gtcgttgc  ttgttttttgc 13140
aacttagaaaa aattttccgtt ttccattttc  atccatccca  gtttcattttc  agtccatttt 13200
cattcgttca  atacacttgtt tctataattt ggttacactt ttttcactt  ttatattttt 13260
ctgtttttat ttgttactac ttataatcat  aaatattttt tattgtatca  gtttccaaat 13320
tttgcctcttgc  ctgttgcctt  ttgttctcttgc  aatttttttgc  ttatcttgc  aacaatgtt 13380
taatttttttctt acttataattt ttagatattt  gatgtttataat  tataatgtt  tataatattt 13440
catgtatgtt  caaagaaaaat  atgttgcattt  tttttttttt  13500
tttatataat ttgcgttataat ttgttgc  aaaaatgttgc  aatttagaaac  tttttttttt  13560
ttttgtttaa taaaactgtt  taattttgc  tgatgttgc  aataaaaactt  ttatataat 13620
actgttttttataatataat  gatgttaca  ttggatccat  ttggatccat  ttggatccat 13680
tagaaaaaaat  gtgttttttataatataat  gatgttataat  aactgtgtca  gtacacgtc 13740
gcgttgcacaa  agtgtgc  tgccgttgc  agaagacagg  ggcttgc  gggatttttgc 13800
cagggttgc  tggccatggat  ttggatccat  tcaggccctt  gttttttttt  13860
atccctggca  aactgttttgc  cgatgggttttgc  ttggatccat  ttggatccat 13920
ccctggcaaa  cgtccatttc  ccagggttttgc  ttggatccat  ttggatccat 13980
caaacggatc  ttggatccat  agtggatttttgc  ttggatccat  ttggatccat 14040
atttttttttgc  ttggatccat  ttggatccat  ttggatccat  ttggatccat 14100
ttatcttataat  agaactgtt  aaaaatgttactt  tttttttttt  14160
atttttttataat  gtttttttataat  tttttttttt  14220
aaagtgtttt  gggagagaag  atagataat  tagacttttgc  cactcgaccc  tggttgc 14280
cttgagag  cccacggc  ttgttgc  ttggatccat  ttggatccat  ttggatccat 14340
aaaactaactt  ttactgttactt  tcacatttgc  caatgttgc  ttggatccat 14400
gttggatccat  ttggatccat  ttggatccat  ttggatccat  ttggatccat 14416

```

```

SEQ ID NO: 2      moltype = DNA  length = 14409
FEATURE          Location/Qualifiers
source           1..14409
                mol_type = other DNA
                organism = synthetic construct

```

```

SEQUENCE: 2
ggggaaatcc ctctttccata ttaagaacat  aaaaatcaac  agggaaaaata  agttcaccaa  60
cccgaaaccc  cacatccctt  atgaaaccc  cggggtaagg  agcacttctt  ttttccaaat 120
gaatcaccac  atctgttagat  tgcaagggtc  caagagatata  agaatttggaa  atggacagag 180
gggtgacact  aactgtgttttgc  ctttttttttgc  ttggatccat  ttggatccat  ttggatccat 240
taatgtcaagg  tataatgttgc  ttggatccat  ttggatccat  ttggatccat  ttggatccat 300
atttttttttgc  ttggatccat  ttggatccat  ttggatccat  ttggatccat  ttggatccat 360
tcctttttgc  ttggatccat  ttggatccat  ttggatccat  ttggatccat  ttggatccat 420
catcttagcag  aggtatgttgc  accttctactt  ttggatccat  ttggatccat  ttggatccat 480
cttcttccat  ttggatccat  ttggatccat  ttggatccat  ttggatccat  ttggatccat 540
tgccgttgc  ctttttttttgc  ttggatccat  ttggatccat  ttggatccat  ttggatccat 600
tttcttttttgc  ttggatccat  ttggatccat  ttggatccat  ttggatccat  ttggatccat 660
tcagggttgc  ttggatccat  ttggatccat  ttggatccat  ttggatccat  ttggatccat 720
atgacactca  ctttttttttgc  ttggatccat  ttggatccat  ttggatccat  ttggatccat 780
gaatgttttgc  ttggatccat  ttggatccat  ttggatccat  ttggatccat  ttggatccat 840
agggttgc  ttggatccat  ttggatccat  ttggatccat  ttggatccat  ttggatccat 900
tctaaggaa  ttggatccat  ttggatccat  ttggatccat  ttggatccat  ttggatccat 960

```

-continued

---

tgttgctgct	gtattggagg	aggaacata	at	ggtttgctt	gaccagca	attctggaaa	1020
ggaggggacag	actgttgtt	tttgtaaggg	cttgc	ccatc	tcatatttgg	atgat	tttc
caacctggat	tgtatgtt	gcttggaga	tcataattt	tttgat	tttgat	ttgg	tttgc
tgttgagggg	gtcttata	aatgtttca	gcataagct	cagg	tgttc	attgact	cca
gattactca	aagaaggaca	aagatctta	tggatctg	caga	agaaca	tata	accacag
actcttgtaa	cagg	tgcaaa	tttctgat	atg	gca	act	gg
gcatcaagg	ttcc	ctcaag	tttttattt	tc	atg	atc	tttgc
tcatcgact	ctct	taaggac	aatagcat	ttt	tttg	tttg	ggagg
gccttcttct	caatcaa	att	cctagctca	gc	agg	gggt	ccacca
ctggcagcat	caatcata	cct	cttccat	ttg	cta	ctt	atattgaa
aggagtgtc	caga	aaatctg	tttgtgag	cag	tttgc	aca	attttt
cagtactcat	aca	agctctc	tcc	act	tgct	tttgc	tttgc
gcagtgtcc	tagatgc	agg	aaattt	tcc	aga	ccct	tttaag
ctgaaaatgg	ac	ctgggg	agg	tttgc	tttgc	tttgc	tttgc
tgaggaaag	c	tttt	tttt	tttgc	tttgc	tttgc	tttgc
caaaacatat	gqa	actc	tttgc	tttgc	tttgc	tttgc	tttgc
ttgggttagc	aat	gttatt	tcc	act	atg	aca	cccc
tgaatgcaca	ag	tttt	tcc	at	gtt	tttgc	tttgc
cgagggtgg	gtt	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc
gaatattcag	aat	tttt	tttgc	tttgc	tttgc	tttgc	tttgc
taatgattag	gat	gcac	act	gct	atc	atc	tttt
tcaagggtt	ataa	atc	cc	tag	atg	atg	atc
tgttctcaa	ca	agg	cc	gg	gg	gg	gg
atgagttgaa	at	ttt	tttgc	tttgc	tttgc	tttgc	tttgc
aaacgaaaat	tca	aaat	gtc	tttgc	tttgc	tttgc	tttgc
aataaatactt	gaaa	aaaa	acat	gtt	tttt	tttt	tttt
ctocaacccg	ctaa	aaac	gggg	tttgc	tttgc	tttgc	tttgc
caatatgaga	aag	tttt	tttgc	tttgc	tttgc	tttgc	tttgc
ccacaccaag	tg	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc
ctatcaaac	aaa	aaac	agat	tttgc	tttgc	tttgc	tttgc
teactgttca	cag	ccaa	aaac	cc	tttgc	tttgc	tttgc
gcaactgaaa	tg	ccaa	aaac	tttgc	tttgc	tttgc	tttgc
cttggccca	ctg	cttccc	ccaa	cc	tttgc	tttgc	tttgc
ataaacatta	aa	at	tttgc	tttgc	tttgc	tttgc	tttgc
gtcaaccaat	gtt	cata	aa	tttgc	tttgc	tttgc	tttgc
ttgtttttgt	at	tttgc	aa	tttgc	tttgc	tttgc	tttgc
gttatttccc	ctt	gatt	ca	tttgc	tttgc	tttgc	tttgc
accagttcaa	cc	actt	aa	tttgc	tttgc	tttgc	tttgc
ttgtctttat	at	tttgc	aa	tttgc	tttgc	tttgc	tttgc
agcatcaaca	taa	tttgc	aa	tttgc	tttgc	tttgc	tttgc
tagcaacaaa	ta	ca	tttgc	tttgc	tttgc	tttgc	tttgc
acaaaacctc	aa	ag	tttgc	tttgc	tttgc	tttgc	tttgc
cttccattaa	tc	at	tttgc	tttgc	tttgc	tttgc	tttgc
gtcatgaa	act	tttgc	aa	tttgc	tttgc	tttgc	tttgc
aaaactccct	aa	act	tttgc	tttgc	tttgc	tttgc	tttgc
aatcttattt	c	at	tttgc	tttgc	tttgc	tttgc	tttgc
agattggat	aa	at	tttgc	tttgc	tttgc	tttgc	tttgc
taaaatgtc	tt	tttgc	aa	tttgc	tttgc	tttgc	tttgc
aattcttcg	aa	tttgc	aa	tttgc	tttgc	tttgc	tttgc
tgcataattt	at	tttgc	aa	tttgc	tttgc	tttgc	tttgc
agaaatagga	ga	ata	tttgc	tttgc	tttgc	tttgc	tttgc
ttctttaacc	tg	gac	ac	tttgc	tttgc	tttgc	tttgc
atatgtatgg	gt	cac	at	tttgc	tttgc	tttgc	tttgc
cacaaaataa	aa	aca	at	tttgc	tttgc	tttgc	tttgc
agtcacca	ttt	gtt	tttgc	tttgc	tttgc	tttgc	tttgc
aacactagat	ttt	gtt	tttgc	tttgc	tttgc	tttgc	tttgc
attagtattt	ttt	gtt	aa	tttgc	tttgc	tttgc	tttgc
taaacatgt	ttt	tttgc	aa	tttgc	tttgc	tttgc	tttgc
gataactctat	at	at	tttgc	tttgc	tttgc	tttgc	tttgc
ttaagtttgc	g	act	tttgc	tttgc	tttgc	tttgc	tttgc
ttgttactt	gt	at	tttgc	tttgc	tttgc	tttgc	tttgc
gettgtatag	ta	ac	tttgc	tttgc	tttgc	tttgc	tttgc
tggaaagtaca	ttt	tttgc	aa	tttgc	tttgc	tttgc	tttgc
tagccaaata	tg	cc	aa	tttgc	tttgc	tttgc	tttgc
atatttaggg	tc	gt	tttgc	tttgc	tttgc	tttgc	tttgc
tccaaacgtt	ttt	tttgc	aa	tttgc	tttgc	tttgc	tttgc
acaaaacact	aaa	act	tttgc	tttgc	tttgc	tttgc	tttgc
aaaaagtaa	a	ata	tttgc	tttgc	tttgc	tttgc	tttgc
taactctat	ttt	tttgc	aa	tttgc	tttgc	tttgc	tttgc
tactctat	ttt	tttgc	cc	tttgc	tttgc	tttgc	tttgc
tggacttttgc	ttt	tttgc	tttgc	tttgc	tttgc	tttgc	tttgc
aatgttccca	tc	at	tttgc	tttgc	tttgc	tttgc	tttgc
cgaggccgt	cc	cg	ac	tttgc	tttgc	tttgc	tttgc
tgcacccagt	gg	cat	tttgc	tttgc	tttgc	tttgc	tttgc
egtc当地	tt	tc	aa	tttgc	tttgc	tttgc	tttgc
tctaggtcgc	tc	tc	aa	tttgc	tttgc	tttgc	tttgc
gccccctag	ttt	tttgc	aa	tttgc	tttgc	tttgc	tttgc

-continued

---

tatattttgg	gattaaatta	tataggaatt	agtaattttt	ctctcttatt	tcttcctttt	5580
tgttcaata	atggaaatc	taacatcatt	taagttttt	tgttagaaaat	attaaaagtt	5640
gatgaattha	tgatacttag	tgaataatta	gagtagaaaaa	ataaagtaaa	gccccaaaaaa	5700
gaaaatttgt	gatatgaaga	tacatgcct	gcatgccccc	ggcacgotta	ggtcaaacac	5760
tgatagttt	aactgaaggc	gggaaacgac	aatctgtcc	ccatcaagct	tgatatcgaa	5820
ttctctgacg	cggggggatc	cactagtct	agagcggcc	cgtaaactgc	aggcgacgg	5880
atccccgggt	accggactcg	aattcaaaatt	tattatgtgt	ttttttccg	tggtcgagat	5940
tgtgtattat	tcttagtta	ttaacaagact	tttagctaa	atttgaaga	atttacttta	6000
agaaaatctt	aacatctgag	ataatttcag	caatagatta	tattttcat	tactctagca	6060
gtatatttgc	agatacattc	caacatata	ggttgttaga	aaaatgcac	tatatatata	6120
tatatttattt	tttcaattt	aagtgcatg	tatataatata	atataatata	atataatgtgt	6180
gtgtgtatat	ggtcaagaa	atttttatac	aaatatacac	gaacacatata	attttgacaaa	6240
atcaaagtat	tacactaaac	aatggatgg	tgcattggcca	aaacaaatata	gtagattaaa	6300
aattccagcc	tccaaaaaaaaa	aatccaagtg	ttgttaagca	ttatataatata	atagtagatc	6360
ccaaatccat	gtacaattcc	acactgatc	aatttttaaa	gttgaatata	tgacgttaga	6420
ttttttat	gtcttacctg	accattact	aataacattc	atacgtttc	attttgaataa	6480
tcctctataa	tttatattgaa	tttggcacat	aataagaaac	ctaattgggt	atttttttta	6540
ctagtaattt	tctggtgatg	ggctttctac	tagaaagctc	tcggaaaatc	ttggaccaaa	6600
tccatattcc	atgacttcg	ttgttaacc	tattgtttt	cacaaacata	ctatcaatata	6660
cattgcaacg	gaaaaggtagc	aagtaaaaaca	ttaaattcga	tagggaaatg	atgttagagg	6720
tttgggaagac	aggcccagaa	agagattat	ctgacttgtt	ttgtgtatag	tttcaatgt	6780
tcataaagga	agatggagac	ttgagaagtt	ttttttggac	tttggtttgc	tttggggc	6840
gttttttttt	ttgtatattc	actttttgtt	gctttatgtt	tgtatattt	tcgtggactc	6900
tttagttttt	tttagacgtc	taacttttt	gggtttatgt	tttggttttaa	cataattgtaa	6960
cagatgactt	gtgtgcgac	taattttt	acattaaaca	tagttctgtt	tttttgaatgt	7020
tcttattttc	atttttttt	gaatgttt	tatttttctt	tatttataat	tcttagtaaaa	7080
ggaaaattttt	gtttttttat	gaaaatataa	tatattccac	agtttccact	aatttttatgc	7140
atttagcagt	acaattttca	aaattttcca	ttttttttca	tgaatttatac	cattatataat	7200
taactaaatc	caaggtaaaa	aaaaggatgt	aaagctctat	agtaagtaaa	atataaattc	7260
cccatataagg	aaggggccaag	tccaccaggo	aagtaaaaatc	agcaagcacc	actccacccat	7320
cacacaatata	cactcataga	taacgatata	atttcatggaa	ttatcttca	cgtggcattt	7380
ttccagcggt	tcaagcgat	aagggtctca	acacctctcc	ttaggcctt	gtggccgtt	7440
ccaaataaaa	ttaacctcac	acatatccac	actccaaaatc	caacgggtta	gatcttagtc	7500
caacttgaatc	tcatgtatcc	tagaccctc	gatcaactca	aaagcttgc	tcattttgtt	7560
tatcattata	tatagatgac	caaaggacta	gacccaaatc	cagtcaacaca	aagagtaaag	7620
aagaacaatg	gtttcttca	tgctcttcc	cgctactatg	tttgccttc	cggctcaggc	7680
caactatggc	gtcttttca	acggactaa	gtctccgct	gccttccag	ccacccgca	7740
ggctaaacac	gacattactt	ccatcacaag	caacggcgga	agagttact	gcatgcagg	7800
gttgcctccg	atggaaaaga	agaagtttg	gactctctt	taccccttc	accccttccg	7860
ttccgggtgt	cgcgtaact	gcatggcgg	catggacaaa	aacccaaaca	tcaacgaatg	7920
cattccat	aactgttgc	gttaaccaga	agttgaatgt	tttggggag	aacgcatgt	7980
aaccgggtac	actccatcg	acatcttctt	gtcttgcata	cgagttctc	tcagcgagtt	8040
cgtgcggat	gttgcggatc	ttcttgcgtt	atcttgcata	ttcttgcgtt	tttgcgttcc	8100
atcttcaatgg	tgatgttcc	ttgttgcata	tgatgttgc	tttgcgttcc	tttgcgttcc	8160
gtttcgccagg	aaccaggcca	tcttcgttgc	ttgttgcata	tttgcgttcc	tttgcgttcc	8220
tgcagagagc	ttcagagatgt	ggggagccga	ttcttgcata	tttgcgttcc	tttgcgttcc	8280
gegttattca	ttaacacgaca	tgaacagcgc	tttgcgttcc	tttgcgttcc	tttgcgttcc	8340
ccagaactact	caaggttctc	tctttgcgtt	tttgcgttcc	tttgcgttcc	tttgcgttcc	8400
cgtgcgttca	gacgtttagc	ttttttggca	tttgcgttcc	tttgcgttcc	tttgcgttcc	8460
tagccgttac	aaacgaccta	ctaggctgt	tttgcgttcc	tttgcgttcc	tttgcgttcc	8520
gtacaacact	gttgcgttcc	tttgcgttcc	tttgcgttcc	tttgcgttcc	tttgcgttcc	8580
ccagttcagg	agagaatgt	cccttcacat	tttgcgttcc	tttgcgttcc	tttgcgttcc	8640
tgtactccaga	accttccat	tccgttacat	tttgcgttcc	tttgcgttcc	tttgcgttcc	8700
cccaagttctt	gagaacttcg	acggtagctt	tttgcgttcc	tttgcgttcc	tttgcgttcc	8760
catcaggagc	ccacacttca	tggacatct	tttgcgttcc	tttgcgttcc	tttgcgttcc	8820
cagaggagag	tattactgtt	ctggacacca	tttgcgttcc	tttgcgttcc	tttgcgttcc	8880
ggcccggttt	accttttctc	tctatggaa	tttgcgttcc	tttgcgttcc	tttgcgttcc	8940
cgttgcctaa	ctaggtcagg	tttgcgttcc	tttgcgttcc	tttgcgttcc	tttgcgttcc	9000
cttcaatata	gttataaca	accaccaact	tttgcgttcc	tttgcgttcc	tttgcgttcc	9060
tggAACCTCT	tcttaacttgc	catccatgt	tttgcgttcc	tttgcgttcc	tttgcgttcc	9120
ggacggaaatc	ccaccacaga	acaacaaatc	tttgcgttcc	tttgcgttcc	tttgcgttcc	9180
gagccacgtg	tccatgttcc	tttgcgttcc	tttgcgttcc	tttgcgttcc	tttgcgttcc	9240
tcttatgttc	tcttggatac	atcgtatgt	tttgcgttcc	tttgcgttcc	tttgcgttcc	9300
tattactcaa	atccctgcag	tgaaggaaa	tttgcgttcc	tttgcgttcc	tttgcgttcc	9360
accaggatcc	acttgcggat	acccgtttag	tttgcgttcc	tttgcgttcc	tttgcgttcc	9420
tagagggtat	atggatgtt	caatttccat	tttgcgttcc	tttgcgttcc	tttgcgttcc	9480
tgtgagggtat	gtttctgtt	ccccttattc	tttgcgttcc	tttgcgttcc	tttgcgttcc	9540
cttctccaaat	acagttccat	ctacagctac	tttgcgttcc	tttgcgttcc	tttgcgttcc	9600
cggttactttt	gaaagtgcac	atcttcaatc	tttgcgttcc	tttgcgttcc	tttgcgttcc	9660
aaacttttagt	gggactgcag	gaggatgtt	tttgcgttcc	tttgcgttcc	tttgcgttcc	9720
aacactcgat	gttgcgttcc	acccgtttag	tttgcgttcc	tttgcgttcc	tttgcgttcc	9780
ctccaccaat	cagcttgcgt	tgaaaactaa	tttgcgttcc	tttgcgttcc	tttgcgttcc	9840
caacttggc	acttacccat	tttgcgttcc	tttgcgttcc	tttgcgttcc	tttgcgttcc	9900
gaaaggataaa	cacggccaagc	gttgcgttcc	tttgcgttcc	tttgcgttcc	tttgcgttcc	9960
caaaagacatc	aacaggcagc	cagaacgttgc	tttgcgttcc	tttgcgttcc	tttgcgttcc	10020
aggaggcgcac	gtatgtttca	aggagaacta	tttgcgttcc	tttgcgttcc	tttgcgttcc	10080

---

-continued

---

ctaccctacc	tacctgttacc	agaagatcga	tgagtccaaa	ctcaaagcct	tcaccaggta	10140
tcaacttaga	ggctacatcg	aagacagccaa	agacccctgaa	atctactcga	tcaggttacaa	10200
tgccaagcac	gagaccgtga	atgtccccagg	tactggttcc	ctctggccac	tttctgcucca	10260
atctccatt	gggaagtgtg	gagagcctaa	caagatgcgc	ccacacccct	agtggaaatcc	10320
tgacttggac	tgctctgtca	gggatggcga	gaagtgtgcc	caccattctc	atcacttctc	10380
cttggacatc	gatgtggat	gtactgaccc	gaatgaggac	ctcggagct	gggtcatctt	10440
caagatcaag	acccaagacg	gacacgcaag	acttggcaac	cttggatcc	tcgaagagaa	10500
accattggc	ggtaagctc	tcgtctgt	gaagagagca	gagaagaatg	ggaggggacaa	10560
acgtgagaaa	ctcgaatggg	aaactaacat	cgtttacaag	gaggccaaag	agtccgtgga	10620
tgttttgc	tgtaactccc	aatatgtaca	gttgcacccg	gacaccaaca	tcgcctatgt	10680
ccacggcga	gacaaacgtc	tcgcacat	tctgtggat	tacttgcctg	agttgtccgt	10740
gatccctgt	gtgaacgctg	ccatcttca	ggaaacttgg	ggacgtatct	ttaccgcatt	10800
ctccttgtac	gatgccagaa	acgtcatca	gaacccgtac	ttcaacaatg	gcctcagctg	10860
ctggaatgtg	aaagggtatc	tggacgttg	ggaacacgaa	aatcagcgtt	ccgtctcggt	10920
tgtgcctgag	tgggaagctg	aagtgttca	agagggttga	gtctgtccag	gtagaggctta	10980
cattctccgt	tgacggcctt	aaacgggg	atacggtgg	ggttgcgtga	ccatccacca	11040
gatcgagaac	aacaccgcg	agcttaagtt	cttcaactgc	gtcgaggaaag	aaatctatcc	11100
caacaacacc	gttacttgc	acgactacaa	tgtgaatcga	gaagagtacg	gagggtgcct	11160
cactagcgt	aacagagggt	acaacaaagc	tccttccgtt	cctgtctact	atgcctccgt	11220
gtacgaggag	aaatctaca	catatggcag	acgtgagaac	ccttgcgtat	tcaacagagg	11280
ttacaggag	tacacaccac	ttccagttgg	ctatgttacc	aaggagottt	agtacttcc	11340
ttagacccgac	aaagtgttgg	tcgagatcgg	tggaaacccgg	ggaaaccttca	tcgtggacag	11400
cgtggagctt	ctcttgcgtt	aggaataatg	agatcccgat	ccttgcgttcc	atttttgagg	11460
gtcttttact	gaataatgt	tgtatctaa	aatgtatgtc	gtatagtc	atagtgagcg	11520
aggaaagtagt	cgggctattt	aactatgact	tgagctccat	ctatgaataa	ataaatcage	11580
atatgtatgt	tttgcgttgc	gtacttca	tgtctgttca	gctaatttga	tatgttttgc	11640
acttggcacc	tataaatat	ctgaaatgt	ttactctgtt	gctaaattaa	ctagatttga	11700
ttagtgttatt	atataaaaa	ggcatttaat	catagatcatc	tttagacaat	tgtcacgtc	11760
taccagaaaa	gaaattgtat	ttgtttttgg	gtctttcaga	ctgacaagat	cgatctgaag	11820
tctaaacaat	tctaaagaggt	atcatgtac	aatgttccgc	cacaatattt	aattgtacgt	11880
cagccccggc	ggccgcacat	atcgtgaat	ttctcatgc	agggccattt	ttggacgtgaa	11940
ttagacacgc	tctaaataaa	gatttccgg	tttagataat	ttgtttatgt	ctttcgccta	12000
taataatcgac	ggatcgtaat	ttgtcggtt	atcaaatgtt	actttcattt	tataataacg	12060
ctgcggacat	ctacattttt	gaattggaaa	aaaatttggta	attactctt	cttttctcc	12120
atattgcaca	tctactatc	tgtctgtat	tgttagatcc	ccggacatga	agccatcaaa	12180
aagttaggact	aattttagaa	agcaagctaa	ttcaaaagaa	tgaaggcactg	tttagtgcgt	12240
gacacgtgtt	ggcgccgg	actgcacact	actaaccaca	caagtgcact	cagtgcgaa	12300
gttgcttaaa	aatthaatgtt	atccgcactt	ataaaagaag	gatagagatg	aaggaaaaaa	12360
cacagaaaaat	acaatccctt	atagaagaa	aaggctgaa	gaagcaaaacg	caaacattag	12420
aagtcttcc	ttccctcaat	ttcccttttc	aatttccctt	tttactaaat	atttcctct	12480
tgcattatata	aagccctcta	tgacaatgc	aaactttaact	ctccctttgtt	gggaacttat	12540
cagtcaactg	ctcttaatata	aatttcttctt	cctatctatt	atgaatattt	actacaagaa	12600
atatgtatgt	tttgcgggg	tttttgcac	gggacattaa	ccctggcaaa	tttcccgagg	12660
actaagccaa	ggaaaaaccct	ggccaaat	cattttggaa	gggtggggac	acttacattt	12720
acacagggg	tttgcctcg	caaaaatata	aaacccctgg	caaaaaaaag	acggggaaat	12780
gaattttaaa	acagcatgtt	gttttccac	agccaaacac	acgggtatgc	cctcggtttc	12840
tgtaaagotg	acggaaatctt	cccttcaatgt	aacacgcacat	gaccatgcac	tgcaaaaagc	12900
tgtgcggccc	ggacgtgaca	gggggtgttac	ccctcgaa	ttgggtgcac	ccccctggcaa	12960
aaaggatcc	ctgctttcc	agctacacgg	ttctgtctat	atagtgac	ctaggagggt	13020
agcctttgc	tctgtgtttt	tgccgggg	attccgtgat	ttatccctgt	ggttttttta	13080
cactatata	ccaaaccgcg	ttgtttatctt	catgtctact	tttgcgtt	tgaaaacttag	13140
aaaatatttt	gttttccatt	ttccatctca	ccatgttccat	tttgcgtt	tttgcgtt	13200
ttcatacaact	ttttttat	tttgcgttaca	tttttttccat	tttgcgtt	tttgcgtt	13260
tattttgttac	tactttat	cataaataattt	ttttattgtt	tcagtgccca	attttgcctc	13320
tcctctgtgc	tccttgcctc	ctgatattgt	tctcttta	ttcaacaatg	tagtaatattt	13380
tctacttata	attttagata	tatgtatgtt	atataatata	tttgcgtt	tttgcgtt	13440
tgtcaatggaa	aatatgtatgt	tttgcgttgc	atatgtgtt	ataatataat	atgttttat	13500
atatttcgaa	ttttgttgc	aataaaaactg	ttaatttgc	aactgtataa	ttttttgtt	13560
taataataact	gtttatattt	gtatgtatgt	ttaataaaaa	ctgtttat	aaaactgtt	13620
atataataat	tatgtatgtt	acattttat	aaactgtttat	aaaacagt	tttgcgtt	13680
aatgttaaaa	cttagagaaaa	aaatgttataa	taaaaactgtt	tcgtacacgc	agcgcgtcag	13740
aaaagtgtgc	agatgcgtca	tggtgaaagac	aggggctaa	acagggat	tgacaggaa	13800
ttttgcgg	gattttgc	gggtcagccc	ctcggtttt	tgccagggg	gaaatccctg	13860
gcaaaactgt	tttgcgttgg	cggttttccc	agggtttcc	ccccctggca	aatccctggc	13920
aaacgtccat	ttcccggg	ttttttttcc	ttttccagggg	aaatccggcc	ttggcaaacga	13980
gtcttttctt	tgtgtgtt	acttttgc	tagttttcc	tgtattttat	tttgcgtt	14040
atggcttgc	taccatatttgc	cattataat	tttgcgttgg	tttgcgtt	tttgcgtt	14100
ctaatagaac	tggaaaagag	tatataata	acttcatc	tttgcgtt	tttgcgtt	14160
ttagcttatt	atatatctt	attattatgt	tttgcgtt	tttgcgtt	tttgcgtt	14220
tttggggag	aatatgtatgt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14280
gagccccagaa	ggctgtgtac	ggcccttta	cctccacca	tcagttggc	tttgcgtt	14340
acgttactga	ctatcacatt	gaccaagtgt	ccaaatgtt	cacccat	tttgcgtt	14400
tctgaagggg						14409

-continued

---

source	1..14436
	mol_type = other DNA
	organism = synthetic construct
<b>SEQUENCE: 3</b>	
ggggaaaatcc ctcttcata ttaagaacat aaaaatca aaaaaaaaata agttcaccaa	60
ccccgaaccag cacaatctct atgaaacctg cggggtaacg agcacttcta ttttccaat	120
gaatcaccac atctgttagat tgcaaaaggc caagagataa aatggaaaa atggacagag	180
gggtgcacact aactgtatct cctagatcta gcattgtatt atcaaattt ctgttccaa	240
taatgcagg tatacagaaa gtacgtggg ctttacattt ctcaggatg taaggacaa	300
atttacccat catgtcgac acatcttgc ccatgtata ctttcatgt ctgttgcgt	360
tcttttggt ggtgcacaaat ctttttagaa acttgacaca tcttggaaatc tgcttgatgg	420
catctagcag aggtatgttc accttctactt tcttggaaatc tccttgcgt tcctttctg	480
cttcttccat ttttttggt tggaaattgtc caaggttggg atggaaagg gataagaggc	540
tgccgtaatg cagaattact agaagaaggc ccacctgtat gaaaattttt gttaggaagc	600
tttcttccat gtgcacat ctcacccat ttttgcgtt tagaaatgg cttgacagg	660
tcaggtgcgg gtgcgtctac tgggtggggt acttgcattt gggtgtcaga cctcaagggt	720
atgacactca cattttcgg atttgcaca gtttgcgtt gcaatttgc agaatttgg	780
gaatgagott gtttcaactg agtagccat cggcccatct gatttgcgt actctgtat	840
aaggcttgg tcttgcgtt aaatggcata ttttggatgg tcttgcgtt cactaactct	900
tctaaggaaag gttaaggagg agtctcgtt gtttgcgtt ttttgcgtt ctgttgcgt	960
tgttgcgtt gtattggagg aggaacat ttttgcgtt gtttgcgtt gtttgcgtt	1020
ggaggggacag actgttggt ttttgcgtt ctttgcgtt ttttgcgtt atgatttgc	1080
caaaccttggat ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	1140
tgttgcgtt gtcttgcgtt aatgttgcgtt gtttgcgtt ttttgcgtt	1200
gattactgcg aagaaggaca aagatctgtt ttttgcgtt ttttgcgtt ttttgcgtt	1260
actcttgcgtt ctttgcgtt ttttgcgtt atggcaagctt gagttactgtt gtttgcgtt	1320
gcattcaatgtt ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	1380
tcatcgactt ctttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	1440
gccttcttctt caatcaattt ctttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	1500
ctggcagcat caatcatactt ctttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	1560
aggagttgtgtc ctttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	1620
cgtactcat acaagtcgtt ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	1680
gcagtggcc ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	1740
ctgaaaatgg accttggggc aaggttagt tagccatctt ttttgcgtt ttttgcgtt	1800
tgggggggggg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	1860
caaacaatattt gtttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	1920
tttgggttgcgtt aatgtttagt ttttgcgtt ttttgcgtt ttttgcgtt	1980
tgaatgcaca agtttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	2040
cgggggggggg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	2100
gaatatttcgtt aatgtttagt ttttgcgtt ttttgcgtt ttttgcgtt	2160
taatgtttagt ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	2220
tcgaagggtt ataaatcacc ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	2280
tgttcttcaatgg aatgtttagt ttttgcgtt ttttgcgtt ttttgcgtt	2340
atgagttggat ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	2400
aaacggaaaat ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	2460
aataataactt gtttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	2520
cttcaacccgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	2580
caatatgttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	2640
ccacaccaatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	2700
ctatcaaaatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	2760
tcactgttcaatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	2820
gcaactgttcaatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	2880
ctttggcccaatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	2940
ataaaccatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3000
gtcaaccatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3060
tttgggggggggg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3120
gttattttccatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3180
accagggttcaatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3240
tttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3300
agcatcaaca ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3360
tagcaaccaatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3420
acaaaacccatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3480
tttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3540
gttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3600
aaaactccatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3660
aatcttccatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3720
agattggatgtt ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3780
taaaattgttcaatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3840
tttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3900
tgcataatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	3960
agaaatgttcaatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	4020
tttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	4080
atatgttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	4140
cacaaaatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	4200
agtccaaatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	4260
aaacactatgttcaatgg ttttgcgtt ttttgcgtt ttttgcgtt ttttgcgtt	4320

---

-continued

---

attagtcat tttggaaagt aaagtggaca aatttgagat ttcatccctt aatgacatta	4380
taaacatgtt tttttccat gaccctttt caatgttaact acaattttt ccttagtta	4440
gatactctat atatgcatgt tacgttagttt atgaaaacat acctaagtgt ttgtgtatgg	4500
ttaagttgc gactacctt gatatcaaa tcctcatctc caatctcata caaaagatcac	4560
ttgtcacttg gtacctgaac ctgtcgtt tgcaaggatgt agttttctt gaagccacac	4620
gettgtatag taaccagaag ccaggaggga gtcctctaag gctctaactc gtatttccg	4680
tggaaagtaca tttttttct taaagaaaac agagatgtt aatatttctt	4740
tagccaaata ggaccatcat gaaaaacaaa actttcttc taagtattt atgcaactac	4800
atattttaggg tgcgtttgtat cgctaaaaa ataagggtct agacaacaca aaaatatttt	4860
tccaacgtttt gattttaaaat atggctgaga gacaatacaa aataaggaaat gatgaactgg	4920
acaaaaacactt aactcaatgt aactcatca aactttttt cagtgcttaa	4980
aaaaaaagtaaa aatacaatat tattcttattttttt attatctcac accttcttc	5040
tactcatttgc tttcaatttca cctctccat gggcacctt gtttgcggc gagagtgcgt	5100
tggacttttgc tttttttctt tttgtcattt atttcttctt ttttgcggc aattttatca	5160
aatgttccca tttcatctt actctttttt gttatgtttt tttttttttt ccaactccaa	5220
cgaggccctg cccgcaccac catcatcagc accttgcgc ggcctcacgc cgcaaggccc	5280
tgcacccagg ggcacatagg gtcacgcctc ctctttaaag gtgtctctt ttttttatgt	5340
cgtcataatgtt ttgtcatttca accttgcgtt ttccatgaa tccctttact tttttttttt	5400
tctagggtcgc tctggccgggt tccaaaccctt gccccaaaaaa aatgttgcgtt ggttagggaaag	5460
gcggcccttag tttgtatcaa aataatcat gctaagatat tgataactgc tatgtatagg	5520
tatattttgg gattaaattna tataggaatt agtaattttt ctcttcttattt tttttttttt	5580
tgttcaataa atggaaatc taatcatattt taagtttttt tgtagaaaaat attaaaatgtt	5640
gatgaaatttt tgataacttag tgaataattna gatgtttttt gccccaaaaaa	5700
aaaaattttgtt gatgttgcgtt gatgttgcgtt gatgttgcgtt gatgttgcgtt	5760
caaaactttcc taaattttttt ctactttttt atgtttttt aatgttgcgtt gatgttgcgtt	5820
ctgtatccca tcaatgttgc tategttgc ctgcagccgc gggggatccac tagttctaga	5880
gcggcccgctg ttaacttgcagg tcgacggatcc cccgggttacc gatgttgcgtt tcaatgttgc	5940
tatgtgtttt ttttccgtgg tcgatgttgcgtt gatgttgcgtt ttagtttataa caagactttt	6000
agctaaaaat tggaaatgtt tttttttttt tttttttttt tttttttttt tttttttttt	6060
tagattatat ttttccattttc tttttttttt tttttttttt tttttttttt tttttttttt	6120
tgttagaaaa aatgttgcgtt atatataat tttttttttt tttttttttt tttttttttt	6180
ataatataatata tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6240
tatacaccggaa cacatattttt tttttttttt tttttttttt tttttttttt tttttttttt	6300
atggccaaaaa caaatatgtt gatgtttttt tttttttttt tttttttttt tttttttttt	6360
taaagccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6420
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6480
aaatccatata tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6540
aagaaacactt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6600
aaagcttccgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6660
tagttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6720
aaatccatata tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6780
acttgcgtttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6840
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6900
tatgtttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	6960
ctttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7020
ttaaacatag tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7080
ttttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7140
attccacat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7200
tttattttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7260
gctctataat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7320
taaaatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7380
catggat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7440
cctctccctt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7500
caaaatccaa tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7560
cacttccaaat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7620
caaaacccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7680
taactatgtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7740
cttgcgtttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7800
cgccggaaaga gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7860
tctctttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7920
ggacataatcc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	7980
tgaagttactt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	8040
tttgacacat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	8100
tgtttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	8160
gcagtttgc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	8220
aggattttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	8280
taactacccaa tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	8340
gaccacat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	8400
cgatccat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	8460
gtggggat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	8520
aaactacacc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	8580
tgattttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	8640
ggacattttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	8700
ccaaacttacc tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	8760
tggttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	8820
cagcataact tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt	8880

-continued

---

catggcctct	ccagttggat	tcagggggcc	cgagtttacc	tttcctctct	atgaaactat	8940
ggaaaacgc	gtccacaac	aacgtatcg	tgtcaacta	ggtcagggtg	tctacagaac	9000
cttgc	acccgttcc	gaagccctt	caatacggt	atcaacaacc	agcaactttc	9060
cgttcttgac	ggaacagagt	tcgcctatgg	aaccttctt	aacttgocat	ccgctgttta	9120
cagaaagagc	ggaaccgtt	attccgttga	cgaaatccca	ccacagaaca	acaatgtgcc	9180
acccaggca	ggatttcc	acagggttag	ccacgtgtcc	atgttccgt	ccggatttc	9240
caacagt	gtgagcatca	tcagagctcc	tatgttctc	tggatacatc	gtatgtgtca	9300
gttcaacaac	atcatcgat	ccgatagat	tactcaa	cctgcagtg	agggaaactt	9360
tctcttcaac	ggttctgtca	tttcaggacc	aggattact	ggtggagacc	tcgttagact	9420
caacagcgt	gaaataaca	ttcagaatag	agggtatata	gaagtccaa	ttcacttccc	9480
atccacatct	accagatata	gagttcggt	gaggtagtgc	tctgtgacc	ctattcac	9540
caacgttat	ttgggttatt	catccatctt	cttcaataca	gttccagcta	cagctactc	9600
cttggataat	ctccatcca	gcatgttcc	ttactttgaa	agtgccaat	cttttacatc	9660
ttcaactcggt	aatccgtgg	gtttagaa	ctttagtgg	actgcaggag	tgattatcga	9720
cagattcgag	ttcattccag	ttactgcac	actcggaggct	gatgtacaacc	ttgagagagc	9780
ccagaaggct	gtgaacgc	tcttac	ccaccaatcg	cttgggttga	aaactaacgt	9840
tactgactat	cacattgacc	aagtgtccaa	cttggtacc	tacccatcg	atgagttctg	9900
cctcgacgag	aagcgtgaac	tctccggagaa	agttaaacac	gccaagcg	tcagcgacga	9960
gaggaatctc	ttcagaact	ccaaacttca	agacatcaac	aggcagccag	aacgtgggt	10020
gggtggaa	accgggatca	ccatccaagg	aggcgcacgt	gttcaagg	agaactacgt	10080
caccctctcc	ggaacttgc	acgagtgtca	ccctac	ttgttacc	agatcgatga	10140
gtccaaactc	aaagcattca	ccaggatata	acttagaggc	tacatcg	acagccaa	10200
ccttgc	aaatctc	tactcgat	ggtacatgc	caagcacgag	accgtgtat	10260
ttgttccctc	ttggccat	tcgc	ccatgggg	aaatgtgg	agccta	10320
atgcgttca	cccccttgc	ttactgcac	tttgcactgc	tcctgcagg	atggcagaa	10380
gttgtcc	catttc	acttcc	ggacatcgat	gttggatgt	ctgac	10440
tgaggac	ggagtgtgg	tcat	ttca	gatcaaggac	caagacggac	10500
ttggca	ccat	ttcg	tttgcgtt	gaagcttgc	ctcg	10560
gagagc	aaagaatgg	gggacaa	tttgcgtt	gaatgg	ggaa	10620
ttacaaggag	cccaaa	ggatgt	tttttgcgt	aactcc	atgatcgat	10680
gcaaggc	accacatcg	ccatgtca	ccgcgcagac	aaacgcgt	acagoatcg	10740
tgaggc	tttgcgt	tgccgt	ccctgg	aacgcgtc	tcttcgagg	10800
acttgg	ctgtat	ccgcattc	tttgcgt	gccagaa	ac	10860
cggt	aaaaat	ccat	tttgcgt	ggtcat	gtcg	10920
acaga	aaat	ccat	tttgcgt	gcctg	gtgg	10980
gggtt	ggatgt	tttgcgt	tttgcgt	accgc	ttaagg	11040
cggt	gggt	tttgcgt	tttgcgt	cgaga	accgc	11100
caact	gggt	tttgcgt	tttgcgt	accgc	accgc	11160
gaat	gggt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	11220
ttcc	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	11280
tgaga	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	11340
tgta	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	11400
aacc	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	11460
ttcc	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	11520
gtat	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	11580
gtcc	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	11640
ctgtt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	11700
ctctg	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	11760
ataca	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	11820
tttc	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	11880
gtc	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	11940
tcat	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12000
gaata	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12060
aaaat	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12120
attgg	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12180
agatttcc	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12240
aagaaatgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12300
aacc	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12360
aaaga	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12420
gtc	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12480
ttcc	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12540
ctaa	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12600
atct	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12660
atatt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12720
tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12780
gccttgc	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12840
caa	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12900
acg	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	12960
tcgg	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	13020
tgct	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	13080
ccgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	13140
gctc	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	13200
gttc	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	13260
tttc	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	13320
tattgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	13380
ctta	tttgcgt	tttgcgt	tttgcgt	tttgcgt	tttgcgt	13440

-continued

---

tatatgtatgt tataattttg catgatctgt caaagaaaat atgatgttc tacttgcatg	13500
atgtgttata atatatgtt tttcgaattt tggtttaat aaaactgtt	13560
aattagaaac tgtataattt ttttggtaaaaactgtt taatgttgc tgatctgtt	13620
aataaaaactg ttatataaa actgtttata tataatataat gatgttaca ttttaaaac	13680
tgtttataaa acagtttagt tagaaaaat gttaaaacta gagaaaaaaaaa ttttataaa	13740
aactgtgtca gtacacgcgc gcgtcagaaa agtgtgcaga tgcgtcagtg agaagacagg	13800
ggctaagaca gggattttga cagggaattt tgccaggat ttggccagg tcagccctc	13860
gttttttgc caggggtgaa atccccggca aactgttttg cgatgggcgt ttggccagg	13920
gatccagccc ctggcaaaat ccctggcaaa cgtccattt ccagggtctt ttgttcttt	13980
cccaaggaaat ccggccctgg caaaggagct tggttctgt agtgattact ttgcattag	14040
tttttctgtt atttaattt atgtttagt gtttgatc ccatttgcat tataagttt	14100
aggggttagcg ttgaaaatgtt ttttctcta atagaactgg aaaagatgtt ttttataact	14160
tcatcaactt ggatacattt attttattta gcttattata ttttcttattt attaatgttaa	14220
ttaactattt ttatctctgc aaagtggat gggagagaag atagataatg tagactctt	14280
cactcgaggc tgatgacaaac cttggagag cccagaaggc tggtaacgcc ctcttacact	14340
ccaccaatca gttggcttg aaaaactaacg ttactgacta tcacattgac caagtgtcca	14400
acttggtcac ctaccttagc gatgagttct gaaggg	14436
SEQ ID NO: 4 moltype = DNA length = 27	
FEATURE Location/Qualifiers	
source 1..27	
mol_type = other DNA	
organism = synthetic construct	
SEQUENCE: 4 tttaaactat cagtgttga cacacac	27
SEQ ID NO: 5 moltype = DNA length = 27	
FEATURE Location/Qualifiers	
source 1..27	
mol_type = other DNA	
organism = synthetic construct	
SEQUENCE: 5 ttcccccct tcagttaaa ctatcag	27
SEQ ID NO: 6 moltype = DNA length = 27	
FEATURE Location/Qualifiers	
source 1..27	
mol_type = other DNA	
organism = synthetic construct	
SEQUENCE: 6 ttccctaaat tagtcctact ttttgat	27
SEQ ID NO: 7 moltype = DNA length = 47	
FEATURE Location/Qualifiers	
source 1..47	
mol_type = other DNA	
organism = synthetic construct	
SEQUENCE: 7 gcacgccttag ttccctaaat tagtcctact ttttgatgtc aaacact	47
SEQ ID NO: 8 moltype = DNA length = 47	
FEATURE Location/Qualifiers	
source 1..47	
mol_type = other DNA	
organism = synthetic construct	
SEQUENCE: 8 tgtgtcaaacc ttccctaaat tagtcctact ttttgatgtt taaaactg	47
SEQ ID NO: 9 moltype = DNA length = 1127	
FEATURE Location/Qualifiers	
source 1..1127	
mol_type = other DNA	
organism = synthetic construct	
SEQUENCE: 9 ttctttggc caactccaaac gaggccgtgc cgcgaccacc atcatcacga ccttatggcg	60
gcctcacggcc gcaaggccct gcacccctgt gcatcagggg tcatgcctcc ttcttaaagg	120
tgtctctttt ttgttatgtc gtcataactgt tgtaattca ctatgttta ttcatgtt	180
ccctttactt gtgggttagt ctatgtcgct ctggccgggtt ccaaccctag cccaaaaaaaaa	240
aatgaaatgg gtggaaagg cggccctgtt gatgtttaaa ataaatcatg ctaatgtt	300
gataactgtc atgttaggtt atatgttggg attaatgtt atatgttata gtaatgtt	360
tctcttttattt ttccctttt gtcataataa ttgttgcattt aatgttgcattt aatgttgcattt	420
gttagaaata tttaaaatgtt atgttgcattt gatgttgcattt gatgttgcattt gatgttgcattt	480
taaagtaaag cccaaaaaaaaa aatgttgcattt atgttgcattt atgttgcattt atgttgcattt	540
gcacgccttag ttccctaaat tagtcctact ttttgatgtc aaacactgtc atgttgcattt	600
gaaggccggaa aacgacaatc tgatccctgtt caatgttgcattt atgttgcattt atgttgcattt	660
ggatccact agttcttagag cggccgcgtt aatgttgcattt atgttgcattt atgttgcattt	720

---

-continued

---

```

agctcgaatt caaatttatt atgtgtttt ttcccggtt cgagattgtg tattattctt 780
tagtttac aagacttta gctaaaaattt gaaaagaattt actttaaaggaa aatcttaaca 840
tctggatcaa ttccgcaat agattattt ttccattact ctgcgtat ttgcagat 900
caatcgcaac atatatggtt gtttagaaaaa atgcactata tatataatata ttatgtt 960
aattaaaagt gcatgtata taatataat atatataat atgtgtgtgt gtatatggtc 1020
aaagaaatcc ttatcacaaat atacacgaa acatataattt gacaaaatca aagtattaca 1080
ctaaacaatg agttggtgca tggccaaaac aaatatgttag attaaaa 1127

SEQ ID NO: 10      moltype = DNA length = 1386
FEATURE           Location/Qualifiers
source            1..1386
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 10
tttgcggcg agagtcgtat ggactttgt tgttccctt ttgcctat tttctttctt 60
ttcatttta atttttcaa atgtccccat catcatcta ctccttctt ttatgtttt 120
tttcttgcg caactccaac gaggccgtgc cgccgaccacc atcatcaca ccttatggcg 180
gcctcacgcc gcaaggccct gcacccagggt gcacccagggt tcacgcctcc ttcttaaagg 240
tgtctcttctt ttgttatgtc gtcaagggtg tgcttaattt cctagaattt ttcaatgaat 300
ccctttaactt gtgggttagt ctaggcgtc ctggccgggtt ccaacccttag cccaaaaaaa 360
aatgaaatgg gtagggaaagg cgggcccgtt ttgaattaaa ataaatcatg ctaagatatt 420
gataactgtc atgtataggt atattttggg attaaattat ataggaaatta gtaattttc 480
tctcttattt cttccctttt gttcaataaa ttggaattt aacatcattt aagttttat 540
gttagaaaaata taaaatggt gataggatg gaataattag agtagaaaaaa 600
taaagtaaag cccaaaaaaag aaaattgggt atatgaagat acatgttagt catgcccag 660
gcacgcgttag tttgtgtgtc aaacactgt agttttaaact gaaggccgga aacgacaatc 720
tgatccccat caagcgttgc atcgaatttcc tgcagcccg gggatccact agttctagag 780
cggccgcgtt aactgcgtt ccacggatcc cccgggttccgc agctcgaattt caaattttt 840
atgtttttt tttccgtgtt cgagattgtg tattattctt tagttttttt aagactttt 900
gtctttttt gaaagaattt actttaaaggaa aatcttaaca tcttgagataa ttccggat 960
agattttttt tttcattact ctgcgtat ttgcgtat caatcgaac atatatggtt 1020
tttagaaaaaa atgcactata tatataatata ttatgtt tttttttt aattaaaatgt gcatgtata 1080
taatataat atatataat atgtgtgtgt gtatatggtc aaagaaatc ttatcacaaat 1140
atacacgaaac acatataattt gacaaaatca aagtattaca ctaaacaatg agttggtgca 1200
tggccaaacaaatgtt gatatttttcc cccggcccttcc aaaaaaaaaatc caatgtgtgt 1260
aaagcattat atatataat atgtatgttgc aattttgttgc aattccacac tgatcgat 1320
ttttaaagggtt aatatctgac gttaggatttt ttatgttgc tacctgacca ttatcaata 1380
acattc                                         1386

SEQ ID NO: 11      moltype = DNA length = 23
FEATURE           Location/Qualifiers
source            1..23
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 11
ctaaattatgtt cttttttt gat                                         23

SEQ ID NO: 12      moltype = DNA length = 25
FEATURE           Location/Qualifiers
source            1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 12
tttgcggcg agagtcgtat ggactttttt 25

SEQ ID NO: 13      moltype = DNA length = 25
FEATURE           Location/Qualifiers
source            1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 13
gaatgttattt agttaaatgggt caggt                                         25

SEQ ID NO: 14      moltype = DNA length = 1406
FEATURE           Location/Qualifiers
source            1..1406
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 14
tttgcggcg agagtcgtat ggactttgt tgttccctt ttgcctat tttctttctt 60
ttcatttta atttttcaa atgtccccat catcatcta ctccttctt ttatgtttt 120
tttcttgcg caactccaac gaggccgtgc cgccgaccacc atcatcaca ccttatggcg 180
gcctcacgcc gcaaggccct gcacccagggt gcacccagggt tcacgcctcc ttcttaaagg 240
tgtctcttctt ttgttatgtc gtcaagggtg tgcttaattt cctagaattt ttcaatgaat 300
ccctttaactt gtgggttagt ctaggcgtc ctggccgggtt ccaacccttag cccaaaaaaa 360
aatgaaatgg gtagggaaagg cgggcccgtt ttgaattaaa ataaatcatg ctaagatatt 420

```

-continued

---

gataactgct atgtataggt atatttggg attaaatttat ataggaatta gtaattttc	480
tctcttattt cttccctttt gtcacaataa ttggaattct aacatcattt aagttttat	540
gtagaaaata taaaagttt atgaattttt gatacttagt gaataattag agtagaaaaaa	600
taaaagtaaag cccaaaaaaag aaaattggtg atatgaagat acatgcttag catgcccag	660
gcacgcttag ttccttaaat tagtctact ttttgatgtc aaacactgt agtttaaact	720
gaaggcgaaa aacgacaatt tgatccccat caagctttagt atcgaattcc tgcaagccgg	780
gggatccact agttctagag cggeccgctt aactgcagg cgacggatcc ccgggttaccg	840
agotcgaatt caaattttt atgtgtttt tttccgttgtt cgagatttg tattttctt	900
tagtttattt aagactttt aactttaaattt gaaagaattt actttaagaa aattttaaca	960
tcttgagataa tttcagcaat agtttattttt tttcattact cttagcgtat ttttgagat	1020
caatcgcacaaat atatatgtt gtttagaaaaaa atgcactata atatatataa ttatttttc	1080
aattttaaatg gcatgatata taatataat atatatataat atgtgtgtgt gtatgttgc	1140
aaagaaaattt ttatacaat atacacgaaat acatataattt gacaaaatca aagtattaca	1200
ctaaacaatg agttgtgtca tggccaaaatc aaatatgttag attttttttt ccagcctcca	1260
aaaaaaaaatc caagtgtgtt aatgtttttt tagatccaa atttttgtac	1320
aatttccacac tgatcgaatt tttttttttt aatatctgac gtaggattt ttaatgttgc	1380
tacctgacca tttactaataa acatcc	1406

SEQ ID NO: 15 moltype = DNA length = 14436  
 FEATURE Location/Qualifiers  
 source 1..14436  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 15

ggaaaatcc ctcttcata ttaagaacat aaaaatcaac agggaaaata agttcaccaa	60
cccgaaaccag cacatccctt atgaacacgt cggggtaacg agcacttcta ttttccaaat	120
gaatcaccc atctgttagat tgcaaggatca caagagataa agaattgaaa atggacagag	180
gggtgacact aactgtatgc ccttagatcta gcattgttattt atcaaattta ctgttccaa	240
taatgcacgg tatacagaaa gtacccgtt ctttacattt ctcaggatg taaggaacaa	300
atttacccat caatgtgtac acatttctgtccat ctttccatgt ctttttgcgt	360
tcctttgtt ggtgcacaaatc tccttttagaa acttgacacaa tcttggaaatc tgcttgatgg	420
catcttagcag aggtatgttccatcttactt tccttgcaggatc tccttttgcgt	480
cttcttcata tttttttgtt tggattgttgc caagggttggaa atggaaaggaa gataagaggc	540
tgcggtaaatg cagaattact agaagaaggtt ccacctgtat gaaaattttt gtttaggaagc	600
tttctctttt gtgcacatctt ctcatccctt ttttcagggtt tagatgttagtgcgt	660
tcagggtcggg tgctgtgtac tgggtggatgtt gttgtcaga cctcaagggtt	720
atgacactca catttttccg attttgcaca gtttgcgtt gcaatttttgc agaattttgg	780
gaatgagttt ggttcaactg agtagccatc cgcccccattt gattttgcgt actctgtat	840
aaggctcttgc tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	900
tctaaaggaaatg tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	960
tgttgcgtttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1020
ggagggacac agttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1080
caacccgtat tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1140
tgttgggggg gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1200
gattactgca aaaaaggaca aagatctgtat tggttgcattt cagaagaaca tataccacag	1260
actctttgttca cttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1320
gcatcaatgtt tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1380
tcatcgactt cttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1440
gecttcttctt caatcaattt cttttttttt gttttttttt gttttttttt gttttttttt	1500
ctggcagcat caatcatactt cttttttttt gttttttttt gttttttttt gttttttttt	1560
aggagggtgtt cttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1620
cagttactcat cttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1680
gcagtgttcc tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1740
ctgaaaaatgg acctggggac aaggtttttt gttttttttt gttttttttt gttttttttt	1800
tggggggggg gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1860
caaaacaaatcat gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1920
tttgggtttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	1980
tgaatgcaca agttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	2040
cgggggtttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	2100
gaatattttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	2160
taatgtttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	2220
tcgaagggtt ataaatccatc tagattttttt gttttttttt gttttttttt gttttttttt	2280
tgttctttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	2340
atgggtttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	2400
aaacggaaat tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	2460
aataataactt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	2520
ctccaaacccggg tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	2580
caatatgaga aagttttttt gttttttttt gttttttttt gttttttttt gttttttttt	2640
ccacaccaag tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	2700
ctatcaaaac aaaaacagctt tttttttttt gttttttttt gttttttttt gttttttttt	2760
tcactgttca cttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	2820
gcaactgaaa tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	2880
ctttggccca tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	2940
ataaaatccat tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	3000
gtcaaccaat tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	3060
ttttttttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt	3120

-continued

---

gttattttccc	cttgattcat	aagcaagtct	cttatccctag	gttaggagga	tttatcccta	3180
accaggctaa	ccacttaatc	caaccctaaa	ttaaattact	aagcggaaaat	taacataagg	3240
ttgtctttat	atgattaagc	aacacataca	ccaattatc	atgaacaaaa	tcgatcatta	3300
agcatcaaca	taaattaagc	gcaaagataa	ttaatcaacg	actaagcgt	catggattag	3360
tagcaacaaa	tacagagtaa	ttgggtggaga	tgaaaaactg	atcaatattc	aatagtaata	3420
acaaaacctc	aaagagagtt	gtgttgttatt	ctcaagagaa	aacaacgcgt	gagacttagc	3480
cttccattaa	tcagtagaaa	acgaaattgt	agaaaacgaa	ttttattcta	tgtgaacaaat	3540
gtgcataac	agtaataaaa	actggattg	caaaacccct	aaattattct	tctctccaaa	3600
aaaactccct	aaactaaaac	cctggtgta	ttatatacggt	cctcagcccc	aaagcttaca	3660
aatcttattt	cagtcacaa	ccataaacg	aataaaataa	aatctggaca	agataagata	3720
agatggatg	aaataaaatc	tggacgaaat	aaatctggaa	taagtagaa	tttgataaaa	3780
taaaattgtc	tgcttttc	aagtcaago	ccaattccgg	attcaagccc	aattttttat	3840
aattcttctg	aaattaaatt	aaaaatacga	aattagtca	gtaggeccaa	atgataaaaac	3900
tgcatataat	atttgacaat	taaggctta	cagtattaa	aatagtgcac	aaaagggtta	3960
agaaatagga	gaataatgac	acatcacca	tatggggagc	aattctaaa	tgcatttgag	4020
ttctttaacc	tgagacacag	tgcaatgag	tctccaagg	ttcattgtgc	cttttattat	4080
atatgtatgg	gtcaactacat	tggecttgc	aaagaaactg	aatttgggg	attaaagaaa	4140
cacaataaa	aaacaaat	aactgttta	tagaaatgt	gcctatgtt	tcttggaaaa	4200
agtccaaac	tttgcgtt	ggataaaaat	catattacc	acttgcgt	tgttcaatca	4260
aacactagat	ttggataaaa	tctcactct	agatatacc	caaggatata	tatgaccaac	4320
attagtcat	tttagaaatg	aaagtggaca	aatttgagat	ttcattccct	aatgacatta	4380
taaacatgt	ttttttccat	gaccctttt	caatgttaat	ccttagtta	4440	
gatactatcat	atatgcatgt	taatgttgc	atgaaatgt	actcaatgtt	tttgttatgg	4500
ttaaatgttc	gactacccct	gatataac	tccatcttc	caatctcata	caaaagatac	4560
ttgtcaatgc	gtacctgaac	cttgcgtt	tgcaatgtt	agtttttct	gaaggcacac	4620
gctttagatag	taaccagaag	ccaggaggga	gtcctctaa	gtctcaatctc	gtatttccg	4680
tggaaatgt	ttttttttct	taaagaaaa	agagatagt	taccaatgt	aatattttt	4740
tagccaaata	gggacatcat	agaaaaacaa	acttttcttc	taatgttata	atgcaactac	4800
atattttatgg	tcgcgttgc	tcgtttaaaa	ataagggct	agacaacaca	aaaatattt	4860
tccaacgtt	gatttttaaa	atggctgaga	gacaatacaa	aataaaatg	gatgaactgg	4920
acaaaacctt	aaactactgt	aactctatca	acttttgcata	actttttgtt	cagtgtctaa	4980
aaaatggat	aaatacaat	tattctatt	ttttactttt	attatctcac	accttcttc	5040
tactcatttgc	tttcaattc	cctcccttgc	gggcacccgt	gtttgtggc	gagagtgcgt	5100
tggacttttgc	ttgtttcc	tttgcattt	atttctttct	tttcattttt	aattttatca	5160
aatgttccat	tcatcatct	actcttctt	gttatgttt	ttttctttgg	ccaaactccaa	5220
cgaggccgt	ccgcgaccac	catcatcag	accttgcatt	ggccctcagc	cgcaaggcccc	5280
tgcacccagg	ggcgcattagg	gtcatgcct	cttcttaa	gtgtctct	tttggatgt	5340
cgtcaatgt	ttgcataattc	accttgcatt	tttcaatgt	cccccttact	ttgtgggttag	5400
tcttaggtcgc	tctgcgggt	tccaaaccct	gccccaaaaa	aaatggaaat	ggtagggaaag	5460
gccccccctag	tggatattaa	ataaaatcat	gtcaatgtt	tgataactgc	tatgtatagg	5520
tatattttttgc	gattaaat	tataatgtt	agtaattttt	ctcttattt	tcttccttt	5580
tgttcaataat	attggaaatc	taatcatatt	taagttttt	tgttagaaaat	attaaaatgtt	5640
gatgaatttgc	tgtatcttgc	tgtatatttt	gatgtatattt	ttatgttata	caagacttt	5700
gaaaattttgtt	gatataaga	tatcatgtt	gtatgttgc	ggcacgttta	gtttcttaaa	5760
ttatgttccat	ttttttatgt	caaactgt	tagtttaac	tgaaggccgg	aaacgcataat	5820
ctgtatccca	tcaagcttgc	tatgttattt	ctgcagcccc	ggggatccac	tagttctaga	5880
gccccccctgcgt	taacttgcgg	tcgcggatc	ccccgggttac	gagctcgat	tcaatatttt	5940
tatgtgttttgc	ttttccgttgc	tcgatgttgc	ttatgttattt	ttatgttata	caagacttt	6000
agotaaaaat	tggaaatgtt	tacttttgc	aaatcttac	atcttgcata	atttcagccaa	6060
tagattttat	tttttatttgc	tctatgttgc	tttttgcaga	tcaatcgca	catatatgtt	6120
tgtttagaaaa	aatgcactat	atatatata	attatttttt	caataaaaat	tgcatgtat	6180
ataatataat	tatataatata	tatgttgcgt	tgtatgttgc	caaaagaaat	cttataacaa	6240
tataacacgaa	cacatattat	tgacaaaatc	aaatgttac	actaaacat	gagttttgtc	6300
atggccaaaa	caaataatgt	gattaaaaat	tccagccttc	aaaaaaaaat	ccaagtgttgc	6360
taaaacatca	tatataatata	gtatgttcc	aaattttgc	caatccca	ctgatcgat	6420
ttttaaatgtt	gatataatgt	cgtaggttt	ttttatgttgc	ttacctgcac	atttactata	6480
aacatttata	cggttttatttgc	tgaaatattc	tatataattt	tattgttata	ggccacataat	6540
aagaaacacta	attgggttgc	tattttacta	gtaaatttgc	gttgcatttttgc	tttctactat	6600
aaaggcttcg	gaaaatcttgc	gaccaatattc	atattccat	acttgcatttgc	tttgcatttgc	6660
tagtttttgc	aaacatatact	tcaatatttgc	tgcaacggaa	aaatggatca	ttttatgttgc	6720
aatccgtat	ggaaatgttgc	tagggatgttgc	ggaaatgcgg	ccccggaaaa	gattttatgt	6780
acttgcatttgc	tgtatgttgc	tcaatgttgc	taaaggaaat	tggagacttgc	agaatgttttgc	6840
tttgcatttttgc	ttttatgttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	6900
tatgttatttttgc	aatattttgc	tgtatgttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	6960
ttttgcatttttgc	ttttatgttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	7020
ttaaataatgttgc	ttttatgttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	7080
ttttctatata	ttataatgttgc	ttttatgttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	7140
attccacatgt	ttcacatata	tttgcatttgc	tttgcatttgc	tttgcatttgc	tttgcatttgc	7200
ttatgtat	atcatatcat	tataatattaa	ctaaatccaa	ggtaaaaaaa	aggtatgtaaa	7260
gctctatgt	aaatggatata	taaatttccca	tttttttttttgc	tttttttttttgc	tttttttttttgc	7320
taaaatgtat	ttttatgttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	7380
catggatata	tctccacgt	ggcatttttgc	cagcggttca	agccgatata	ggtctcaaca	7440
cctctcccttgc	ggcccttgc	ggccgttacca	agtttttttttgc	tttttttttttgc	tttttttttttgc	7500
caaaatccaa	cggtgtatgt	cctgttcc	ttgtatgttgc	tttttttttttgc	tttttttttttgc	7560
cactccaaat	ttttatgttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	7620
caaaacactc	ttttatgttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	tttttttttttgc	7680

-continued

---

tactatgggt	gcctctccgg	ctcaggccac	tatggtcgct	ccttcaacg	gacttaagtc	7740
cttcgcgtcc	ttcccgacca	cccgaaaggc	taacaacgac	attacttcca	tcacaacgaa	7800
cggcggaga	gttaactgca	tgcaagggtg	gcctcccgatt	ggaaaagaaga	agtttgagac	7860
tctcttta	cttcctgacc	ttaccgatc	cgggtgtcg	gtcaactgca	tgcaaggccat	7920
ggacaacaac	ccaaacatca	acgaatgcat	tccatacaac	tgcttgagta	acccagaagt	7980
tgaagtactt	ggtggagaac	gcattgaaaac	cggttacact	cccatcgaca	tctcttgc	8040
cttgacacag	tttctgtca	gcegatctgt	gccagggtgt	gggttgcgtc	tcggactgt	8100
tgacatcat	ttgggtatct	ttggccatc	taatggat	gcatcttgg	tgcaaatgta	8160
gcagttgatc	aaccagagga	tcgaagagtt	cgccaggaaac	caggccatct	ctaggttgg	8220
aggattgagc	aatcttca	aaatctatgc	agagagcttc	agagagttgg	aagccgatcc	8280
tactaaacca	gtctcccg	agggaaatgg	tattcaattc	aacgacatga	acagcgctt	8340
gaccacgct	atccccattgt	tcgcgatc	gaactaccaa	gttcccttct	tgccgtgt	8400
cgttcaagca	gtaatcttc	acctcagcgt	gttccgagac	gttagegtgt	ttgggcaag	8460
gtggggatc	gtatgtc	ccatcaatag	ccgttacaaac	gaccatcta	ggctgtatgg	8520
aaactacacc	gaccacgt	ttcggtgtt	caacactgtt	ttggagcgtt	tctgggttcc	8580
tgattctaga	gattgatta	gatacaacca	gttcaggaga	gaattgaccc	tcacagttt	8640
ggacattgt	tctcttcc	cgaactatga	ctccagaacc	taccctatcc	gtacagtgtc	8700
ccaaactacc	agagaatctt	atactaaacc	agttcttgag	aacttcgacg	gtacgttccg	8760
ttgttctgc	caaggatctcg	aagggttcat	caggagccca	cacttgatgg	acatcttga	8820
cacgatatact	actacaccg	atgtccacag	aggagagtat	tactgttctg	gacaccagat	8880
catggctct	ccagttggat	tcaggggco	cgagtttacc	tttccctct	atggaactat	8940
ggggaaacccc	gttccacaac	aacgtatctgt	tgctcaacta	ggtcagggtg	tctacagaa	9000
cttgcgttcc	accttgc	gaagacccct	caatatecggt	atcaacaacc	agcaacttcc	9060
cgttcttgac	ggAACAGAGT	tcgcgtatgg	aaaccttctt	aacttgcatt	ccgctgttta	9120
cagaaagagc	ggAACCGTTG	atctcttgg	cgaaatccca	ccacagaaaca	acaatgtgc	9180
acccaggca	ggattttccc	acaggtttag	ccacgtgtcc	atgttccgtt	ccggatttc	9240
caacacgttcc	gtggacatc	tcagacgtt	tatgttcttct	tggatatactc	gtatgtgt	9300
gttcaacaac	atcatcgat	ccgtatgtat	tactcaatc	cctgcgtgt	agggaaactt	9360
tctcttca	gttctgtca	ttttaggacc	aggatttact	gttggagacc	tcgttagact	9420
caacagcagt	ggaaataaca	ttcagaatag	agggtatatt	gaagttccaa	ttcacttccc	9480
atccacatct	aggatgttac	gaggttgcgt	gaggatgtt	tctgtgaccc	ctattcacct	9540
caacgttata	ttgggttatt	catccatctt	cttccaaata	gttccagctt	cagctactc	9600
cttggataat	ctccaaatcc	gcgatttcgg	ttactttgaa	agtgcctatg	cttttacatc	9660
ttcaactcggt	aacatcg	gtgttagaaa	ctttagtgg	actgcaggag	tgattatcg	9720
caagatcgat	ttccatccag	ttactcgac	actcgagggt	gactacaacc	ttgagagagc	9780
ccagaaggct	gtgaacggcc	tctttacctc	caccaatcag	cttgggttga	aaactaacgt	9840
tactgactat	cacattgacc	aagtgtccaa	cttggtacc	tacccatgg	atgagttctg	9900
cctcgcacg	aagcgtgaa	tctccggagaa	agttaaacac	gccttgcgtc	tcagcgtac	9960
gaggaatctt	tttgcgttca	ccatcttca	agacatcaac	aggcagcgt	aacgtgggt	10020
gggttggaa	accggatca	ccatccagg	aggcgacgt	gttgcgttca	agaaactacgt	10080
cacccttcc	ggaaacttcc	acgagtgtca	cccttaccc	ttgttacc	agatgtat	10140
gtccaaactc	aaagccttca	ccaggtatca	acttagaggc	tacatcgaa	acagccaaga	10200
ccttgcata	tactcgat	gttgcgttca	caagcgttac	accgtgtat	tcccggttac	10260
ttgttccat	ttggccatctt	ctgcccata	teccatttgg	aaatgttgg	agccataacag	10320
atgcgttca	cacccgttgc	ggatcttca	cttggacttc	tcttgcagg	atggcgat	10380
gtgtgtcc	catttcatc	acttcttctt	ggacatcgat	gttggatgt	ctgacactgaa	10440
tgaggaccc	ggagtgttgg	tcatcttca	gatcaacgtt	caagacggac	acgtcaagact	10500
ttggcaactt	ggatgttccg	aaagaaaaac	attggtcgtt	gaaggtctcg	ctcggtgt	10560
gagagcagag	aaaaatgtt	ggggacaaac	tgagaaactt	gaatggaaa	ctaacatcg	10620
ttacaaggag	ggccaaagat	ccgttgcgt	tttggatgt	aactccat	atgatcgat	10680
gcaagccgac	accaacatcg	ccatgtatc	cgccgcgtac	aaacgtgtgc	acagcatcg	10740
tgaggaccc	ttgcgttgc	tttgcgtat	tttgcgttgc	aaacgttgc	tcttcgagga	10800
acttgcgg	ctgtatctt	ccgtatctc	cttgcgtat	gccagaaac	tcataagaa	10860
cggtgactt	aaacatggcc	tcagtcgt	gaatgtgaa	ggtcgtatgt	acgtggagga	10920
acagaacaaat	cacggatcc	tcctcggtgt	gccttgcgtt	gaaggtgtt	tgttccaa	10980
gggttgcgtt	tttgcgttca	gaggatgtt	tctccgtgt	accgttata	aggaggat	11040
cggtgagggt	ttgcgttgc	tccacggat	cgagaacaaac	accggacgt	ttaatgttcc	11100
caactcgatc	gaggagaaaa	tctatccaa	caacaccgtt	acttgcac	actacactgt	11160
gaatcgaa	gagttcggt	gtgcgttac	tagccgttac	agaggatata	acgttgttac	11220
ttccgttcc	gttgcgtat	cctcggtgt	cgaggatgtt	tttgcgtat	tttgcgttac	11280
tgagaaccc	ttgcgttca	acagatgtt	cgaggatct	acaccatctc	cgttgcgtt	11340
tgtttacca	gagtttgcgt	acttcttca	gaccgtat	gtgttgcgtt	agatgtgt	11400
aaccggat	accttgcatc	ttggacaggt	ggagcttctc	tttgcgttgc	tttgcgttac	11460
tcccggttcc	tgttgcgttca	tttgcgttgc	taagtgttgc	tttgcgttgc	tttgcgttac	11520
gtatgtgtt	atagtcttca	gttgcgttgc	aaatgttgc	gttgcgttgc	tttgcgttac	11580
gtcccatctt	ttgtatata	aatatcgat	tgatgtttt	ttttgtgt	tttgcgttac	11640
ctgttgcgtt	aatttgtat	gttgcgttgc	tggcgtat	aaatatgtt	tttgcgttac	11700
cttgcgttac	aaatataact	gattatgtt	gtgttattata	tacaaaag	tttgcgttac	11760
atacatctt	gacaaatgtt	cacggatcc	caaaaaaa	attgttgc	ttttgggt	11820
tttcgttgc	acaagatcg	tctgtatgt	aaacaatttct	aaggttgc	atgtgtac	11880
gtccgttcc	aatatgtt	tgatgttgc	cccgccgtt	ccatgtatc	gttgcgttac	11940
tcatcttgc	ccccatgtt	acgtgtatgt	agacacgttgc	aaataaagat	tttgcgttac	12000
gaataatgtt	tttattgtt	tcgcgttata	atacgtatgtt	tcgtatgtt	tcgttgcgtt	12060
aaaatgttact	ttcattttat	aataacgttgc	cggtatgttgc	cattttgtt	tttgcgttac	12120
attgttgc	attcttctt	tttgcgttca	tttgcgttgc	tttgcgttac	tttgcgttac	12180
agatgttcc	gacatgtatc	catcaaaa	taggtactat	tttgcgttac	tttgcgttac	12240

-continued

aaggaaagtga	aggcacgctt	agtgtgagac	acgtgttgag	cgcgattact	gccactca	12300
aaccacacaa	gtgcactcg	tgccagggtt	gcttaaaaat	taagtgtt	cgca	12360
aaagaaggat	agagatgaag	gaaaaaacac	agaaaataca	atccc	tata	12420
gctagaagaa	gcaaacgca	acatttaga	tcattcc	cctcaattc	cattttca	12480
ttcccccttt	actaaatatt	ctcccttgc	aattataa	gc	cctatga	12540
ctaaacttc	cttttgtggg	aacttatcg	tcaactgtc	ttaatataat	ttctcttct	12600
atcttattg	aattttcact	acaagaaat	tgcccattt	ccaggattt	ttgacagg	12660
catatacccc	tggcaattt	cccaaggact	aagccaagg	aaccctgg	aaaatgac	12720
tttgagaaggc	tgggaccact	tacatttaca	cagggttt	tccctcgaa	aaatac	12780
gccttggc	aaaaaaagc	ggggaaatgaa	ttttaaaaca	gcatttttt	ttcacacac	12840
caaacacac	ggatgaccc	cgttttcc	aaagctgac	gaatttcc	ataagtc	12900
acgacatgac	catgca	aaaaa	gctg	cgatc	agggtttcc	12960
tcggaaatgg	cttgcagcc	ctggcaaaa	ggaatccct	ctttcc	tacaccgtt	13020
tgtctatata	gtgtgacta	ggggattt	catttttt	gttttttgc	gaggggc	13080
ccgtgatgt	tccctgggt	tatata	gttttttt	ttatagcc	aaccgtgt	13140
gctcagttt	gtgttttgc	aactttagaa	ttttttcg	ttccattttc	atc	13200
gttcattttc	agtccattat	cattcagtt	atacactt	tttataattt	ggta	13260
tttca	tttatatttt	ctgtttttt	ttgttactac	ttat	aaatatttt	13320
tattgtatca	gtgtccaa	at	ttgttttcc	ttgtgtct	ttttt	13380
cttaagttt	aaaaagtt	ttat	tttttttt	tttagatata	tttttttata	13440
tatatgtgt	tataat	tttgc	catgttgc	caa	agaaaat	13500
atgtgttata	atata	tttgc	ttttttttt	tttttttt	ttttttttt	13560
aattttaaac	tgttataatt	tttttttta	taaaactttt	taat	ttttttttt	13620
aattttttttt	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	13680
tgttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	13740
aacttgcgtca	gtacacgac	gtcgtcagaa	agtg	gtcgtcag	agaagac	13800
gottaagaca	ggggat	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	13860
gtttttttttt	caggggttga	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	13920
gattcagcc	ctggcaaaat	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	13980
cccaaggaaat	ccggccctgg	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	14040
ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	14100
aggggttagcg	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	14160
tcatca	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	14220
tttaactatt	ttatctctgc	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	14280
caactcgaggc	tgtgtaca	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	14340
ccaccatca	gtttggctt	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	14400
acttggtcac	tttcc	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	14436

SEQ ID NO: 16      moltype = AA   length = 1307  
 FEATURE      Location/Qualifiers  
 source      1..1307  
 mol\_type = protein  
 organism = unidentified

SEQUENCE: 16

MTQFEGFTNL	YQVSKTLRFE	LIPQGKTLKH	IQEQQFIEED	KARNDHYKEL	KPIIDRIYKT	60
YADQCLQLVQ	LDWENLSSAI	DSYRKEKTEE	TRNALIEEQ	TYRNNAIHDF	IGRTDNLTDA	120
INKRHAEIYK	GLFKAEFLFG	KVLQLGTT	TTEHENALLR	SFDKFTTYS	GFYENRKNVF	180
SAEDISTAAIP	HRIVQDNFPK	FKENCHIFTR	LITAVPSLRE	HFPENVKAIG	IFVSTSIEEV	240
FSPFPFYNQOLL	TQTQDLYNQ	LLGGISREAG	TEIKIGLNEV	LNLAIQKNDE	TAHIIASLPH	300
RFIPLFLKQIL	SDRNTLSFIL	EEFKSDEEV	QSFCKYKTL	RNENVLETAE	ALFNELN	360
LTHIFISHKK	LETISSLCD	HWDTLRNALY	ERRISELTGK	ITKS	AKEVKVQ	420
QEIIASAAGKE	LSEAFQKQTS	EILSHAHAL	DQPLPTTLKK	QEEKEILKSQ	LDSLGLYHL	480
LDWFAVDESN	EVDPPEFSRAL	TGIKLEMEPS	LSPYNKARNY	ATKKPYSVEK	FKLNFQMP	540
ASGWDVNKEK	NNGAILFVKN	GLYYLGIMPK	QKGRYKALSF	EPTEKTSEGF	DKMYYDYFPD	600
AAKMIPKST	QLKAVTAHFNG	THTTPILLSN	NFIEPLEITK	EIYDLNNPEV	EPKKQ	660
KKTGDQKGYR	EALCWKDF	RDFLSKYKT	TSIDLSSLRP	SSQYKDLGEY	YAELNPL	720
ISFQRIAEKE	IMDAVETGKL	YLFOIY	NKDF	AKGHHGKP	NLHTLYWTGLFS	780
LNGQAELFYR	PKSRMCRM	RLGEKMLNKK	LKDQKTP	IPD	TLYQELYD	840
EARALLPNVI	TKEVSHIEII	DRRFTSDKF	FHV	PITLNYQ	AANSPSKFNQ	900
ETPIIGIDRG	ENRSHIYITVI	DSTGKILEQR	SLNTIQ	QFDY	QK	960
VGTIKD	LQKG	YLSQVIHEIV	DLM	YQAVV	NLFGKF	1020
DKLNCLV	LKD	YPAEKVGVL	NPYQLTDQFT	SFAKMG	TQSG	1080
DPFVW	KTIKN	GDFILHFKMN	RNLSF	QRLP	GFMPA	1140
EKNETQFD	AK	GP	TPI	IV	AWD	1200
PKLLENDSSH	AIDTMVALIR	VPVIENHRPT	GRYRDLYPAN	ELIALLEEKG	IVFRDGSN	1260
DADANGAYHI	ALKGQ	LLNH	NGISNQDWLA	YIQELRN	1307	

SEQ ID NO: 17      moltype = DNA   length = 40  
 FEATURE      Location/Qualifiers  
 source      1..40  
 mol\_type = other DNA  
 organism = synthetic construct

SEQUENCE: 17

ccaggcacgc	ttagtgtgt	tgtcaaacac	tgtatagttt			40
------------	-----------	------------	------------	--	--	----

SEQ ID NO: 18      moltype = DNA   length = 5757

-continued

---

FEATURE	Location/Qualifiers
source	1..5757
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 18	
ggaaaaatcc ctcttcata ttaagaacat	aaaatcaac agggaaaaata agttcaccaa 60
cccgaaaccag cacaatctt atgaaacctg	cggggttaagc agcacttcta ttttccaat 120
gaatcaccac atctgttagat tgcaaggc	caagagataa agaattgaaa atggacagag 180
gggtgacact aactgtatgc	cctagatcta gcattgtatt atcaaattt ctgttccaa 240
taatgcacgg tatacagaatg	gtatctgggt ctttacattt ctcaggaaatg taaggaaacaa 300
atttacatcatat caatgtgac	acattttttt ccatgtcaat cttttcatgt ctttgagct 360
tccttttgtt ggtgcacaac	tccttttagaa acttgacaca tcttggaaatc tgcttgatgg 420
catcttagcg aggtatgttc	acctctactt tcctgaaggct ctccaaatgc tcctttctg 480
cttcttccat ttttttttt tggaaatgtt	caagggttggaa atggaaaggaa gataagggc 540
tgcgtttagt cagaattact	agaagaagggt ccacatgtcat gaaaattttt gttaggaagc 600
tttctttttt gtgcactat	ctatcttcctt tttttttttt cttttttttt tagaatgttgc ctttgcacgg 660
tcagggtcggt gtgtgtctac	tgggggggtt acttgaattt ggttgcaga cctcaagggt 720
atgacactca catttttccgg	atttttccgg gtttgcaca gtttgcaga gcaatttttc agaatttttg 780
gaatgagctt ggttcaactt	agttagccat cggccatcattt gtttgcatac actctgaatg 840
aaggcttctt tttttttttt	aaatggcata tttttttttt tttttttttt tttttttttt cttttttttt 900
tctaaggaaatg	gttaaggagg agtctcgtt gtttgcatac tttttttttt tttttttttt 960
tgttgcgtct gtattttttttt	gtttttttttt gtttgcatac tttttttttt tttttttttt 1020
ggaggggacag acgtttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 1080
caacctggat tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 1140
tgttgggggg gtctattata	aatgttttgc gtttgcatac tttttttttt tttttttttt 1200
gattactgtca aagaaggaca	aagatctgtt gtttgcatac tttttttttt tttttttttt 1260
actcttgcataa cagggtcgaaa	ttttttttttt gtttgcatac tttttttttt tttttttttt 1320
gcatcaagggtt tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 1380
tcatcgactt ctcttgcataa	ttttttttttt gtttgcatac tttttttttt tttttttttt 1440
gccttcttctt caatcaattt	ttttttttttt gtttgcatac tttttttttt tttttttttt 1500
ctggcgcgtt caatctactt	ttttttttttt gtttgcatac tttttttttt tttttttttt 1560
aggagttgtt cagaatctgtt	ttttttttttt gtttgcatac tttttttttt tttttttttt 1620
cagtactcat acaagtttctt	ttttttttttt gtttgcatac tttttttttt tttttttttt 1680
gcagtgggtcc tagatgttgc	ttttttttttt gtttgcatac tttttttttt tttttttttt 1740
ctgaaatatttgg acctgtttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 1800
tgaggaaaatg tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 1860
caaaacatat ggaacttctt	ttttttttttt gtttgcatac tttttttttt tttttttttt 1920
ttgggttagca aatgttattttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 1980
tgaatgcaca atttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 2040
cgagggttggat gtttgcatac	ttttttttttt gtttgcatac tttttttttt tttttttttt 2100
gaatatttttttttcaatcaca	ttttttttttt gtttgcatac tttttttttt tttttttttt 2160
taatgtttagt gatgcactt	ttttttttttt gtttgcatac tttttttttt tttttttttt 2220
tcgaagggtt ataaatccat	ttttttttttt gtttgcatac tttttttttt tttttttttt 2280
tgttcttcaaa caagcacaat	ttttttttttt gtttgcatac tttttttttt tttttttttt 2340
atgagttgttcaat tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 2400
aaacgaaaat tcaaaatgttcaat	ttttttttttt gtttgcatac tttttttttt tttttttttt 2460
ataataatgttcaat tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 2520
cttcaacccggg tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 2580
caatatgaga aatgtttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 2640
ccacaccaatg tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 2700
ctatcaaac aaaaacatgttcaat	ttttttttttt gtttgcatac tttttttttt tttttttttt 2760
tcactgttca cagcaaaatca	ttttttttttt gtttgcatac tttttttttt tttttttttt 2820
gcaactgttca tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 2880
ctttggccca tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 2940
ataaaatccat aaaaatgttcaat	ttttttttttt gtttgcatac tttttttttt tttttttttt 3000
gtcaacccat tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 3060
ttttttttttt aattttttttcaatcaca	ttttttttttt gtttgcatac tttttttttt tttttttttt 3120
gttatttttttttcaatcaca	ttttttttttt gtttgcatac tttttttttt tttttttttt 3180
accagtccat tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 3240
ttgtcttttat atgatgttcaat	ttttttttttt gtttgcatac tttttttttt tttttttttt 3300
agcatcaaca tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 3360
tagcaacaaa tacagatgttcaat	ttttttttttt gtttgcatac tttttttttt tttttttttt 3420
acaaaacccat aagagatgttcaat	ttttttttttt gtttgcatac tttttttttt tttttttttt 3480
cttccattaa tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 3540
gtgtcatgttcaat tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 3600
aaaactccat aactaaaaatcaat	ttttttttttt gtttgcatac tttttttttt tttttttttt 3660
aatcttattttt cttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 3720
agatgttggat aataaaatcaat	ttttttttttt gtttgcatac tttttttttt tttttttttt 3780
taaaaatgttcaat tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 3840
aatttttttttcaatcaca	ttttttttttt gtttgcatac tttttttttt tttttttttt 3900
tgcataatcaat tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 3960
agaaatagtaga aataatgttcaat	ttttttttttt gtttgcatac tttttttttt tttttttttt 4020
tttttttttttcaatcaca	ttttttttttt gtttgcatac tttttttttt tttttttttt 4080
atatgttggat gtttgcatac	ttttttttttt gtttgcatac tttttttttt tttttttttt 4140
cacaaaatataa aacaaaatgttcaat	ttttttttttt gtttgcatac tttttttttt tttttttttt 4200
agtccaaatcaat tttttttttt	ttttttttttt gtttgcatac tttttttttt tttttttttt 4260

---

-continued

---

aacactagat	ttggataaaaa	tctcaactcct	agatataacct	caagggataa	tatgaccaa	4320
attagtatt	tttagaaagt	aaatggaca	aatttggagat	ttcatttc	aatgacatta	4380
taaacatgt	tttttccat	gaccctttt	caatgttaat	acaattttc	ccttagtta	4440
gataactctat	atatgcattgt	tacgttagtg	atgaaaacat	actctaagt	tttgttatgg	4500
ttaagttgc	gactacctct	gatataaaac	tcctcatctc	caatctcata	caaaagatac	4560
ttgtcaacttg	gtaccgtgaac	cttgcactgtt	tgcaagtgtt	agtttcttc	gaagccacac	4620
gettgtatag	taaccagaag	ccaggaggga	gtccctctaag	gtcttaactc	gtatttccg	4680
tggaaagtaca	tttttttct	taaagaaaaac	agagatgtt	taccaatgtat	aatatttttt	4740
tagccaaata	ggaccatcat	agaaaaacaaa	actcttctt	taagtattta	atgcaactac	4800
atatttaggg	tgcgtttgat	tcgcgttttt	ataagggtct	agacaacaca	aaaatatttt	4860
tccaaacgtt	gatttttttt	atgggtgaga	gacaataaaat	aataaagaaat	gtgaaactgg	4920
acaaaaaccc	aaaaacttgc	aactcaactga	atctcataca	actttttttt	cagtgtctaa	4980
aaaaagtaaa	aatacaatat	tatccctatt	tttactttt	attatctcac	accttcttc	5040
taactcaatttgc	tttcaatttgc	ccttcccaat	gggcacccat	gtttgtggc	gagagtgcga	5100
tggacttttg	ttgttttttct	tttgcattt	atttctttt	tttcattttt	aatttattca	5160
aatgttccca	tcatcatctt	actcttctt	gttatgtttt	ttttttttt	ccaactccaa	5220
cgaggcccgt	ccgcgcaccac	catcatcact	accttatggc	ggccctcacgc	cgcaaggccc	5280
tgccacccatgt	ggccatcaggg	gtcatgcctc	ccttcttaaag	gtgtcttc	tttggtatgt	5340
cgtccaaatgt	ttgtcaatttgc	accttagattt	tttcaatgtt	ttcccttact	tgtgggttag	5400
tcttaggtgc	tctggcccggt	tccaaacccat	gccccaaaaaa	aatgaaatg	gttaggaaag	5460
gggggcctag	tttgaattaa	aataaattat	gctaagatgt	tgataactgc	tatgtatagg	5520
tatattttttgg	gattaaat	tataggaatt	agtaattttt	ctctttttt	tcttcctttt	5580
tgttcaataa	attggaaattc	taatccattt	taatggattt	tgttagaaat	attaaaagtt	5640
gatgaatttttgc	tgatacttag	tgataatttt	gatggaaaaa	ataaaagttaa	gccccaaaaaa	5700
gaaaattttgtt	gatataaga	tacatgttta	gcatgccccca	ggcacgttta	gtgtgtt	5757

SEQ ID NO: 19	moltype = DNA length = 40
FEATURE	Location/Qualifiers
source	1..40
	mol_type = other DNA
	organism = synthetic construct

SEQUENCE: 19		
ttcccgac atgaagccat	caaaaagtag gactaatttt	40

SEQ ID NO: 20	moltype = DNA length = 2233
FEATURE	Location/Qualifiers
source	1..2233
	mol_type = other DNA
	organism = synthetic construct

SEQUENCE: 20						
caaaaagtag	gactaatttt	ggaaagcaag	ctaattcaag	aaagtgaagg	cacgtttagt	60
gtgagacacg	tgttgagcgc	gattactgcc	actcaactac	cacacaatgt	cactcagtgc	120
gaagggtgt	ttaaaat	ttgttgcgt	acttataaaa	gaaggataga	gtgaaaggaa	180
aaaacacaca	aaatataat	ccttataat	gacaaaggct	agaagaagca	aacgcaaaaca	240
ttagaagtca	ttccttccct	caatccctt	tttcaattt	ccctttact	aatatttctc	300
ctcttgcatt	tataaagcct	cctatgacaa	tgacaagcta	aactcttctt	ttgtggaaac	360
ttatcgttca	actgtctta	ataatattt	tcttccttatc	tatttgcattt	attcaactaca	420
agaatatgc	ccatggccca	gggatttttt	acagggacat	taacccctgg	caaattttcc	480
agggactaa	ccaaaggaaac	ccctggcaaa	atgacatttgc	agaagggtgg	gaccactac	540
atttacacag	gggtttgtcc	ctcgcaaaaa	tacaaaaggcc	ttggcaaaaa	aaagagcggg	600
aatatgttca	taaaacagca	tggttgcgtt	acacagccca	acacacgggt	atgcctcggt	660
tttctgtttaa	gtgtgcggaa	tcttccttata	agtcaacac	acatgaccat	gcaactcaaa	720
aaqctgtgc	gccccagacgt	gacaggggt	tttcccttgc	gaaatggctt	gcagccctg	780
gcaaaaaagg	atccctgtt	ttcttagctac	accgttctgc	tcataatgt	gaagctttag	840
ggtttagcc	tgacttgcgtt	gttttgcgt	gggcatttc	tgatgtttt	cctgggttt	900
tttacactat	atagccaaac	cgcgttttgc	ttctctatgt	cgttgttgc	tttttgcata	960
ttagaaaaat	tttcggtttgc	cattttccat	cttaccaggat	cattttcgt	ccattatcat	1020
tcagttcata	cactgttct	ataatttgtt	aaactctttt	tcacttattat	tatttttctg	1080
tttttttttt	ttaactactt	ttaacatata	tatttttttt	tgtatgttgc	tccaaattttgc	1140
cctccctctgt	ctgtctcttgc	ctctctgtt	tttttttttt	tttttttttt	tttttttttt	1200
tttttttttttt	tataattttgc	tatattat	tttttttttt	tttttttttt	tttttttttt	1260
gatctgtca	agaaaaatat	atgttttttgc	tttgcattat	tttgcattat	tttgcattat	1320
atataatattt	cgatatttttgc	tgttataata	actgttttttgc	tgttataata	tatgtatgtt	1380
tgttataata	actgttttttgc	tgttataata	tttttttttt	tttttttttt	tttttttttt	1440
tttttttttttt	tatataat	ttatataat	tttttttttt	tttttttttt	tttttttttt	1500
aaaaaaatgtt	aaaactatgtt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	1560
tcagaaaaatgt	gtgcagatgc	gtcactgttttgc	tttttttttt	tttttttttt	tttttttttt	1620
ggaatttttgc	cagggttttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	1680
cctggcaaaac	tgatgttgcgt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	1740
tggccaaatgt	ccatggccca	tttttttttt	tttttttttt	tttttttttt	tttttttttt	1800
acgagttgtt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	1860
tttttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	1920
tttctctata	gaaactggaaa	tttttttttt	tttttttttt	tttttttttt	tttttttttt	1980
tttattttat	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2040
gtgatgttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2100
gagagagcccc	agaaggctgt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	2160

---

-continued

---

```

actaacgtta ctgactatca cattgaccaa gtgtccaaact tggtcacctt ccttagcgat 2220
gagttctgaa ggg                                         2233

SEQ ID NO: 21          moltype = DNA  length = 1127
FEATURE           Location/Qualifiers
source            1..1127
                   mol_type = other DNA
                   organism = synthetic construct

SEQUENCE: 21
tttctttggc caactccaaac gaggccgtgc cgcgaccacc atcatcaga ccttatggcg 60
gcctcaagcc gcaaggccc gcacccagtgc gatcagggg tcatgcctcc ttcttaagg 120
tgtctctttt ttgttatgtc gtcaaatgtg tgtaattca cctagaattt ttcaatgaat 180
ccctttactt gtgggttagt cttagtcgt ctggccgggtt ccaaccctag cccaaaaaaa 240
aatgaaatgg gttaggaagg cgggcctagt ttgaattaaa ataaatcatg ctaagatatt 300
gataactgtc atgttataggt atatTTGGG attaaattat ataggaattt gtaattttc 360
tctcttattt ttccctttt ttcaataaa ttgaaatttctt aacatcatttt aagttttt 420
gtagaaaata tttaaaagttt atgaatttt gataacttagt gaataatttt agtagaaaaaa 480
taaaatggaaag cccaaaaaaaag aaaattttttt gatggatcatcgtt catgccccag 540
gcacgcttag tttgtgtgtc aaacttttttcc aaatttttttgc tttttttttt gtttttttt 600
gaaggccggaa aacgacaatc ttatccat caagcttgcat atcgatcc ttccatcc 660
gggatccact agttcttagt cggccggcgtt aacttcgtt ccacggatcc ccgggttacc 720
agctcgaaat caaattttt atgtttttttttt ttccctttttt ccgggtttttt 780
tagttttttt aagactttttt gcttaaaattt gaaagaattt acttttaagaa aatcttaaca 840
ttctgatataa ttccatcaat agattttttt ttccatcaat ccacgtttt tttttttttt 900
caatcgcaac atatatggtt tttagaaaaaa atgcactata tatataatata ttatTTTT 960
aattttttttt gatgtatataa taatataatata atatataatata atgtgtgtt gtatgtgt 1020
aaagaaatcc ttatataatata atacacgaaat acatataatata gacaaaatca aagtattaca 1080
ctaaacaatg agtttttgca ttggccaaac aaatattgtat attaaaaa 1127

```

---

The invention claimed is:

**1-15.** (canceled)

**16.** A transgenic soybean plant cell comprising a modified MON87701 transgenic locus comprising the DNA molecule set forth in SEQ ID NO: 2.

**17.** A transgenic soybean plant part comprising the soybean plant cell of claim **16**.

**18.** The transgenic soybean plant part of claim **17**, wherein said soybean plant part is a seed.

**19.** A transgenic soybean plant comprising the soybean plant cell of claim **16**.

**20.** A method for obtaining a bulked population of inbred seed comprising selfing the transgenic soybean plant of

claim **19** and harvesting seed comprising the modified MON87701 transgenic locus from the selfed soybean plant.

**21.** A method of obtaining hybrid soybean seed comprising crossing the transgenic soybean plant of claim **19** to a second soybean plant which is genetically distinct from the first soybean plant and harvesting seed comprising the modified MON87701 transgenic locus from the cross.

**22.** A DNA molecule comprising the DNA sequence of SEQ ID NO: 2.

**23.** A processed transgenic soybean plant product comprising the DNA molecule of claim **22**.

\* \* \* \* \*