

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

(19) World Intellectual Property Organization

International Bureau



(10) International Publication Number

WO 2013/169802 A1

(43) International Publication Date
14 November 2013 (14.11.2013)

WIPO | PCT

(51) International Patent Classification:
C12N 15/87 (2006.01) *C12N 5/10* (2006.01)
C12N 15/00 (2006.01)

(74) Agent: PASTERNAK, Dahna, S.; Pasternak Patent Law, 1900 Embarcadero Rd., Suite 211, Palo Alto, CA 94303 (US).

(21) International Application Number:
PCT/US2013/039979

(22) International Filing Date:
7 May 2013 (07.05.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/643,812 7 May 2012 (07.05.2012) US

(71) Applicants: SANGAMO BIOSCIENCES, INC. [US/US]; Point Richmond Tech Center, 501 Canal Blvd., Suite A100, Richmond, CA 94804 (US). DOW AGROSCIENCES LLC [US/US]; 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US).

(72) Inventors: COST, Gregory, J.; c/o Sangamo BioSciences, Inc., Point Richmond Tech Center, 501 Canal Blvd., Suite A100, Richmond, CA 94804 (US). URNOV, Fyodor; c/o Sangamo BioSciences, Inc., Point Richmond Tech Center, 501 Canal Blvd., Suite A100, Richmond, CA 94804 (US). AINLEY, William M.; c/o Dow Agrosciences LLP, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). PETOLINO, Joseph F.; c/o Dow Agrosciences LLP, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). SAMUEL, Jayakumar Pon; c/o Dow Agrosciences LLP, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). WEBB, Steven R.; c/o Dow Agrosciences LLP, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US). SASTRY-DENT, Lakshmi; c/o Dow Agrosciences LLP, 9330 Zionsville Road, Indianapolis, IN 46268-1054 (US).

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

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

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(54) Title: METHODS AND COMPOSITIONS FOR NUCLEASE-MEDIATED TARGETED INTEGRATION OF TRANSGENES



FIG. 1A

(57) Abstract: Disclosed herein are methods and compositions for homology-independent targeted insertion of donor molecules into the genome of a cell.

WO 2013/169802 A1

**METHODS AND COMPOSITIONS FOR NUCLEASE-MEDIATED
TARGETED INTEGRATION OF TRANSGENES**

5

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims the benefit of U.S. Provisional Application No. 61/643,812, filed May 7, 2012, the disclosure of which is hereby incorporated by reference in its entirety.

10

STATEMENT OF RIGHTS TO INVENTIONS

MADE UNDER FEDERALLY SPONSORED RESEARCH

[0002] Not applicable.

TECHNICAL FIELD

15 [0003] The present disclosure is in the field of genome engineering, particularly targeted modification of the genome of a cell.

BACKGROUND

[0004] Integration of foreign DNA into the genome of organisms and cell lines is a widely utilized method for interrogation and manipulation of biological systems. Traditionally, transgene insertion is targeted to a specific locus by provision of a plasmid carrying a transgene, and containing substantial DNA sequence identity flanking the desired site of integration. Spontaneous breakage of the chromosome followed by repair using the homologous region of the plasmid DNA as a template results in the transfer of the intervening transgene into the genome. *See, e.g.*, Koller *et al.* (1989) *Proc. Nat'l. Acad. Sci. USA* 86(22):8927-8931; Thomas *et al.* (1986) *Cell* 44(3):419-428. The frequency of this type of homology-directed targeted integration can be increased by up to a factor of 10^5 by deliberate creation of a double-strand break in the vicinity of the target region (Hockemeyer *et al.* (2009) *Nature Biotech.* 27(9):851-857; Lombardo *et al.* (2007) *Nature Biotech.* 25(11):1298-1306; Moehle *et al.* (2007) *Proc. Nat'l Acad. Sci. USA* 104(9):3055-3060; Rouet *et al.* (1994) *Proc. Nat'l Acad. Sci. USA* 91(13):6064-6068.

[0005] A double-strand break (DSB) or nick for can be created by a site-specific nuclease such as a zinc-finger nuclease (ZFN) or TAL effector domain nuclease (TALEN), or using the CRISPR/Cas9 system with an engineered crRNA/tract RNA (single guide RNA) to guide specific cleavage. *See, for example, Burgess (2013)*

5 Nature Reviews Genetics 14:80-81, Urnov *et al.* (2010) *Nature* 435(7042):646-51; United States Patent Publications 20030232410; 20050208489; 20050026157; 20050064474; 20060188987; 20090263900; 20090117617; 20100047805; 20110207221; 20110301073 and International Publication WO 2007/014275, the disclosures of which are incorporated by reference in their entireties for all purposes.

10 In many organisms, transgene insertion can be accomplished via homology-directed repair (HDR) processes, which require the inserted transgene to include regions of homology to the site of insertion (cleavage). However, some organisms and cell lines lack traditional HDR process and targeted integration occurs primarily via the homology-independent non-homologous end joining (NHEJ) DNA repair machinery.

15 As such, to date, in organisms and cell lines (*e.g.*, CHO cells) that are recalcitrant to HDR processes, only relatively short (<100 bp) oligonucleotides have been integrated via homology-independent pathways following nuclease-mediated cleavage of the target locus. *See, e.g., Orlando et al. (2010) Nucleic Acids Res. 38(15):e152 and U.S. Patent Publication No. 20110207221.*

20 [0006] Thus, there remains a need for compositions and methods for homology-independent targeted integration of transgenes, including larger transgenes, directly into the site of cleavage, for example in organisms and cell lines that lack, or are deficient in, traditional homology-driven approaches.

25

SUMMARY

[0007] Disclosed herein are methods and compositions for homology-independent targeted integration of a transgene.

30 [0008] In one aspect, described herein are double-stranded donor polynucleotides for integration into an endogenous locus of choice following *in vivo* cleavage of the donor using at least one nuclease. The donor polynucleotides include an exogenous sequence (transgene) to be integrated into the endogenous locus and contain at least one target site for a nuclease, for example two paired nuclease binding sites separated by a “spacer” sequence separating near edges of binding sites. The

spacer can be of any size, for example, between 4 and 20 base pairs (or any value therebetween). Donors having multiple nuclease target sites may have the same or different target sites, for example, two of the same paired sites flanking the transgene or two different paired sites flanking the transgene. The donor nucleotides do not require the presence of homology arms flanking the transgene sequence. The only chromosomal homology that may be present in the donor sequence is(are) the nuclease binding site(s). In embodiments in which the nuclease target sites exhibit homology to the genome, the homology to the genome is less than 50 to 100 (or any number of base pairs between 50 and 100) contiguous base pairs in length. In certain embodiments, where the nuclease used to cleave the donor is not the same as the nuclease used to cleave the chromosome, there may be no homology between the chromosomal locus cleaved by the nuclease(s) and the donor sequence. In addition, the nuclease target site(s) are not within the transgene and, as such, cleavage of the donor polynucleotide by the nuclease(s) that bind(s) to the target site(s) does not modify the transgene. In certain embodiments, the donor nucleic acid comprises two target sites and the spacer sequence between the two target sites is non-naturally occurring, for example when the spacer sequence does not occur in a genomic sequence between the two target sites present in the genome. In certain embodiments, the donor molecules are integrated into the endogenous locus via homology-independent mechanisms (e.g., NHEJ). In other embodiments, the double-stranded donor comprises a transgene of at least 1 kb in length and nuclease target site(s) 3' and/or 5' of the transgene for *in vivo* cleavage. In certain embodiments, the nuclease target site(s) used to cleave the donor are not re-created upon integration of the transgene, for example when the spacer between paired target sites is not present in and/or does not exhibit homology to an endogenous locus. The donor molecule may be, for example, a plasmid. In certain embodiments, the donor is integrated following nuclease-mediated cleavage of the endogenous locus. In any nuclease-mediated integration of the donor molecule, the one or more of the nucleases used to cleave the donor may be the same as one or more of the nucleases used to cleave the endogenous locus. Alternatively, one or more of the nucleases used to cleave the donor may be different from one or more of the nucleases used to cleave the endogenous locus.

[0009] In some embodiments, the donor is contained on a plasmid. The donor may be integrated following nuclease-mediated cleavage where the sequence to be

integrated (donor or transgene) is flanked in the plasmid by at least two nuclease cleavage sites. In other embodiments, the donor is contained on a plasmid, wherein the donor may be integrated following nuclease-mediated cleavage where the sequence to be integrated (donor or transgene) is the plasmid comprising a single nuclease cleavage site. In certain embodiments, the sequence of the nuclease cleavage sites in the donor plasmid is the same as the sequence of the nuclease cleavage site in the chromosomal locus to be targeted. In embodiments in which the cleavage sites are the same as between the donor and the genome, the sequences separating the cleavage sites may be the same or different. In certain embodiments, the sequences separating the cleavage sites (spacers) are different in the donor as compared to the genome such that following cleavage of the donor, the target sites is(are) not re-created and the donor cannot be cleaved again by the same nuclease(s). In other embodiments, the nuclease cleavage sites flanking the donor on the donor-containing plasmid are different from the cleavage site in the chromosome. In further embodiments, the nuclease cleavage sites flanking the donor in the donor-containing plasmid are not the same, and also may be different from the nuclease cleavage site in the chromosome. In further embodiments, the donor may be contained on a plasmid flanked by at least two nuclease cleavage sites and may be integrated into a deletion in the chromosome created by the action of two nucleases. In this embodiment, the nuclease cleavage sites flanking the donor on the plasmid and the nuclease cleavage sites in the chromosome may either be the same or may be different.

[0010] The sequence of interest of the donor molecule may comprise one or more sequences encoding a functional polypeptide (e.g., a cDNA), with or without a promoter. In certain embodiments, the nucleic acid sequence comprises a sequence encoding an antibody, an antigen, an enzyme, a growth factor, a receptor (cell surface or nuclear), a hormone, a lymphokine, a cytokine, a reporter, an insect resistant gene, a herbicide tolerance gene, a transcription factor, sequestration protein or functional fragments of any of the above and combinations of the above. The sequence of interest of the donor molecule may comprise one or more sequences that encode an RNA molecule that encodes a functional or structural RNA, for example, an RNAi, sRNAi, and/or mRNAi. In embodiments in which the functional polypeptide encoding sequences are promoterless, expression of the integrated sequence is then ensured by transcription driven by an endogenous promoter or other control element

in the region of interest. In other embodiments, a “tandem” cassette is integrated into the selected site in this manner, the first component of the cassette comprising a promoterless sequence as described above, followed by a transcription termination sequence, and a second sequence, encoding an autonomous expression cassette.

5 Additional sequences (coding or non-coding sequences) may be included in the donor molecule, including but not limited to, sequences encoding a 2A peptide, SA site, IRES, *etc.* In certain embodiments, the donor nucleic acid (transgene) comprises sequences encoding functional RNAs for example, miRNAs or shRNAs.

[0011] In another aspect, described herein are methods of integrating a donor

10 nucleic acid (*e.g.*, a donor molecule as described herein) into the genome of a cell via homology-independent mechanisms. The methods comprise creating a double-stranded break (DSB) in the genome of a cell and cleaving the donor molecule using one or more nucleases, such that the donor nucleic acid is integrated at the site of the DSB. In certain embodiments, the donor nucleic acid is integrated via non-homology 15 dependent methods (*e.g.*, NHEJ). As noted above, upon *in vivo* cleavage the donor sequences can be integrated in a targeted manner into the genome of a cell at the location of a DSB. The donor sequence can include one or more of the same target sites for one or more of the nucleases used to create the DSB. Thus, the donor sequence may be cleaved by one or more of the same nucleases used to cleave the

20 endogenous gene into which integration is desired. In certain embodiments, the donor sequence includes different nuclease target sites from the nucleases used to induce the DSB. DSBs in the genome of the target cell may be created by any mechanism. In certain embodiments, the DSB is created by one or more (*e.g.*, a dimerizing pair of) zinc-finger nucleases (ZFNs), fusion proteins comprising a zinc finger binding 25 domain, which is engineered to bind a sequence within the region of interest, and a cleavage domain or a cleavage half-domain. In other embodiments, the DSB is created by one or more TALE DNA-binding domains (naturally occurring or non-naturally occurring) fused to a nuclease domain (TALEN). In still further embodiments, cleavage is performed using a nuclease system such as CRISPR/Cas 30 with an engineered crRNA/tracr RNA.

[0012] Furthermore, in any of the methods described herein, the first and second cleavage half-domains may be from a Type IIS restriction endonuclease, for example, *FokI* or *StsI*. Furthermore, in any of the methods described herein, at least

one of the fusion proteins may comprise an alteration in the amino acid sequence of the dimerization interface of the cleavage half-domain, for example such that obligate heterodimers of the cleavage half-domains are formed. Alternatively, in any of the methods described herein the cleavage domain may be a naturally or non-naturally occurring (engineered) meganuclease.

[0013] In any of the methods described herein, the cell can be any eukaryotic cells, for example a plant cell or a mammalian cell or cell line, including COS, CHO (e.g., CHO-S, CHO-K1, CHO-DG44, CHO-DUXB11, CHO-DUKX, CHOK1SV), VERO, MDCK, WI38, V79, B14AF28-G3, BHK, HaK, NS0, SP2/0-Ag14, HeLa,

10 HEK293 (e.g., HEK293-F, HEK293-H, HEK293-T), and perC6 cells as well as insect cells such as *Spodoptera frugiperda* (Sf), or fungal cells such as *Saccharomyces*, *Pichia* and *Schizosaccharomyces*. In certain embodiments, the cell line is a CHO, MDCK or HEK293 cell line. Suitable cells also include stem cells such as, by way of example, embryonic stem cells, induced pluripotent stem cells, hematopoietic stem 15 cells, neuronal stem cells and mesenchymal stem cells. Furthermore, the cell may be arrested in the G2 phase of the cell cycle. In some embodiments of the methods described herein, the cell may be one lacking efficient homology-based DNA repair, for example a CHO cell. In certain embodiments, the cells may be primary or non-dividing cells which preferentially use the NHEJ DNA repair pathway. In some 20 embodiments, the cell can be a plant or fungal cell. In other embodiments, the methods described herein may be used in cells with unsequenced genomes. These cells can be used to create cell lines and/or transgenic organisms (e.g., animals or plants) bearing the transgene(s).

[0014] In another aspect, transgenic organisms (e.g., plants or animals) comprising a transgene integrated according to any of the methods described herein are provided. In one embodiment, a cell, cell line or transgenic organism carrying a heterozygous genotype for the selected gene is constructed, while in another embodiment, a homozygous cell, cell line or transgenic organism is made carrying two mutant copies in both alleles of a desired locus.

30 **[0015]** A kit, comprising the methods and compositions of the invention, is also provided. The kit may comprise the nucleases, (e.g. RNA molecules or ZFN, TALEN or CRISPR/Cas system encoding genes contained in a suitable expression vector), or aliquots of the nuclease proteins, donor molecules, suitable host cell lines,

instructions for performing the methods of the invention, and the like. The kit may also comprise donor molecules of interest (e.g. reporter genes, specific transgenes and the like).

[0016] These and other aspects will be readily apparent to the skilled artisan in light of disclosure as a whole.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] **Figure 1, panels A to F**, show capture of a transgene cleaved *in vivo* at an *AAVS1* locus in K562 cells. Figure 1A is a schematic depicting an exemplary donor molecule having two paired binding sites (4 binding sites total), with each pair separated by spacers), which sites flank a transgene to be integrated into the genome. The target sites may be the same or different. Figure 1B is a schematic depicting four different *in vivo* donor cleavage techniques. In the first embodiment, when both the chromosome and the donor plasmid contain a ZFN cleavage site (dark grey region), cleavage of the donor and chromosome are synchronized, allowing efficient integration of the donor into the chromosome. Integration can occur in both forward and reverse orientations, termed “AB” and “BA,” respectively. In the second embodiment, the donor contains more than one nuclease cleavage site. Nuclease action liberates a linear fragment of DNA which is integrated into the chromosome. In the third embodiment, the chromosome is cleaved by more than one nuclease and the donor by one nuclease, resulting in integration of the donor into a deletion in the chromosome. In the fourth embodiment, both the donor DNA and the chromosomal DNA are cleaved by more than one nuclease, resulting in integration of a linear fragment into a deletion in the chromosome. Figure 1C shows a comparison of donors that favor the forward (AB) integration of the transgene and re-create the same nuclease target site (top panels) with donors that favor reverse (BA) integration of the transgene and do not re-create the same nuclease target site (bottom panels). The sequence between the binding sites (spacers) is underlined and the overhangs of the wild-type (top panels) and reverse-complement (bottom panels) spacers created after cleavage are shown on the right of the top and bottom panels. The sequences shown in the top left are SEQ ID NOs:107 and 108. The sequences shown in the top right are SEQ ID NOs:109 and 110. The sequences shown in the bottom left are SEQ ID NOs:111 and 112 and the sequences shown in the bottom right are SEQ ID NOs:113

and 114. Figure 1D is a gel showing detection of targeted integration from donors with wild-type (“w.t.”) sequences (or spacers) between the ZFN binding sites (the same spacer sequences as in the genome) and reverse-complement (“r.c.”) sequences between the ZFN binding sites. As shown, more signal is seen in the BA orientation with reverse-complement overhang-nucleotides

5 Figure 1E is a gel depicting integration of a donor plasmid at *AAVS1* following *in vivo* cleavage by the *AAVS*-specific ZFNs. A donor plasmid either containing or lacking the *AAVS1* ZFN site was co-transfected with the *AAVS1*-specific ZFNs into K562 cells. Integration in both the AB and BA orientations was monitored by PCR of chromosome-donor junctions.

10 Insertion in the AB orientation produces 391 and 423 bp PCR products for the left and right junctions, respectively; insertion in the BA orientation produces 369 and 471 bp junction PCR products for the left and right junctions, respectively. “A1” refers to the PCR reaction designed to amplify the *AAVS1* ZFN site and “NT” refers to those reactions lacking DNA template. Figure 1F depicts both schematics of the triploid

15 *AAVS1* locus and results of interrogation of targeted integration, as assayed by Southern blot. Three clones were assayed in duplicate. Genomic DNA from the three clones was either cut with BglII and probed with an *AAVS1*-specific probe (top gel) or cut with AccI and probed with a transgene specific probe (*bla*).

[0018] **Figure 2, panels A and B**, show that *in vivo* donor cleavage promotes

20 transgene capture at several loci in two different cell types. Figure 2A shows that targeted integration via NHEJ at sites in the *IL2R γ* , *CCR5*, and glutamine synthetase (GS) loci is more efficient when the donor plasmid is co-cleaved with the chromosome rather than being cut prior to transfection. One junction PCR for each orientation was performed for all three loci, the left junction for the BA orientation

25 and the right junction for the AB orientation. The experimental conditions that were used are labeled as follows: “N” refers to transfections wherein the donor lacked a ZFN site; “ERV” refers to the sample in which the donor was pre-cut with EcoRV prior to introduction into the cell; “Y” refers to a transfection where the donor and the targeted gene both contained the ZFN site; and “NT” refers to PCR reactions with no

30 template DNA. The amplicon size expected from PCR amplification of successfully integrated donors is shown below each lane in base pairs (“Expected size, bp”). The picture shown is a color-inverted image of an ethidium bromide-stained gel. Figure 2B depicts that donor cleavage does not need to be done with the same ZFNs as those

used to cut the target site in the chromosome. Junction-specific PCR assays were used to detect transgene integration into the chromosomal target and to detect the orientation of the integrated transgenes. These assays demonstrated that the transgene could integrate in either orientation following ZFN cleavage. Experimental conditions 5 are labeled as follows: GS, GS-specific ZFNs or donor plasmid with the *GS* ZFN cleavage site; A1, *AAVS1*-specific ZFNs or donor with the *AAVS1*ZFN cleavage site.

[0019] **Figure 3, panels A to C,** depict high-frequency targeted transgene integration at the *GS* locus in CHO-K1 cells. Figure 3A is a schematic of the *GS* locus showing the transgene integrated in the BA orientation. Figure 3B shows a Southern blot assay of cell clones for targeted integration of the donor at the *GS* locus. The exonic *GS* probe also detects two *GS* pseudogenes. The same panel of clones was assayed for total transgene integration by probing for the *E. coli bla* gene (Figure 3C). Integration of the transgene at the *GS* locus is seen along with transgene integration elsewhere in the genome in three of the eight clones analyzed.

10 [0020] **Figure 4, panels A and C,** depict disruption of alpha-(1,6)-fucosyltransferase (*FUT8*) in CHO-K1 cells by ZFN- and TALEN-mediated targeted insertion of a monoclonal antibody transgene. Figure 4A depicts area-proportional Venn diagrams showing concordance between clones screened for transgene insertion by junction-specific PCR and for IgG expression. Figure 4B is a schematic of the 15 *FUT8* locus containing the inserted transgene (depicted in light grey), labeling the orientation of the transgene in either the “AB” or “BA” nomenclature, and Figure 4C shows Southern blot confirmation of integration at *FUT8*. Integrants containing the transgenes inserted in the BA and AB orientations are indicated.

20 [0021] **Figure 5, panels A to C,** show ZFN activity for experiments described in the Examples. Figure 5A shows ZFN cleavage at the *AAVS1* locus (top) and of the donor plasmid (bottom). The corresponding lanes in Figure 5A from Figure 1D are indicated above the gel (ex.: “1,12”), as is the presence (“Y”) or absence (“N”) of the Surveyor™ nuclease enzyme. The percentage of molecules modified is shown below the lanes with signal. Figure 5B shows ZFN cleavage at *IL2Rγ*, *CCR5*, and *GS* using 25 the corresponding gene-specific ZFNs. As described above for 5A, the corresponding lanes in 5B from Figure 2A are shown above the gel, as is the presence (“Y”) or absence (“N”) of the Surveyor™ nuclease enzyme. While the gels across the top of Figure 5B depict the results from integration into the gene loci, the gels on the bottom

of the figure depict the results of cleavage in the donors. Figure 5C shows ZFN cleavage at GS where the gel on the left depicts the results from the gene locus in CHO cells, while the gel on the right depicts the results of ZFN cleavage of the donor. As above, the corresponding lanes in Figure 5C from Figure 2B is shown above the gel, as is the presence ("Y") or absence ("N") of the Surveyor™ nuclease enzyme. 5 Arrows indicate the expected cleavage products.

[0022] **Figure 6** is a graph depicting homology-directed targeted integration of a GFP encoding transgene in HEK 293 and CHO-K1 cells. The percentage of cells that are GFP-positive is shown in light grey (HEK 293 cells) or dark grey (CHO-K1 cells). The amount of donor used is indicated below each grouping. 10

[0023] **Figure 7** shows partial DNA sequence of junction PCR products from CHO K1 clones into which *in vivo* cleaved donors were integrated into the *AAVS1* locus. Chromosomal sequence is shown in plain text, and donor sequence is shown in italics. ZFN binding sites are underlined and in bold. Microhomology is 15 shaded in grey. The expected allele sequences in the AB orientation are shown across the top of the AB group and are defined as perfect ligation of the 5' overhangs; the expected allele sequences in the BA orientation are shown across the top of the BA group and are defined as removal of the 5' overhangs followed by ligation. Sequence identifiers are indicated in the Figure.

[0024] **Figure 8, panels A to C**, show DNA sequences of junction PCR assays from CHO K1 cell pools with transgene integrations at *AAVS1* (Figure 8A), *CCR5* (Figure 8B), *GS* (Figure 8C), and *IL2R γ* (Figure 8C). Chromosomal sequence is shown in plain text, donor sequence is shown in italics. ZFN binding sites are underlined and in bold. The expected allele sequences are shown as above, and are 20 also as defined above for Figure 7. Identical sequences isolated more than once are indicated and sequence identifiers are indicated in the Figure.

[0025] **Figure 9** shows DNA sequences of junction PCRs from *GS* single cell-derived clones. Chromosomal sequence is shown in plain text, donor sequence is shown in italics. ZFN binding sites are underlined and in bold. The expected alleles in 30 the AB orientation are defined as perfect ligation of the 5' overhangs; the expected alleles in the BA orientation are defined as removal of the 5' overhangs followed by ligation. Sequence identifiers are indicated in the Figure.

[0026] **Figure 10, panels A and B,** show DNA sequences of junction PCRs from single cell-derived clones with integrations at *FUT8*. Chromosomal sequence is shown in plain text, donor sequence is shown in italics. Figure 10A shows integration following cleavage with *FUT8*-targeted ZFNs (ZFN binding sites are underlined and in bolded). Figure 10B shows integration following cleavage with *FUT8*-targeted TALENs (TALEN binding sites are underlined and bolded). Sequence identifiers are indicated in the Figure.

[0027] **Figure 11** shows a plasmid map of pDAB109350.

[0028] **Figure 12** shows a plasmid map of pDAB109360.

10 [0029] **Figure 13** shows a plasmid map of pDAS000153.

[0030] **Figure 14** shows a plasmid map of pDAS000150.

[0031] **Figure 15** shows a plasmid map of pDAS000143.

[0032] **Figure 16** shows a plasmid map of pDAS000164.

[0033] **Figure 17** shows a plasmid map of pDAS000433.

15 [0034] **Figure 18** shows a plasmid map of pDAS000434.

[0035] **Figure 19, panels A and B,** depict exogenous marker-free, sequential transgene stacking at an endogenous AHAS locus in the wheat genome of *Triticum aestivum* using ZFN-mediated, NHEJ-directed DNA repair. Figure 19A depicts the first transgene stack; Figure 19B depicts the second transgene stack.

20 [0036] **Figure 20, panels A and B,** depict exogenous marker-free, sequential transgene stacking at an endogenous AHAS locus in the wheat genome of *Triticum aestivum* using ZFN-mediated, HDR-directed DNA repair. Figure 20A depicts the first transgene stack; Figure 20B depicts the second transgene stack.

[0037] **Figure 21** shows a plasmid map of pDAS000435.

25 [0038] **Figure 22** shows a plasmid map of pDAB107827.

[0039] **Figure 23** shows a plasmid map of pDAB107828.

[0040] **Figure 24** shows a plasmid map of pDAS000340.

[0041] **Figure 25** shows a plasmid map of pDAS000341.

[0042] **Figure 26** shows a plasmid map of pDAS000342.

30 [0043] **Figure 27** shows a plasmid map of pDAS000343.

[0044] **Figure 28, panels A and B,** show the locations of the primers and their position relative to the start and stop codon of Fad3C. Figure 28A shows the location of the primer sites for the wild type Fad3C locus. Figure 28B shows the

location of the primer sites to confirm donor integration, and the possible orientations by which the donor could integrate within the Fad3C locus.

[0045] **Figure 29, panels A and B,** shows sequences alignments of various targeted integrations. Figure 29A shows a sequence alignment amplified from the junction of the tGFP cassette of pDAS000341 with Fad3C at the double strand break as recognized by ZFN 28051-2A-28052. Sequences shown are SEQ ID NOs:480 to 493 from top to bottom. The ":" indicates the deletions located at the cut sites. Figure 29B shows a sequence alignment amplified from the junction of the tGFP cassette of pDAS000343 with Fad3C at the double strand break as recognized by ZFN 28051-2A-28052 and ZFN 28053-2A-28054. The ":" indicates the deletions located at the cut sites. Sequences shown are SEQ ID NOs:494 to 507 from top to bottom.

[0046] **Figure 30, panels A and B,** show a sequence alignment amplified from the junction of the hph cassette of pDAS000340 with FAD3C at the double strand break as recognized by ZFN 28051-2A-28052. The ":" indicates the deletions located at the cut sites. Sequences shown are SEQ ID NOs:508 to 523 from top to bottom. Figure 30A shows sequences for the 5' junction and the sequences shown in Figure 30B are for the 3' junction.

[0047] **Figure 31, panels A and B,** show a sequence alignment amplified from the junction of the hph cassette of pDAS00034 with FAD3C at the double strand break as recognized by ZFN 28051-2A-28052 and 28053-2A-28054. The ":" indicates the deletions located at the cut sites. Sequences shown are SEQ ID NOs:524 to 532 from top to bottom. The sequences shown in Figure 31A are for the 5' junction and the sequences shown in Figure 31B are for the 3' junction.

[0048] **Figure 32** depicts the relation of the ZFNs designed to bind the genomic locus of transgenic insert in Corn Event DAS-59132. Six ZFNs (E32 ZFN1-6) were identified from the yeast assay and four ZFNs were advanced for evaluation in plants.

[0049] **Figure 33** shows a plasmid map of pDAB105906.

[0050] **Figure 34** shows a plasmid map of pDAB111809.

[0051] **Figure 35** shows a plasmid map of pDAB100655.

[0052] **Figure 36** depicts a graph showing evaluation of transiently expressed ZFNs in plants. Four ZFNs were evaluated in maize callus by transiently expressing the ZFNs and an internal control ZFN directed to the IPPK2 gene. After Next

Generation Sequencing of PCR amplified fragments from the region surrounding the ZFN cleavage sites, the sequenced PCR amplified fragments were scored for the presence of sequence variants resulting from indels. The relative frequency of indels from each of the four E32 ZFN pairs as compared to IPPK2 ZFN activity are depicted. Event 32 ZFN6 which contains the 25716 ('716) and 25717 ('717) zinc finger binding domains cleaved the genomic locus of transgenic Corn Event DAS-59132 at 380 times the efficiency of the control IPPK2 zinc finger nuclease.

5 [0053] **Figure 37** depicts a graph of the ZFN locus disruption of Corn Event DAS-59132.

10 [0054] **Figure 38** is a schematic depicting the experimental system used for donor integration into the ELP of maize genome.

[0055] **Figure 39** is a graph illustrating the cleavage of genomic target DNA by eZFNs. DNA was isolated from each treatment group (6 replicates each) as indicated. TAQMANTM assays were used to measure cleavage of the target DNA.

15 Cleavage activity of the eZFNs is relative to the Donor DNA alone treatments. eZFNs (eZFN1 and eZFN3) levels were 1:1 or 1:10 ratios relative to the Donor DNA. Statistical groupings are indicated by lower case letters.

[0056] **Figure 40** illustrates the primer binding sites within the ELP loci of the corn genome.

20 [0057] **Figure 41** illustrates the primer binding sites of the pDAB100651 fragment for copy number evaluation.

[0058] **Figure 42** shows the cleavage activity of the eZFNs is relative to the Donor DNA alone treatments. The eZFN (eZFN1 and eZFN3) cleavage levels were 1:1 or 1:10 ratios relative to the donor DNA. Statistical groupings are indicated by lower case letters.

25 [0059] **Figure 43, panels A and B**, show the junction sequence from in-out PCR reactions. The left and right sequences are partial sequences of the AAD1 and ELP, respectively. The sequence expected from an insertion restoring the eZFN binding site is shown in the blue font. The eZFN binding site is highlighted green and deletions are black bars. The sequence for the direct orientation (Figure 43A) and the reverse orientation (Figure 43B) are shown. The sequences are in blocks according to the PCR reaction from which they were cloned.

DETAILED DESCRIPTION

[0060] Disclosed herein are compositions and methods for nuclease-mediated homology-independent (*e.g.*, NHEJ capture) targeted integration of a transgene.

While insertion of oligonucleotides can be performed via simple co-transfection of

5 DNA with compatible 5' overhangs, it has now been shown that NHEJ capture of transgene-size fragments (*e.g.*, > 0.5 kb) is greatly facilitated by *in vivo* nuclease-mediated cleavage of the donor plasmid in addition to cleavage of the chromosome. In this way, transgenes of larger size (*e.g.*, between 1 and 14 kb or longer in length) can be integrated in a targeted manner into organisms and cell lines, such as Chinese
10 hamster ovary (CHO) cells, which are recalcitrant to HDR-based integration. For example, *in vivo* donor cleavage allowed targeted integration at high frequency (6%) in unselected CHO cells, a cell type otherwise recalcitrant to targeted insertion of large DNA sequences.

[0061] Co-cleavage of the chromosome and transgene-containing double-

15 stranded donor as described herein results in successful integration into any endogenous target locus in a selected host cell. The methods and compositions described herein allow for efficient non-homology-driven targeted integration that is not generally achievable by simple co-transfection of pre-cut donors.

[0062] Thus, the compositions and methods described herein allow for

20 homology-independent targeted integration of large transgenes into sites of nuclease-cleavage, including into deletions created by engineered nucleases such as ZFNs and/or TALENs. Alternately, a donor plasmid with nuclease sites flanking a transgene to be integrated can be used such that the transgene portion is liberated upon nuclease cleavage and efficiently integrated at a targeted location. Further, use
25 of the methods and compositions of the invention allow for nuclease-mediated *in vivo* cleavage of a large donor molecule such as a bacterial or yeast artificial chromosome permits the targeted integration of large transgenes in mammalian and plant cells.

Finally, the *in vivo* cleavage compositions and methods described will find use in the targeted genetic modification of other organisms and cells, especially those which

30 perform homology-direct DNA repair poorly.

General

[0063] Practice of the methods, as well as preparation and use of the compositions disclosed herein employ, unless otherwise indicated, conventional

techniques in molecular biology, biochemistry, chromatin structure and analysis, computational chemistry, cell culture, recombinant DNA and related fields as are within the skill of the art. These techniques are fully explained in the literature. *See*, for example, Sambrook *et al.* MOLECULAR CLONING: A LABORATORY MANUAL, 5 Second edition, Cold Spring Harbor Laboratory Press, 1989 and Third edition, 2001; Ausubel *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 and periodic updates; the series METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, Third edition, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol. 304, 10 "Chromatin" (P.M. Wassarman and A. P. Wolffe, eds.), Academic Press, San Diego, 1999; and METHODS IN MOLECULAR BIOLOGY, Vol. 119, "Chromatin Protocols" (P.B. Becker, ed.) Humana Press, Totowa, 1999.

Definitions

15 [0064] The terms "nucleic acid," "polynucleotide," and "oligonucleotide" are used interchangeably and refer to a deoxyribonucleotide or ribonucleotide polymer, in linear or circular conformation, and in either single- or double-stranded form. For the purposes of the present disclosure, these terms are not to be construed as limiting with respect to the length of a polymer. The terms can encompass known analogues of natural nucleotides, as well as nucleotides that are modified in the base, sugar and/or phosphate moieties (e.g., phosphorothioate backbones). In general, an analogue of a particular nucleotide has the same base-pairing specificity; *i.e.*, an analogue of A will base-pair with T.

20 [0065] The terms "polypeptide," "peptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues. The term also applies to amino acid polymers in which one or more amino acids are chemical analogues or modified derivatives of a corresponding naturally-occurring amino acids.

25 [0066] "Binding" refers to a sequence-specific, non-covalent interaction between macromolecules (e.g., between a protein and a nucleic acid). Not all 30 components of a binding interaction need be sequence-specific (e.g., contacts with phosphate residues in a DNA backbone), as long as the interaction as a whole is sequence-specific. Such interactions are generally characterized by a dissociation

constant (K_d) of 10^{-6} M⁻¹ or lower. "Affinity" refers to the strength of binding: increased binding affinity being correlated with a lower K_d .

[0067] A "binding protein" is a protein that is able to bind non-covalently to another molecule. A binding protein can bind to, for example, a DNA molecule (a DNA-binding protein), an RNA molecule (an RNA-binding protein) and/or a protein molecule (a protein-binding protein). In the case of a protein-binding protein, it can bind to itself (to form homodimers, homotrimers, *etc.*) and/or it can bind to one or more molecules of a different protein or proteins. A binding protein can have more than one type of binding activity. For example, zinc finger proteins have DNA-binding, RNA-binding and protein-binding activity.

[0068] A "zinc finger DNA binding protein" (or binding domain) is a protein, or a domain within a larger protein, that binds DNA in a sequence-specific manner through one or more zinc fingers, which are regions of amino acid sequence within the binding domain whose structure is stabilized through coordination of a zinc ion.

15 The term zinc finger DNA binding protein is often abbreviated as zinc finger protein or ZFP.

[0069] A "TALE DNA binding domain" or "TALE" is a polypeptide comprising one or more TALE repeat domains/units. The repeat domains are involved in binding of the TALE to its cognate target DNA sequence. A single "repeat unit" (also referred to as a 20 "repeat") is typically 33-35 amino acids in length and exhibits at least some sequence homology with other TALE repeat sequences within a naturally occurring TALE protein. *See, e.g.*, U.S. Patent Publication No. 20110301073, incorporated by reference herein in its entirety.

[0070] Zinc finger and TALE binding domains can be "engineered" to bind to 25 a predetermined nucleotide sequence, for example via engineering (altering one or more amino acids) of the recognition helix region of a naturally occurring zinc finger. Similarly, TALEs can be "engineered" to bind to a predetermined nucleotide sequence, for example by engineering of the amino acids involved in DNA binding (the repeat variable diresidue or RVD region). Therefore, engineered DNA binding 30 proteins (zinc fingers or TALEs) are proteins that are non-naturally occurring. Non-limiting examples of methods for engineering DNA-binding proteins are design and selection. A designed DNA binding protein is a protein not occurring in nature whose design/composition results principally from rational criteria. Rational criteria for

design include application of substitution rules and computerized algorithms for processing information in a database storing information of existing ZFP and/or TALE designs and binding data. See, for example, U.S. Patents 6,140,081; 6,453,242; and 6,534,261; see also WO 98/53058; WO 98/53059; WO 98/53060; 5 WO 02/016536 and WO 03/016496 and U.S. Publication Nos. 20110301073, 20110239315 and 20119145940.

[0071] A "selected" zinc finger protein or TALE is a protein not found in nature whose production results primarily from an empirical process such as phage display, interaction trap or hybrid selection. See *e.g.*, U.S. Patent Nos. 5,789,538; US 10 5,925,523; US 6,007,988; US 6,013,453; US 6,200,759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970 WO 01/88197 and WO 02/099084 and U.S. Publication Nos. 20110301073, 20110239315 and 20119145940.

[0072] "Recombination" refers to a process of exchange of genetic 15 information between two polynucleotides, including but not limited to, donor capture by non-homologous end joining (NHEJ) and homologous recombination. For the purposes of this disclosure, "homologous recombination (HR)" refers to the specialized form of such exchange that takes place, for example, during repair of double-strand breaks in cells via homology-directed repair mechanisms. This process 20 requires nucleotide sequence homology, uses a "donor" molecule to template repair of a "target" molecule (*i.e.*, the one that experienced the double-strand break), and is variously known as "non-crossover gene conversion" or "short tract gene conversion," because it leads to the transfer of genetic information from the donor to the target. Without wishing to be bound by any particular theory, such transfer can involve 25 mismatch correction of heteroduplex DNA that forms between the broken target and the donor, and/or "synthesis-dependent strand annealing," in which the donor is used to resynthesize genetic information that will become part of the target, and/or related processes. Such specialized HR often results in an alteration of the sequence of the target molecule such that part or all of the sequence of the donor polynucleotide is 30 incorporated into the target polynucleotide. For HR-directed integration, the donor molecule contains at least 2 regions of homology to the genome ("homology arms") of least 50-100 base pairs in length. *See, e.g.*, U.S. Patent Publication No. 20110281361.

[0073] In the methods of the disclosure, one or more targeted nucleases as described herein create a double-stranded break in the target sequence (e.g., cellular chromatin) at a predetermined site, and a “donor” polynucleotide, having homology to the nucleotide sequence in the region of the break, can be introduced into the cell.

5 The presence of the double-stranded break has been shown to facilitate integration of the donor sequence. The donor sequence may be physically integrated or, alternatively, the donor polynucleotide is used as a template for repair of the break via homologous recombination, resulting in the introduction of all or part of the nucleotide sequence as in the donor into the cellular chromatin. Thus, a first sequence
10 in cellular chromatin can be altered and, in certain embodiments, can be converted into a sequence present in a donor polynucleotide. Thus, the use of the terms “replace” or “replacement” can be understood to represent replacement of one nucleotide sequence by another, (i.e., replacement of a sequence in the informational sense), and does not necessarily require physical or chemical replacement of one
15 polynucleotide by another.

[0074] In any of the methods described herein, additional pairs of zinc-finger proteins or TALEN can be used for additional double-stranded cleavage of additional target sites within the cell.

20 **[0075]** Any of the methods described herein can be used for insertion of a donor of any size and/or partial or complete inactivation of one or more target sequences in a cell by targeted integration of donor sequence that disrupts expression of the gene(s) of interest. Cell lines with partially or completely inactivated genes are also provided.

25 **[0076]** Furthermore, the methods of targeted integration as described herein can also be used to integrate one or more exogenous sequences. The exogenous nucleic acid sequence can comprise, for example, one or more genes or cDNA molecules, or any type of coding or noncoding sequence, as well as one or more control elements (e.g., promoters). In addition, the exogenous nucleic acid sequence (transgene) may produce one or more RNA molecules (e.g., small hairpin RNAs
30 (shRNAs), inhibitory RNAs (RNAis), microRNAs (miRNAs), etc.).

[0077] “Cleavage” refers to the breakage of the covalent backbone of a DNA molecule. Cleavage can be initiated by a variety of methods including, but not limited to, enzymatic or chemical hydrolysis of a phosphodiester bond. Both single-stranded

cleavage and double-stranded cleavage are possible, and double-stranded cleavage can occur as a result of two distinct single-stranded cleavage events. DNA cleavage can result in the production of either blunt ends or staggered ends. In certain embodiments, fusion polypeptides are used for targeted double-stranded DNA

5 cleavage.

[0078] A "cleavage half-domain" is a polypeptide sequence which, in conjunction with a second polypeptide (either identical or different) forms a complex having cleavage activity (preferably double-strand cleavage activity). The terms "first and second cleavage half-domains;" "+ and – cleavage half-domains" and "right and 10 left cleavage half-domains" are used interchangeably to refer to pairs of cleavage half-domains that dimerize.

[0079] An "engineered cleavage half-domain" is a cleavage half-domain that has been modified so as to form obligate heterodimers with another cleavage half-domain (e.g., another engineered cleavage half-domain). *See, also,* U.S. Patent 15 Publication Nos. 2005/0064474, 20070218528 and 2008/0131962, incorporated herein by reference in their entireties.

[0080] The term "sequence" refers to a nucleotide sequence of any length, which can be DNA or RNA; can be linear, circular or branched and can be either single-stranded or double stranded. The term "donor sequence" or "transgene" refers 20 to a nucleotide sequence that is inserted into a genome. A donor sequence can be of any length, for example between 2 and 100,000,000 nucleotides in length (or any integer value therebetween or thereabove), preferably between about 100 and 100,000 nucleotides in length (or any integer therebetween), more preferably between about 2000 and 60,000 nucleotides in length (or any value therebetween) and even more 25 preferable, between about 3 and 15 kb (or any value therebetween).

[0081] "Chromatin" is the nucleoprotein structure comprising the cellular genome. Cellular chromatin comprises nucleic acid, primarily DNA, and protein, including histones and non-histone chromosomal proteins. The majority of eukaryotic cellular chromatin exists in the form of nucleosomes, wherein a 30 nucleosome core comprises approximately 150 base pairs of DNA associated with an octamer comprising two each of histones H2A, H2B, H3 and H4; and linker DNA (of variable length depending on the organism) extends between nucleosome cores. A molecule of histone H1 is generally associated with the linker DNA. For the purposes

of the present disclosure, the term "chromatin" is meant to encompass all types of cellular nucleoprotein, both prokaryotic and eukaryotic. Cellular chromatin includes both chromosomal and episomal chromatin.

[0082] A "chromosome," is a chromatin complex comprising all or a portion of the genome of a cell. The genome of a cell is often characterized by its karyotype, which is the collection of all the chromosomes that comprise the genome of the cell. The genome of a cell can comprise one or more chromosomes.

[0083] An "episome" is a replicating nucleic acid, nucleoprotein complex or other structure comprising a nucleic acid that is not part of the chromosomal karyotype of a cell. Examples of episomes include plasmids and certain viral genomes.

[0084] A "target site" or "target sequence" is a nucleic acid sequence that defines a portion of a nucleic acid to which a binding molecule will bind, provided sufficient conditions for binding exist.

[0085] An "exogenous" molecule is a molecule that is not normally present in a cell, but can be introduced into a cell by one or more genetic, biochemical or other methods. "Normal presence in the cell" is determined with respect to the particular developmental stage and environmental conditions of the cell. Thus, for example, a molecule that is present only during embryonic development of muscle is an exogenous molecule with respect to an adult muscle cell. Similarly, a molecule induced by heat shock is an exogenous molecule with respect to a non-heat-shocked cell. An exogenous molecule can comprise, for example, a functioning version of a malfunctioning endogenous molecule or a malfunctioning version of a normally-functioning endogenous molecule.

[0086] An exogenous molecule can be, among other things, a small molecule, such as is generated by a combinatorial chemistry process, or a macromolecule such as a protein, nucleic acid, carbohydrate, lipid, glycoprotein, lipoprotein, polysaccharide, any modified derivative of the above molecules, or any complex comprising one or more of the above molecules. Nucleic acids include DNA and RNA, can be single- or double-stranded; can be linear, branched or circular; and can be of any length. Nucleic acids include those capable of forming duplexes, as well as triplex-forming nucleic acids. See, for example, U.S. Patent Nos. 5,176,996 and 5,422,251. Proteins include, but are not limited to, DNA-binding proteins,

transcription factors, chromatin remodeling factors, methylated DNA binding proteins, polymerases, methylases, demethylases, acetylases, deacetylases, kinases, phosphatases, integrases, recombinases, ligases, topoisomerases, gyrases and helicases.

5 [0087] An exogenous molecule can be the same type of molecule as an endogenous molecule, *e.g.*, an exogenous protein or nucleic acid. For example, an exogenous nucleic acid can comprise an infecting viral genome, a plasmid or episome introduced into a cell, or a chromosome that is not normally present in the cell. Methods for the introduction of exogenous molecules into cells are known to those of
10 skill in the art and include, but are not limited to, lipid-mediated transfer (*i.e.*, liposomes, including neutral and cationic lipids), electroporation, direct injection, cell fusion, particle bombardment, calcium phosphate co-precipitation, DEAE-dextran-mediated transfer and viral vector-mediated transfer. An exogenous molecule can also be the same type of molecule as an endogenous molecule but derived from a different
15 species than the cell is derived from. For example, a human nucleic acid sequence may be introduced into a cell line originally derived from a mouse or hamster.

[0088] By contrast, an "endogenous" molecule is one that is normally present in a particular cell at a particular developmental stage under particular environmental conditions. For example, an endogenous nucleic acid can comprise a chromosome, the genome of a mitochondrion, chloroplast or other organelle, or a naturally-
20 occurring episomal nucleic acid. Additional endogenous molecules can include proteins, for example, transcription factors and enzymes.

[0089] A "fusion" molecule is a molecule in which two or more subunit molecules are linked, preferably covalently. The subunit molecules can be the same
25 chemical type of molecule, or can be different chemical types of molecules. Examples of the first type of fusion molecule include, but are not limited to, fusion proteins (for example, a fusion between a ZFP or TALE DNA-binding domain and one or more activation domains) and fusion nucleic acids (for example, a nucleic acid encoding the fusion protein described *supra*). Examples of the second type of fusion
30 molecule include, but are not limited to, a fusion between a triplex-forming nucleic acid and a polypeptide, and a fusion between a minor groove binder and a nucleic acid.

[0090] Expression of a fusion protein in a cell can result from delivery of the fusion protein to the cell or by delivery of a polynucleotide encoding the fusion protein to a cell, wherein the polynucleotide is transcribed, and the transcript is translated, to generate the fusion protein. Trans-splicing, polypeptide cleavage and 5 polypeptide ligation can also be involved in expression of a protein in a cell. Methods for polynucleotide and polypeptide delivery to cells are presented elsewhere in this disclosure.

[0091] A "gene," for the purposes of the present disclosure, includes a DNA region encoding a gene product (see *infra*), as well as all DNA regions which regulate 10 the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites 15 and locus control regions.

[0092] "Gene expression" refers to the conversion of the information, contained in a gene, into a gene product. A gene product can be the direct transcriptional product of a gene (e.g., mRNA, tRNA, rRNA, antisense RNA, ribozyme, structural RNA or any other type of RNA) or a protein produced by 20 translation of an mRNA. Gene products also include RNAs which are modified, by processes such as capping, polyadenylation, methylation, and editing, and proteins modified by, for example, methylation, acetylation, phosphorylation, ubiquitination, ADP-ribosylation, myristilation, and glycosylation.

[0093] "Modulation" of gene expression refers to a change in the activity of a 25 gene. Modulation of expression can include, but is not limited to, gene activation and gene repression. Genome editing (e.g., cleavage, alteration, inactivation, random mutation) can be used to modulate expression. Gene inactivation refers to any reduction in gene expression as compared to a cell that does not include a ZFP as described herein. Thus, gene inactivation may be partial or complete.

[0094] A "region of interest" is any region of cellular chromatin, such as, for 30 example, a gene or a non-coding sequence within or adjacent to a gene, in which it is desirable to bind an exogenous molecule. Binding can be for the purposes of targeted DNA cleavage and/or targeted recombination. A region of interest can be present in a

chromosome, an episome, an organellar genome (*e.g.*, mitochondrial, chloroplast), or an infecting viral genome, for example. A region of interest can be within the coding region of a gene, within transcribed non-coding regions such as, for example, leader sequences, trailer sequences or introns, or within non-transcribed regions, either upstream or downstream of the coding region. A region of interest can be as small as a single nucleotide pair or up to 2,000 nucleotide pairs in length, or any integral value of nucleotide pairs.

5 [0095] "Eukaryotic" cells include, but are not limited to, fungal cells (such as yeast), plant cells, animal cells, mammalian cells and human cells (*e.g.*, T-cells).

10 [0096] The terms "operative linkage" and "operatively linked" (or "operably linked") are used interchangeably with reference to a juxtaposition of two or more components (such as sequence elements), in which the components are arranged such that both components function normally and allow the possibility that at least one of the components can mediate a function that is exerted upon at least one of the other 15 components. By way of illustration, a transcriptional regulatory sequence, such as a promoter, is operatively linked to a coding sequence if the transcriptional regulatory sequence controls the level of transcription of the coding sequence in response to the presence or absence of one or more transcriptional regulatory factors. A transcriptional regulatory sequence is generally operatively linked in *cis* with a coding 20 sequence, but need not be directly adjacent to it. For example, an enhancer is a transcriptional regulatory sequence that is operatively linked to a coding sequence, even though they are not contiguous.

25 [0097] With respect to fusion polypeptides, the term "operatively linked" can refer to the fact that each of the components performs the same function in linkage to the other component as it would if it were not so linked. For example, with respect to a fusion polypeptide in which a ZFP, TALE or Cas DNA-binding domain is fused to an activation domain, the ZFP, TALE or Cas DNA-binding domain and the activation domain are in operative linkage if, in the fusion polypeptide, the ZFP, TALE or Cas DNA-binding domain portion is able to bind its target site and/or its binding site, 30 while the activation domain is able to upregulate gene expression. When a fusion polypeptide in which a ZFP, TALE or Cas DNA-binding domain is fused to a cleavage domain, the ZFP, TALE or Cas DNA-binding domain and the cleavage domain are in operative linkage if, in the fusion polypeptide, the ZFP, TALE or Cas

DNA-binding domain portion is able to bind its target site and/or its binding site, while the cleavage domain is able to cleave DNA in the vicinity of the target site.

[0098] A "functional fragment" of a protein, polypeptide or nucleic acid is a protein, polypeptide or nucleic acid whose sequence is not identical to the full-length protein, polypeptide or nucleic acid, yet retains the same function as the full-length protein, polypeptide or nucleic acid. A functional fragment can possess more, fewer, or the same number of residues as the corresponding native molecule, and/or can contain one or more amino acid or nucleotide substitutions. Methods for determining the function of a nucleic acid (*e.g.*, coding function, ability to hybridize to another nucleic acid) are well-known in the art. Similarly, methods for determining protein function are well-known. For example, the DNA-binding function of a polypeptide can be determined, for example, by filter-binding, electrophoretic mobility-shift, or immunoprecipitation assays. DNA cleavage can be assayed by gel electrophoresis. See Ausubel *et al.*, *supra*. The ability of a protein to interact with another protein can be determined, for example, by co-immunoprecipitation, two-hybrid assays or complementation, both genetic and biochemical. See, for example, Fields *et al.* (1989) *Nature* **340**:245-246; U.S. Patent No. 5,585,245 and PCT WO 98/44350.

[0099] A "vector" is capable of transferring gene sequences to target cells. Typically, "vector construct," "expression vector," and "gene transfer vector," mean any nucleic acid construct capable of directing the expression of a gene of interest and which can transfer gene sequences to target cells. Thus, the term includes cloning, and expression vehicles, as well as integrating vectors.

[0100] A "reporter gene" or "reporter sequence" refers to any sequence that produces a protein product that is easily measured, preferably although not necessarily in a routine assay. Suitable reporter genes include, but are not limited to, sequences encoding proteins that mediate antibiotic resistance (*e.g.*, ampicillin resistance, neomycin resistance, G418 resistance, puromycin resistance), sequences encoding colored or fluorescent or luminescent proteins (*e.g.*, green fluorescent protein, enhanced green fluorescent protein, red fluorescent protein, luciferase), and proteins which mediate enhanced cell growth and/or gene amplification (*e.g.*, dihydrofolate reductase). Epitope tags include, for example, one or more copies of FLAG, His, myc, Tap, HA or any detectable amino acid sequence. "Expression tags" include

sequences that encode reporters that may be operably linked to a desired gene sequence in order to monitor expression of the gene of interest.

[0101] A “safe harbor” locus is a locus within the genome wherein a gene may be inserted without any deleterious effects on the host cell. Most beneficial is a 5 safe harbor locus in which expression of the inserted gene sequence is not perturbed by any read-through expression from neighboring genes. Non-limiting examples of safe harbor loci in mammalian cells are the AAVS1 gene (see U.S. Patent No. 8,110,379), the CCR5 gene (see U.S. Publication No. 20080159996), the Rosa locus (see WO 2010/065123) and/or the albumin locus (U.S. Application No. 13/624,193). 10 Non-limiting examples of safe harbor loci in plant cells are the ZP15 locus (U.S. Patent No 8,329,986)

Nucleases

[0102] Described herein are compositions, particularly nucleases, that are 15 useful for *in vivo* cleavage of a donor molecule carrying a transgene and nucleases for cleavage of the genome of a cell such that the transgene is integrated into the genome in a targeted manner. In certain embodiments, one or more of the nucleases are naturally occurring. In other embodiments, one or more of the nucleases are non-naturally occurring, *i.e.*, engineered in the DNA-binding domain and/or cleavage 20 domain. For example, the DNA-binding domain of a naturally-occurring nuclease may be altered to bind to a selected target site (*e.g.*, a meganuclease that has been engineered to bind to site different than the cognate binding site). In other embodiments, the nuclease comprises heterologous DNA-binding and cleavage domains (*e.g.*, zinc finger nucleases; TAL-effector domain DNA binding proteins; 25 meganuclease DNA-binding domains with heterologous cleavage domains).

A. DNA-binding domains

[0103] In certain embodiments, the composition and methods described herein employ a meganuclease (homing endonuclease) DNA-binding domain for binding to 30 the donor molecule and/or binding to the region of interest in the genome of the cell. Naturally-occurring meganucleases recognize 15-40 base-pair cleavage sites and are commonly grouped into four families: the LAGLIDADG family, the GIY-YIG family, the His-Cyst box family and the HNH family. Exemplary homing

endonucleases include I-SceI, I-CeuI, PI-PspI, PI-Sce, I-SceIV, I-CsmI, I-PanI, I-SceII, I-PpoI, I-SceIII, I-CreI, I-TevI, I-TevII and I-TevIII. Their recognition sequences are known. See also U.S. Patent No. 5,420,032; U.S. Patent No. 6,833,252; Belfort *et al.* (1997) *Nucleic Acids Res.* **25**:3379–3388; Dujon *et al.* (1989) *Gene* **82**:115–118; Perler *et al.* (1994) *Nucleic Acids Res.* **22**, 1125–1127; Jasen (1996) *Trends Genet.* **12**:224–228; Gimble *et al.* (1996) *J. Mol. Biol.* **263**:163–180; Argast *et al.* (1998) *J. Mol. Biol.* **280**:345–353 and the New England Biolabs catalogue.

[0104] In certain embodiments, the methods and compositions described herein make use of a nuclease that comprises an engineered (non-naturally occurring) homing endonuclease (meganuclease). The recognition sequences of homing endonucleases and meganucleases such as I-SceI, I-CeuI, PI-PspI, PI-Sce, I-SceIV, I-CsmI, I-PanI, I-SceII, I-PpoI, I-SceIII, I-CreI, I-TevI, I-TevII and I-TevIII are known. See also U.S. Patent No. 5,420,032; U.S. Patent No. 6,833,252; Belfort *et al.* (1997) *Nucleic Acids Res.* **25**:3379–3388; Dujon *et al.* (1989) *Gene* **82**:115–118; Perler *et al.* (1994) *Nucleic Acids Res.* **22**, 1125–1127; Jasen (1996) *Trends Genet.* **12**:224–228; Gimble *et al.* (1996) *J. Mol. Biol.* **263**:163–180; Argast *et al.* (1998) *J. Mol. Biol.* **280**:345–353 and the New England Biolabs catalogue. In addition, the DNA-binding specificity of homing endonucleases and meganucleases can be engineered to bind non-natural target sites. See, for example, Chevalier *et al.* (2002) *Molec. Cell* **10**:895–905; Epinat *et al.* (2003) *Nucleic Acids Res.* **31**:2952–2962; Ashworth *et al.* (2006) *Nature* **441**:656–659; Paques *et al.* (2007) *Current Gene Therapy* **7**:49–66; U.S. Patent Publication No. 20070117128. The DNA-binding domains of the homing endonucleases and meganucleases may be altered in the context of the nuclease as a whole (*i.e.*, such that the nuclease includes the cognate cleavage domain) or may be fused to a heterologous cleavage domain.

[0105] In other embodiments, the DNA-binding domain of one or more of the nucleases used in the methods and compositions described herein comprises a naturally occurring or engineered (non-naturally occurring) TAL effector DNA binding domain. See, *e.g.*, U.S. Patent Publication No. 20110301073, incorporated by reference in its entirety herein. The plant pathogenic bacteria of the genus *Xanthomonas* are known to cause many diseases in important crop plants. Pathogenicity of *Xanthomonas* depends on a conserved type III secretion (T3S)

system which injects more than 25 different effector proteins into the plant cell. Among these injected proteins are transcription activator-like (TAL) effectors which mimic plant transcriptional activators and manipulate the plant transcriptome (see Kay *et al* (2007) *Science* 318:648-651). These proteins contain a DNA binding domain and a transcriptional activation domain. One of the most well characterized TAL-effectors is AvrBs3 from *Xanthomonas campestris* pv. *Vesicatoria* (see Bonas *et al* (1989) *Mol Gen Genet* 218: 127-136 and WO2010079430). TAL-effectors contain a centralized domain of tandem repeats, each repeat containing approximately 34 amino acids, which are key to the DNA binding specificity of these proteins. In addition, 10 they contain a nuclear localization sequence and an acidic transcriptional activation domain (for a review see Schornack S, *et al* (2006) *J Plant Physiol* 163(3): 256-272). In addition, in the phytopathogenic bacteria *Ralstonia solanacearum* two genes, designated brg11 and hpx17 have been found that are homologous to the AvrBs3 family of *Xanthomonas* in the *R. solanacearum* biovar 1 strain GMI1000 and in the 15 biovar 4 strain RS1000 (See Heuer *et al* (2007) *Appl and Envir Micro* 73(13): 4379-4384). These genes are 98.9% identical in nucleotide sequence to each other but differ by a deletion of 1,575 bp in the repeat domain of hpx17. However, both gene products have less than 40% sequence identity with AvrBs3 family proteins of *Xanthomonas*. See, e.g., U.S. Patent Publication Nos. 20110239315, 20110145940 and 20 20110301073, incorporated by reference in its entirety herein.

[0106] Specificity of these TAL effectors depends on the sequences found in the tandem repeats. The repeated sequence comprises approximately 102 bp and the repeats are typically 91-100% homologous with each other (Bonas *et al*, *ibid*). Polymorphism of the repeats is usually located at positions 12 and 13 and there 25 appears to be a one-to-one correspondence between the identity of the hypervariable diresidues at positions 12 and 13 with the identity of the contiguous nucleotides in the TAL-effector's target sequence (see Moscou and Bogdanove, (2009) *Science* 326:1501 and Boch *et al* (2009) *Science* 326:1509-1512). Experimentally, the natural code for DNA recognition of these TAL-effectors has been determined such that an 30 HD sequence at positions 12 and 13 leads to a binding to cytosine (C), NG binds to T, NI to A, C, G or T, NN binds to A or G, and ING binds to T. These DNA binding repeats have been assembled into proteins with new combinations and numbers of repeats, to make artificial transcription factors that are able to interact with new

sequences and activate the expression of a non-endogenous reporter gene in plant cells (Boch *et al, ibid*). Engineered TAL proteins have been linked to a *FokI* cleavage half domain to yield a TAL effector domain nuclease fusion (TALEN) exhibiting activity in a yeast reporter assay (plasmid based target). *See, e.g.*, U.S. Patent 5 Publication No. 20110301073; Christian *et al* ((2010)<*Genetics* epub 10.1534/genetics.110.120717).

[0107] In other embodiments, the nuclease is a system comprising the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas (CRISPR Associated) nuclease system. The CRISPR/Cas is an engineered nuclease system

10 based on a bacterial system that can be used for genome engineering. It is based on part of the adaptive immune response of many bacteria and archaea. When a virus or plasmid invades a bacterium, segments of the invader's DNA are converted into CRISPR RNAs (crRNA) by the 'immune' response. This crRNA then associates, through a region of partial complementarity, with another type of RNA called

15 tracrRNA to guide the Cas9 nuclease to a region homologous to the crRNA in the target DNA called a "protospacer". Cas9 cleaves the DNA to generate blunt ends at the DSB at sites specified by a 20-nucleotide guide sequence contained within the crRNA transcript. Cas9 requires both the crRNA and the tracrRNA for site specific DNA recognition and cleavage. This system has now been engineered such that the

20 crRNA and tracrRNA can be combined into one molecule (the "single guide RNA"), and the crRNA equivalent portion of the single guide RNA can be engineered to guide the Cas9 nuclease to target any desired sequence (see Jinek *et al* (2012) *Science* 337, p. 816-821, Jinek *et al*, (2013), *eLife* 2:e00471, and David Segal, (2013) *eLife* 2:e00563). Thus, the CRISPR/Cas system can be engineered to create a DSB at a 25 desired target in a genome, and repair of the DSB can be influenced by the use of repair inhibitors to cause an increase in error prone repair.

[0108] In certain embodiments, the DNA binding domain of one or more of the nucleases used for *in vivo* cleavage and/or targeted cleavage of the genome of a cell comprises a zinc finger protein. Preferably, the zinc finger protein is non-

30 naturally occurring in that it is engineered to bind to a target site of choice. *See, for example, See, for example, Beerli *et al.* (2002) *Nature Biotechnol.* 20:135-141; Pabo *et al.* (2001) *Ann. Rev. Biochem.* 70:313-340; Isalan *et al.* (2001) *Nature Biotechnol.* 19:656-660; Segal *et al.* (2001) *Curr. Opin. Biotechnol.* 12:632-637; Choo *et al.**

(2000) *Curr. Opin. Struct. Biol.* **10**:411-416; U.S. Patent Nos. 6,453,242; 6,534,261; 6,599,692; 6,503,717; 6,689,558; 7,030,215; 6,794,136; 7,067,317; 7,262,054; 7,070,934; 7,361,635; 7,253,273; and U.S. Patent Publication Nos. 2005/0064474; 2007/0218528; 2005/0267061, all incorporated herein by reference in their entireties.

5 [0109] An engineered zinc finger binding domain can have a novel binding specificity, compared to a naturally-occurring zinc finger protein. Engineering methods include, but are not limited to, rational design and various types of selection. Rational design includes, for example, using databases comprising triplet (or quadruplet) nucleotide sequences and individual zinc finger amino acid sequences, in
10 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, for example, co-owned U.S. Patents 6,453,242 and 6,534,261, incorporated by reference herein in their entireties.

15 [0110] Exemplary selection methods, including phage display and two-hybrid systems, are disclosed in US Patents 5,789,538; 5,925,523; 6,007,988; 6,013,453; 6,410,248; 6,140,466; 6,200,759; and 6,242,568; as well as WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 and GB 2,338,237. In addition, enhancement of binding specificity for zinc finger binding domains has been described, for example, in co-owned WO 02/077227.

20 [0111] In addition, as disclosed in these and other references, zinc finger domains and/or multi-fingered zinc finger proteins may be linked together using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Patent Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described
25 herein may include any combination of suitable linkers between the individual zinc fingers of the protein.

[0112] Selection of target sites; ZFPs and methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to those of skill in the art and described in detail in U.S. Patent Nos. 6,140,081; 5,789,538; 6,453,242; 6,534,261; 5,925,523; 6,007,988; 6,013,453; 6,200,759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970; WO 01/88197; WO 02/099084; WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 and WO 03/016496.

[0113] In addition, as disclosed in these and other references, zinc finger domains and/or multi-fingered zinc finger proteins may be linked together using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Patent Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual zinc fingers of the protein.

[0114] The CRISPR (clustered regularly interspaced short palindromic repeats) locus, which encodes RNA components of the system, and the cas (CRISPR-associated) locus, which encodes proteins (Jansen *et al.*, 2002. *Mol. Microbiol.* 43: 1565-1575; Makarova *et al.*, 2002. *Nucleic Acids Res.* 30: 482-496; Makarova *et al.*, 2006. *Biol. Direct* 1: 7; Haft *et al.*, 2005. *PLoS Comput. Biol.* 1: e60) make up the gene sequences of the CRISPR/Cas nuclease system. CRISPR loci in microbial hosts contain a combination of CRISPR-associated (Cas) genes as well as non-coding RNA elements capable of programming the specificity of the CRISPR-mediated nucleic acid cleavage.

[0115] The Type II CRISPR is one of the most well characterized systems and carries out targeted DNA double-strand break in four sequential steps. First, two non-coding RNA, the pre-crRNA array and tracrRNA, are transcribed from the CRISPR locus. Second, tracrRNA hybridizes to the repeat regions of the pre-crRNA and mediates the processing of pre-crRNA into mature crRNAs containing individual spacer sequences. Third, the mature crRNA:tracrRNA complex directs Cas9 to the target DNA via Watson-Crick base-pairing between the spacer on the crRNA and the protospacer on the target DNA next to the protospacer adjacent motif (PAM), an additional requirement for target recognition. Finally, Cas9 mediates cleavage of target DNA to create a double-stranded break within the protospacer. Activity of the CRISPR/Cas system comprises of three steps: (i) insertion of alien DNA sequences into the CRISPR array to prevent future attacks, in a process called ‘adaptation’, (ii) expression of the relevant proteins, as well as expression and processing of the array, followed by (iii) RNA-mediated interference with the alien nucleic acid. Thus, in the bacterial cell, several of the so-called ‘Cas’ proteins are involved with the natural function of the CRISPR/Cas system and serve roles in functions such as insertion of the alien DNA etc.

[0116] In certain embodiments, Cas protein may be a "functional derivative" of a naturally occurring Cas protein. A "functional derivative" of a native sequence polypeptide is a compound having a qualitative biological property in common with a native sequence polypeptide. "Functional derivatives" include, but are not limited to, 5 fragments of a native sequence and derivatives of a native sequence polypeptide and its fragments, provided that they have a biological activity in common with a corresponding native sequence polypeptide. A biological activity contemplated herein is the ability of the functional derivative to hydrolyze a DNA substrate into fragments. The term "derivative" encompasses both amino acid sequence variants of polypeptide, 10 covalent modifications, and fusions thereof. Suitable derivatives of a Cas polypeptide or a fragment thereof include but are not limited to mutants, fusions, covalent modifications of Cas protein or a fragment thereof. Cas protein, which includes Cas protein or a fragment thereof, as well as derivatives of Cas protein or a fragment thereof, may be obtainable from a cell or synthesized chemically or by a combination 15 of these two procedures. The cell may be a cell that naturally produces Cas protein, or a cell that naturally produces Cas protein and is genetically engineered to produce the endogenous Cas protein at a higher expression level or to produce a Cas protein from an exogenously introduced nucleic acid, which nucleic acid encodes a Cas that is same or different from the endogenous Cas. In some case, the cell does not naturally 20 produce Cas protein and is genetically engineered to produce a Cas protein.

[0117] Thus, the nuclease comprises a DNA-binding domain in that specifically binds to a target site in any gene into which it is desired to insert a donor (transgene).

25 B. Cleavage Domains

[0118] Any suitable cleavage domain can be operatively linked to a DNA-binding domain to form a nuclease. For example, ZFP DNA-binding domains have been fused to nuclease domains to create ZFNs – a functional entity that is able to recognize its intended nucleic acid target through its engineered (ZFP) DNA binding 30 domain and cause the DNA to be cut near the ZFP binding site via the nuclease activity. See, e.g., Kim *et al.* (1996) *Proc Natl Acad Sci USA* 93(3):1156-1160. More recently, ZFNs have been used for genome modification in a variety of organisms. See, for example, United States Patent Publications 20030232410; 20050208489;

20050026157; 20050064474; 20060188987; 20060063231; and International Publication WO 07/014275. Likewise, TALE DNA-binding domains have been fused to nuclease domains to create TALENs. *See, e.g.*, U.S. Publication No. 20110301073.

[0119] As noted above, the cleavage domain may be heterologous to the 5 DNA-binding domain, for example a zinc finger DNA-binding domain and a cleavage domain from a nuclease or a TALEN DNA-binding domain and a cleavage domain, or meganuclease DNA-binding domain and cleavage domain from a different nuclease. Heterologous cleavage domains can be obtained from any endonuclease or exonuclease. Exemplary endonucleases from which a cleavage domain can be 10 derived include, but are not limited to, restriction endonucleases and homing endonucleases. *See*, for example, 2002-2003 Catalogue, New England Biolabs, Beverly, MA; and Belfort *et al.* (1997) *Nucleic Acids Res.* **25**:3379-3388. Additional enzymes which cleave DNA are known (*e.g.*, S1 Nuclease; mung bean nuclease; pancreatic DNase I; micrococcal nuclease; yeast HO endonuclease; *see also* Linn *et* 15 *al.* (eds.) *Nucleases*, Cold Spring Harbor Laboratory Press, 1993). One or more of these enzymes (or functional fragments thereof) can be used as a source of cleavage domains and cleavage half-domains.

[0120] Similarly, a cleavage half-domain can be derived from any nuclease or portion thereof, as set forth above, that requires dimerization for cleavage activity. In 20 general, two fusion proteins are required for cleavage if the fusion proteins comprise cleavage half-domains. Alternatively, a single protein comprising two cleavage half-domains can be used. The two cleavage half-domains can be derived from the same endonuclease (or functional fragments thereof), or each cleavage half-domain can be derived from a different endonuclease (or functional fragments thereof). In addition, 25 the target sites for the two fusion proteins are preferably disposed, with respect to each other, such that binding of the two fusion proteins to their respective target sites places the cleavage half-domains in a spatial orientation to each other that allows the cleavage half-domains to form a functional cleavage domain, *e.g.*, by dimerizing. Thus, in certain embodiments, the near edges of the target sites are separated by 5-8 30 nucleotides or by 15-18 nucleotides. However any integral number of nucleotides or nucleotide pairs can intervene between two target sites (*e.g.*, from 2 to 50 nucleotide pairs or more). In general, the site of cleavage lies between the target sites.

[0121] Restriction endonucleases (restriction enzymes) are present in many species and are capable of sequence-specific binding to DNA (at a recognition site), and cleaving DNA at or near the site of binding. Certain restriction enzymes (e.g., Type IIS) cleave DNA at sites removed from the recognition site and have separable binding and cleavage domains. For example, the Type IIS enzyme *Fok* I catalyzes double-stranded cleavage of DNA, at 9 nucleotides from its recognition site on one strand and 13 nucleotides from its recognition site on the other. *See, for example, US Patents 5,356,802; 5,436,150 and 5,487,994; as well as Li et al. (1992) Proc. Natl. Acad. Sci. USA* **89**:4275-4279; *Li et al. (1993) Proc. Natl. Acad. Sci. USA* **90**:2764-2768; *Kim et al. (1994a) Proc. Natl. Acad. Sci. USA* **91**:883-887; *Kim et al. (1994b) J. Biol. Chem.* **269**:31,978-31,982. Thus, in one embodiment, fusion proteins comprise the cleavage domain (or cleavage half-domain) from at least one Type IIS restriction enzyme and one or more zinc finger binding domains, which may or may not be engineered.

[0122] An exemplary Type IIS restriction enzyme, whose cleavage domain is separable from the binding domain, is *Fok* I. This particular enzyme is active as a dimer. *Bitinaite et al. (1998) Proc. Natl. Acad. Sci. USA* **95**: 10,570-10,575.

Accordingly, for the purposes of the present disclosure, the portion of the *Fok* I enzyme used in the disclosed fusion proteins is considered a cleavage half-domain.

Thus, for targeted double-stranded cleavage and/or targeted replacement of cellular sequences using zinc finger-*Fok* I fusions, two fusion proteins, each comprising a *Fok* I cleavage half-domain, can be used to reconstitute a catalytically active cleavage domain. Alternatively, a single polypeptide molecule containing a zinc finger binding domain and two *Fok* I cleavage half-domains can also be used. Parameters for targeted cleavage and targeted sequence alteration using zinc finger-*Fok* I fusions are provided elsewhere in this disclosure.

[0123] A cleavage domain or cleavage half-domain can be any portion of a protein that retains cleavage activity, or that retains the ability to multimerize (e.g., dimerize) to form a functional cleavage domain.

[0124] Exemplary Type IIS restriction enzymes are described in International Publication WO 07/014275, incorporated herein in its entirety. Additional restriction enzymes also contain separable binding and cleavage domains, and these are

contemplated by the present disclosure. *See, for example, Roberts *et al.* (2003) Nucleic Acids Res. 31:418-420.*

[0125] In certain embodiments, the cleavage domain comprises one or more engineered cleavage half-domain (also referred to as dimerization domain mutants) that minimize or prevent homodimerization, as described, for example, in U.S. Patent Publication Nos. 20050064474; 20060188987; 20070305346 and 20080131962, the disclosures of all of which are incorporated by reference in their entireties herein. Amino acid residues at positions 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537, and 538 of *Fok I* are all targets for influencing dimerization of the *Fok I* cleavage half-domains.

[0126] Exemplary engineered cleavage half-domains of *Fok I* that form obligate heterodimers include a pair in which a first cleavage half-domain includes mutations at amino acid residues at positions 490 and 538 of *Fok I* and a second cleavage half-domain includes mutations at amino acid residues 486 and 499.

[0127] Thus, in one embodiment, a mutation at 490 replaces Glu (E) with Lys (K); the mutation at 538 replaces Iso (I) with Lys (K); the mutation at 486 replaced Gln (Q) with Glu (E); and the mutation at position 499 replaces Iso (I) with Lys (K). Specifically, the engineered cleavage half-domains described herein were prepared by mutating positions 490 (E→K) and 538 (I→K) in one cleavage half-domain to produce an engineered cleavage half-domain designated “E490K:I538K” and by mutating positions 486 (Q→E) and 499 (I→L) in another cleavage half-domain to produce an engineered cleavage half-domain designated “Q486E:I499L”. The engineered cleavage half-domains described herein are obligate heterodimer mutants in which aberrant cleavage is minimized or abolished. *See, e.g., U.S. Patent Publication No. 2008/0131962, the disclosure of which is incorporated by reference in its entirety for all purposes.* In certain embodiments, the engineered cleavage half-domain comprises mutations at positions 486, 499 and 496 (numbered relative to wild-type *FokI*), for instance mutations that replace the wild type Gln (Q) residue at position 486 with a Glu (E) residue, the wild type Iso (I) residue at position 499 with a Leu (L) residue and the wild-type Asn (N) residue at position 496 with an Asp (D) or Glu (E) residue (also referred to as a “ELD” and “ELE” domains, respectively). In other embodiments, the engineered cleavage half-domain comprises mutations at positions 490, 538 and 537 (numbered relative to wild-type *FokI*), for instance

mutations that replace the wild type Glu (E) residue at position 490 with a Lys (K) residue, the wild type Iso (I) residue at position 538 with a Lys (K) residue, and the wild-type His (H) residue at position 537 with a Lys (K) residue or a Arg (R) residue (also referred to as “KKK” and “KKR” domains, respectively). In other 5 embodiments, the engineered cleavage half-domain comprises mutations at positions 490 and 537 (numbered relative to wild-type FokI), for instance mutations that replace the wild type Glu (E) residue at position 490 with a Lys (K) residue and the wild-type His (H) residue at position 537 with a Lys (K) residue or a Arg (R) residue (also referred to as “KIK” and “KIR” domains, respectively). (See US Patent 10 Publication No. 20110201055). In other embodiments, the engineered cleavage half domain comprises the “Sharkey” and/or “Sharkey” mutations (see Guo *et al*, (2010) *J. Mol. Biol.* 400(1):96-107).

[0128] Engineered cleavage half-domains described herein can be prepared using any suitable method, for example, by site-directed mutagenesis of wild-type 15 cleavage half-domains (*Fok I*) as described in U.S. Patent Publication Nos. 20050064474; 20080131962; and 20110201055.

[0129] Alternatively, nucleases may be assembled *in vivo* at the nucleic acid 20 target site using so-called “split-enzyme” technology (see *e.g.* U.S. Patent Publication No. 20090068164). Components of such split enzymes may be expressed either on separate expression constructs, or can be linked in one open reading frame where the individual components are separated, for example, by a self-cleaving 2A peptide or IRES sequence. Components may be individual zinc finger binding domains or domains of a meganuclease nucleic acid binding domain.

[0130] Nucleases can be screened for activity prior to use, for example in a 25 yeast-based chromosomal system as described in WO 2009/042163 and 20090068164. Nuclease expression constructs can be readily designed using methods known in the art. See, *e.g.*, United States Patent Publications 20030232410; 20050208489; 20050026157; 20050064474; 20060188987; 20060063231; and International Publication WO 07/014275. Expression of the nuclease may be under 30 the control of a constitutive promoter or an inducible promoter, for example the galactokinase promoter which is activated (de-repressed) in the presence of raffinose and/or galactose and repressed in presence of glucose.

[0131] The Cas9 related CRISPR/Cas system comprises two RNA non-coding components: tracrRNA and a pre-crRNA array containing nuclease guide sequences (spacers) interspaced by identical direct repeats (DRs). To use a CRISPR/Cas system to accomplish genome engineering, both functions of these RNAs must be present
5 (see Cong *et al*, (2013) *Scienceexpress* 1/10.1126/science 1231143). In some embodiments, the tracrRNA and pre-crRNAs are supplied via separate expression constructs or as separate RNAs. In other embodiments, a chimeric RNA is constructed where an engineered mature crRNA (conferring target specificity) is fused to a tracrRNA (supplying interaction with the Cas9) to create a chimeric cr-
10 RNA-tracrRNA hybrid (also termed a single guide RNA). (see Jinek *ibid* and Cong, *ibid*).

Target Sites

[0132] As described in detail above, DNA domains can be engineered to bind
15 to any sequence of choice. An engineered DNA-binding domain can have a novel binding specificity, compared to a naturally-occurring DNA-binding domain. Engineering methods include, but are not limited to, rational design and various types of selection. Rational design includes, for example, using databases comprising triplet (or quadruplet) nucleotide sequences and individual zinc finger amino acid
20 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*, for example, co-owned U.S. Patents 6,453,242 and 6,534,261, incorporated by reference herein in their entireties. Rational design of TAL-effector domains can also be performed. *See, e.g.*, U.S. Publication No.
25 20110301073.

[0133] Exemplary selection methods applicable to DNA-binding domains, including phage display and two-hybrid systems, are disclosed in US Patents 5,789,538; 5,925,523; 6,007,988; 6,013,453; 6,410,248; 6,140,466; 6,200,759; and 6,242,568; as well as WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197
30 and GB 2,338,237. In addition, enhancement of binding specificity for zinc finger binding domains has been described, for example, in co-owned WO 02/077227.

[0134] Selection of target sites; nucleases and methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to

those of skill in the art and described in detail in U.S. Patent Application Publication Nos. 20050064474 and 20060188987, incorporated by reference in their entireties herein.

[0135] In addition, as disclosed in these and other references, DNA-binding domains (e.g., multi-fingered zinc finger proteins) may be linked together using any suitable linker sequences, including for example, linkers of 5 or more amino acids. See, e.g., U.S. Patent Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual DNA-binding domains of the protein. See, also, U.S. Publication No. 20110301073.

[0136] As noted above, the DNA-binding domains of the nucleases may be targeted to any gene. In certain embodiments, the nuclease (DNA-binding domain component) is targeted to a “safe harbor” locus, which includes, by way of example only, the AAVS1 gene (see U.S. Patent No. 8,110,379), the CCR5 gene (see U.S. Publication No. 20080159996), the Rosa locus (see WO 2010/065123) and/or the albumin locus (see, U.S. Application No. 13/624,193).

Donors

[0137] Described herein are methods of targeted insertion of any polynucleotides for insertion into a chosen location. Polynucleotides for insertion can also be referred to as “exogenous” polynucleotides, “donor” polynucleotides or molecules or “transgenes.”

[0138] Surprisingly, it is demonstrated herein that double-stranded donor nucleotides (e.g., plasmids) without homology arms flanking the exogenous sequence (transgene) can be effectively integrated into a selected target region of the genome of cell following *in vivo* cleavage of the double-stranded donor. Thus, the double-stranded donors include one or more nuclease binding sites for cleavage of the donor *in vivo* (in the cell). In certain embodiments, the donor includes two nuclease binding sites. In methods in which targeted integration is achieved by making a double-stranded cut in the target region of the genome (see, e.g., U.S. Patent Nos. 7,888,121; 7,951,925; 8,110,379 and U.S. Patent Publication Nos. 20090263900; 20100129869 and 20110207221), one or more of then nucleases used to cleave the target region may also be used to cleave the donor molecule.

[0139] In certain embodiments, the double-stranded donor includes sequences (e.g., coding sequences, also referred to as transgenes) greater than 1 kb in length, for example between 2 and 200 kb, between 2 and 10 kb (or any value therebetween).

The double-stranded donor also includes at least one nuclease target site, for example.

5 In certain embodiments, the donor includes at least 2 target sites, for example for a pair of ZFNs or TALENs. Typically, the nuclease target sites are outside the transgene sequences, for example, 5' and/or 3' to the transgene sequences, for cleavage of the transgene. The nuclease cleavage site(s) may be for any nuclease(s). In certain embodiments, the nuclease target site(s) contained in the double-stranded

10 donor are for the same nuclease(s) used to cleave the endogenous target into which the cleaved donor is integrated via homology-independent methods.

[0140] As noted above, the donor can be cleaved *in vivo* and integrated into the genome in a forward (“AB”) or in a reverse (“BA”) orientation. Targeted integration via *in vivo* donor cleavage that results in a perfectly ligated AB-orientation insertion will recreate the paired nuclease (e.g., ZFN or TALEN) binding sites with the original spacing between the sites. Such recreated sites are potential substrates for a second round of cleavage by the nucleases. Nuclease cleavage at the recreated sites could result in DNA deletion at the transgene-chromosome junctions (as a result of inaccurate NHEJ-based repair) or even transgene excision. In contrast, reverse (BA) orientation insertions result in formation of two different nuclease pair binding sites (e.g., homodimers of the left and right nucleases). If obligate heterodimer (EL/KK, ELD/KKR, etc.) FokI nuclease domains are used, recreated BA sites will not be re-cleavable since the recreated binding sites are both homodimer sites. *See, also, Figure 1C.*

25 **[0141]** Furthermore, changing the nucleotides in the transgene donor nuclease spacer that make up the single-strand 5' overhang as compared to the wild-type (genomic) sequence, to the reverse complement of the wild-type sequence favors BA-orientation insertion of the cleaved donor (via Watson-Crick base-pairing with the overhangs on the cleaved chromosome) which would create an un-recleavable

30 transgene integration (Figure 1C).

[0142] The transgenes carried on the donor sequences described herein may be isolated from plasmids, cells or other sources using standard techniques known in the art such as PCR. Donors for use can include varying types of topology, including

circular supercoiled, circular relaxed, linear and the like. Alternatively, they may be chemically synthesized using standard oligonucleotide synthesis techniques. In addition, donors may be methylated or lack methylation. Donors may be in the form of bacterial or yeast artificial chromosomes (BACs or YACs).

5 [0143] The double-stranded donor polynucleotides described herein may include one or more non-natural bases and/or backbones. In particular, insertion of a donor molecule with methylated cytosines may be carried out using the methods described herein to achieve a state of transcriptional quiescence in a region of interest.

10 [0144] The exogenous (donor) polynucleotide may comprise any sequence of interest (exogenous sequence). Exemplary exogenous sequences include, but are not limited to any polypeptide coding sequence (e.g., cDNAs), promoter sequences, enhancer sequences, epitope tags, marker genes, cleavage enzyme recognition sites and various types of expression constructs. Marker genes include, but are not limited to, sequences encoding proteins that mediate antibiotic resistance (e.g., ampicillin 15 resistance, neomycin resistance, G418 resistance, puromycin resistance), sequences encoding colored or fluorescent or luminescent proteins (e.g., green fluorescent protein, enhanced green fluorescent protein, red fluorescent protein, luciferase), and proteins which mediate enhanced cell growth and/or gene amplification (e.g., dihydrofolate reductase). Epitope tags include, for example, one or more copies of 20 FLAG, His, myc, Tap, HA or any detectable amino acid sequence.

25 [0145] In a preferred embodiment, the exogenous sequence (transgene) comprises a polynucleotide encoding any polypeptide of which expression in the cell is desired, including, but not limited to antibodies, antigens, enzymes, receptors (cell surface or nuclear), hormones, lymphokines, cytokines, reporter polypeptides, growth factors, insect resistant, transcription factors and functional fragments of any of the above. The coding sequences may be, for example, cDNAs.

30 [0146] For example, the exogenous sequence may comprise a sequence encoding a polypeptide that is lacking or non-functional in the subject having a genetic disease, including but not limited to any of the following genetic diseases: achondroplasia, achromatopsia, acid maltase deficiency, adenosine deaminase deficiency (OMIM No.102700), adrenoleukodystrophy, aicardi syndrome, alpha-1 antitrypsin deficiency, alpha-thalassemia, androgen insensitivity syndrome, apert syndrome, arrhythmogenic right ventricular, dysplasia, ataxia telangiectasia, barth

syndrome, beta-thalassemia, blue rubber bleb nevus syndrome, canavan disease, chronic granulomatous diseases (CGD), cri du chat syndrome, cystic fibrosis, dercum's disease, ectodermal dysplasia, fanconi anemia, fibrodysplasia ossificans progressive, fragile X syndrome, galactosemis, Gaucher's disease, generalized 5 gangliosidoses (e.g., GM1), hemochromatosis, the hemoglobin C mutation in the 6th codon of beta-globin (HbC), hemophilia, Huntington's disease, Hurler Syndrome, hypophosphatasia, Klinefelter syndrome, Krabbes Disease, Langer-Giedion Syndrome, leukocyte adhesion deficiency (LAD, OMIM No. 116920), leukodystrophy, long QT syndrome, Marfan syndrome, Moebius syndrome, 10 mucopolysaccharidosis (MPS), nail patella syndrome, nephrogenic diabetes insipidus, neurofibromatosis, Neimann-Pick disease, osteogenesis imperfecta, porphyria, Prader-Willi syndrome, progeria, Proteus syndrome, retinoblastoma, Rett syndrome, Rubinstein-Taybi syndrome, Sanfilippo syndrome, severe combined 15 immunodeficiency (SCID), Shwachman syndrome, sickle cell disease (sickle cell anemia), Smith-Magenis syndrome, Stickler syndrome, Tay-Sachs disease, Thrombocytopenia Absent Radius (TAR) syndrome, Treacher Collins syndrome, trisomy, tuberous sclerosis, Turner's syndrome, urea cycle disorder, von Hippel-Landau disease, Waardenburg syndrome, Williams syndrome, Wilson's disease, Wiskott-Aldrich syndrome, X-linked lymphoproliferative syndrome (XLP, OMIM 20 No. 308240).

[0147] Additional exemplary diseases that can be treated by targeted integration include acquired immunodeficiencies, lysosomal storage diseases (e.g., Gaucher's disease, GM1, Fabry disease and Tay-Sachs disease), mucopolysaccharidosis (e.g. Hunter's disease, Hurler's disease), hemoglobinopathies (e.g., sickle cell diseases, HbC, α -thalassemia, β -thalassemia) and hemophilias. 25

[0148] In certain embodiments, the exogenous sequences can comprise a marker gene (described above), allowing selection of cells that have undergone targeted integration, and a linked sequence encoding an additional functionality. Non-limiting examples of marker genes include GFP, drug selection marker(s) and 30 the like.

[0149] Additional gene sequences that can be inserted may include, for example, wild-type genes to replace mutated sequences. For example, a wild-type Factor IX gene sequence may be inserted into the genome of a stem cell in which the

endogenous copy of the gene is mutated. The wild-type copy may be inserted at the endogenous locus, or may alternatively be targeted to a safe harbor locus.

[0150] In some embodiments, the exogenous nucleic acid sequence (transgene) comprising an agronomic gene or nucleotide sequence encoding a polypeptide of interest may include, for example and without limitation: a gene that confers resistance to a pest or disease (See, e.g., Jones *et al.* (1994) *Science* 266:789 (cloning of the tomato Cf-9 gene for resistance to *Cladosporium fulvum*); Martin *et al.* (1993) *Science* 262:1432; Mindrinos *et al.* (1994) *Cell* 78:1089 (*RSP2* gene for resistance to *Pseudomonas syringae*); PCT International Patent Publication No. WO 96/30517 (resistance to soybean cyst nematode); PCT International Patent Publication No. WO 93/19181); a gene that encodes a *Bacillus thuringiensis* protein, a derivative thereof, or a synthetic polypeptide modeled thereon (See, e.g., Geiser *et al.* (1986) *Gene* 48:109 (cloning and nucleotide sequence of a Bt δ-endotoxin gene; moreover, DNA molecules encoding δ-endotoxin genes can be purchased from American Type Culture Collection (Manassas, VA), for example, under ATCC Accession Nos. 40098; 67136; 31995; and 31998)); a gene that encodes a lectin (See, e.g., Van Damme *et al.* (1994) *Plant Molec. Biol.* 24:25 (nucleotide sequences of several *Clivia miniata* mannose-binding lectin genes)); a gene that encodes a vitamin-binding protein, e.g., avidin (See PCT International Patent Publication No. US93/06487 (use of avidin and avidin homologues as larvicides against insect pests)); a gene that encodes an enzyme inhibitor, e.g., a protease, proteinase inhibitor, or amylase inhibitor (See, e.g., Abe *et al.* (1987) *J. Biol. Chem.* 262:16793 (nucleotide sequence of rice cysteine proteinase inhibitor); Huub *et al.* (1993) *Plant Molec. Biol.* 21:985 (nucleotide sequence of cDNA encoding tobacco proteinase inhibitor I); Sumitani *et al.* (1993) *Biosci. Biotech. Biochem.* 57:1243 (nucleotide sequence of *Streptomyces nitrosporeus* alpha-amylase inhibitor) and U.S. Patent 5,494,813); a gene encoding an insect-specific hormone or pheromone, e.g., an ecdysteroid or juvenile hormone, a variant thereof, a mimetic based thereon, or an antagonist or agonist thereof (See, e.g., Hammock *et al.* (1990) *Nature* 344:458 (baculovirus expression of cloned juvenile hormone esterase, an inactivator of juvenile hormone)); a gene encoding an insect-specific peptide or neuropeptide that, upon expression, disrupts the physiology of the affected pest (See, e.g., Regan (1994) *J. Biol. Chem.* 269:9 (expression cloning yields DNA coding for insect diuretic hormone receptor); Pratt *et al.* (1989) *Biochem.*

Biophys. Res. Comm. 163:1243 (an allostatin in *Diploptera punctata*); and U.S. Patent 5,266,317 (genes encoding insect-specific, paralytic neurotoxins)); a gene encoding an insect-specific venom produced in nature by a snake, a wasp, or other organism (See, e.g., Pang *et al.* (1992) Gene 116:165 (heterologous expression in plants of a gene coding for a scorpion insectotoxic peptide)); a gene encoding an enzyme responsible for a hyperaccumulation of a monoterpane, a sesquiterpene, a steroid, hydroxamic acid, a phenylpropanoid derivative or other molecule with insecticidal activity; a gene encoding an enzyme involved in the modification, including the post-translational modification, of a biologically active molecule, e.g., a glycolytic enzyme, a proteolytic enzyme, a lipolytic enzyme, a nuclease, a cyclase, a transaminase, an esterase, a hydrolase, a phosphatase, a kinase, a phosphorylase, a polymerase, an elastase, a chitinase, or a glucanase, whether natural or synthetic (See, e.g., PCT International Patent Publication No. WO 93/02197 (nucleotide sequence of a callase gene); moreover, DNA molecules containing chitinase-encoding sequences can be obtained, for example, from the ATCC, under Accession Nos. 39637 and 67152; Kramer *et al.* (1993) Insect Biochem. Molec. Biol. 23:691 (nucleotide sequence of a cDNA encoding tobacco hornworm chitinase); and Kawalleck *et al.* (1993) Plant Molec. Biol. 21:673 (nucleotide sequence of the parsley ubi4-2 polyubiquitin gene)); a gene encoding a molecule that stimulates signal transduction (See, e.g., Botella *et al.* (1994) Plant Molec. Biol. 24:757 (nucleotide sequences for mung bean calmodulin cDNA clones); and Griess *et al.* (1994) Plant Physiol. 104:1467 (nucleotide sequence of a maize calmodulin cDNA clone)); a gene that encodes a hydrophobic moment peptide (See, e.g., PCT International Patent Publication No. WO 95/16776 (peptide derivatives of Tachyplesin which inhibit fungal plant pathogens); and PCT International Patent Publication No. WO 95/18855 (synthetic antimicrobial peptides that confer disease resistance)); a gene that encodes a membrane permease, a channel former, or a channel blocker (See, e.g., Jaynes *et al.* (1993) Plant Sci 89:43 (heterologous expression of a cecropin- β lytic peptide analog to render transgenic tobacco plants resistant to *Pseudomonas solanacearum*)); a gene that encodes a viral-invasive protein or complex toxin derived therefrom (See, e.g., Beachy *et al.* (1990) Ann. rev. Phytopathol. 28:451); a gene that encodes an insect-specific antibody or immunotoxin derived therefrom (See, e.g., Taylor *et al.*, Abstract #497, Seventh Int'l Symposium on Molecular Plant-Microbe Interactions (Edinburgh,

Scotland) (1994) (enzymatic inactivation in transgenic tobacco via production of single-chain antibody fragments)); a gene encoding a virus-specific antibody (*See, e.g.*, Tavladoraki *et al.* (1993) *Nature* 366:469 (transgenic plants expressing recombinant antibody genes are protected from virus attack)); a gene encoding a 5 developmental-arrestive protein produced in nature by a pathogen or a parasite (*See, e.g.*, Lamb *et al.* (1992) *Bio/Technology* 10:1436 (fungal endo α -1,4-D-polygalacturonases facilitate fungal colonization and plant nutrient release by solubilizing plant cell wall homo- α -1,4-D-galacturonase); Toubart *et al.* (1992) *Plant J.* 2:367 (cloning and characterization of a gene which encodes a bean 10 endopolygalacturonase-inhibiting protein)); a gene encoding a developmental-arrestive protein produced in nature by a plant (*See, e.g.*, Logemann *et al.* (1992) *Bio/Technology* 10:305 (transgenic plants expressing the barley ribosome-inactivating gene have an increased resistance to fungal disease)).

[0151] In some embodiments, nucleic acids comprising an agronomic gene or 15 nucleotide sequence encoding a polypeptide of interest may also and/or alternatively include, for example and without limitation: genes that confer resistance to an herbicide, such as an herbicide that inhibits the growing point or meristem, for example, an imidazolinone or a sulfonylurea (exemplary genes in this category encode mutant ALS and AHAS enzymes, as described, for example, by Lee *et al.* 20 (1988) *EMBO J.* 7:1241, and Miki *et al.* (1990) *Theor. Appl. Genet.* 80:449, respectively); glyphosate resistance as conferred by, *e.g.*, mutant 5-enolpyruvylshikimate-3-phosphate synthase (EPSPs) genes (*via* the introduction of recombinant nucleic acids and/or various forms of *in vivo* mutagenesis of native 25 EPSPs genes); *aroA* genes and glyphosate acetyl transferase (GAT) genes, respectively); other phosphono compounds, such as glufosinate phosphinothricin acetyl transferase (PAT) genes from *Streptomyces* species, including *Streptomyces hygroscopicus* and *Streptomyces viridichromogenes*); and pyridinoxy or phenoxy propionic acids and cyclohexones (ACCase inhibitor-encoding genes). *See, e.g.*, U.S. Patents 4,940,835 and 6,248,876 (nucleotide sequences of forms of EPSPs which 30 can confer glyphosate resistance to a plant). A DNA molecule encoding a mutant *aroA* gene can be obtained under ATCC accession number 39256. *See also* U.S. Pat. No. 4,769,061 (nucleotide sequence of a mutant *aroA* gene). European patent application No. 0 333 033 and U.S. Pat. No. 4,975,374 disclose nucleotide sequences

of glutamine synthetase genes, which may confer resistance to herbicides such as L-phosphinothricin. Nucleotide sequences of exemplary PAT genes are provided in European application No. 0 242 246, and DeGreef *et al.* (1989) Bio/Technology 7:61 (production of transgenic plants that express chimeric bar genes coding for PAT activity). Exemplary of genes conferring resistance to phenoxy propionic acids and cyclohexones, such as sethoxydim and haloxyfop, include the Acc1-S1, Acc1-S2 and Acc1-S3 genes described by Marshall *et al.* (1992) Theor. Appl. Genet. 83:435. GAT genes capable of conferring glyphosate resistance are described, for example, in WO 2005012515. Genes conferring resistance to 2,4-D, phenoxypropionic acid and 10 pyridyloxy auxin herbicides are described, for example, in WO 2005107437.

15 [0152] Nucleic acids comprising an agronomic gene or nucleotide sequence encoding a polypeptide of interest may also include, for example and without limitation: a gene conferring resistance to an herbicide that inhibits photosynthesis, such as a triazine (psbA and gs+ genes) or a benzonitrile (nitrilase gene). *See, e.g.,* Przibila *et al.* (1991) Plant Cell 3:169 (transformation of *Chlamydomonas* with plasmids encoding mutant psbA genes). Nucleotide sequences for nitrilase genes are disclosed in U.S. Patent 4,810,648, and DNA molecules containing these genes are available under ATCC Accession Nos. 53435; 67441; and 67442. *See also* Hayes *et al.* (1992) Biochem. J. 285:173 (cloning and expression of DNA coding for a 20 glutathione S-transferase).

25 [0153] In some embodiments, nucleic acids comprising an agronomic gene or nucleotide sequence encoding a polypeptide of interest may also and/or alternatively include, genes that confer or contribute to a value-added trait, for example and without limitation: modified fatty acid metabolism, *e.g.*, by transforming a plant with an antisense gene of stearyl-ACP desaturase to increase stearic acid content of the plant (*See, e.g.,* Knultzon *et al.* (1992) Proc. Natl. Acad. Sci. U.S.A. 89:2624); decreased phytate content, *e.g.*, introduction of a phytase-encoding gene may enhance breakdown of phytate, adding more free phosphate to the transformed plant (*See, e.g.,* Van Hartingsveldt *et al.* (1993) Gene 127:87 (nucleotide sequence of an *Aspergillus niger* phytase gene); a gene may be introduced to reduce phytate content- in maize, for example, this may be accomplished by cloning and then reintroducing DNA associated with the single allele which may be responsible for maize mutants characterized by low levels of phytic acid (*See* Raboy *et al.* (1990) Maydica 35:383));

and modified carbohydrate composition effected, *e.g.*, by transforming plants with a gene encoding an enzyme that alters the branching pattern of starch (*See, e.g.*, Shiroza *et al.* (1988) *J. Bacteol.* 170:810 (nucleotide sequence of *Streptococcus* mutant fructosyltransferase gene); Steinmetz *et al.* (1985) *Mol. Gen. Genet.* 20:220 (levansucrase gene); Pen *et al.* (1992) *Bio/Technology* 10:292 (α -amylase); Elliot *et al.* (1993) *Plant Molec. Biol.* 21:515 (nucleotide sequences of tomato invertase genes); Sogaard *et al.* (1993) *J. Biol. Chem.* 268:22480 (barley α -amylase gene); and Fisher *et al.* (1993) *Plant Physiol.* 102:1045 (maize endosperm starch branching enzyme II)).

10 [0154] Construction of such expression cassettes, following the teachings of the present specification, utilizes methodologies well known in the art of molecular biology (see, for example, Ausubel or Maniatis). Before use of the expression cassette to generate a transgenic animal, the responsiveness of the expression cassette to the stress-inducer associated with selected control elements can be tested by introducing 15 the expression cassette into a suitable cell line (*e.g.*, primary cells, transformed cells, or immortalized cell lines).

20 [0155] Furthermore, although not required for expression, exogenous sequences (transgenes) may also include transcriptional or translational regulatory sequences, for example, promoters, enhancers, insulators, internal ribosome entry sites, sequences encoding 2A peptides and/or polyadenylation signals. Further, the control elements of the genes of interest can be operably linked to reporter genes to create chimeric genes (*e.g.*, reporter expression cassettes).

25 [0156] Targeted insertion of a transgene of non-coding nucleic acid sequence may also be achieved. Transgenes encoding antisense RNAs, RNAi, shRNAs and micro RNAs (miRNAs) may also be used for targeted insertions.

30 [0157] In additional embodiments, the donor nucleic acid may comprise non-coding sequences that are specific target sites for additional nuclease designs. Subsequently, additional nucleases may be expressed in cells such that the original donor molecule is cleaved and modified by insertion of another donor molecule of interest. In this way, reiterative integrations of donor molecules may be generated allowing for trait stacking at a particular locus of interest or at a safe harbor locus.

Methods for targeted transgene integration

[0158] The donor molecules disclosed herein are integrated into a genome of a cell via targeted, homology-independent methods. For such targeted integration, the genome is cleaved at a desired location (or locations) using a nuclease, for example, a fusion between a DNA-binding domain (*e.g.*, zinc finger binding domain or TAL effector domain is engineered to bind a site at or near the predetermined cleavage site) and nuclease domain (*e.g.*, cleavage domain or cleavage half-domain). In certain embodiments, two fusion proteins, each comprising a DNA-binding domain and a cleavage half-domain, are expressed in a cell, and bind to sites which are juxtaposed in such a way that a functional cleavage domain is reconstituted and DNA is cleaved in the vicinity of the target site(s). In one embodiment, cleavage occurs between the binding sites of the two DNA-binding domains. One or both of the DNA-binding domains can be engineered. *See, also,* U.S. Patent No. 7,888,121; U.S. Patent Publication 20050064474 and International Patent Publications WO05/084190, WO05/014791 and WO 03/080809.

[0159] The nucleases as described herein can be introduced as polypeptides and/or polynucleotides. For example, two polynucleotides, each comprising sequences encoding one of the aforementioned polypeptides, can be introduced into a cell, and when the polypeptides are expressed and each binds to its target sequence, cleavage occurs at or near the target sequence. Alternatively, a single polynucleotide comprising sequences encoding both fusion polypeptides is introduced into a cell. Polynucleotides can be DNA, RNA or any modified forms or analogues or DNA and/or RNA.

[0160] Following the introduction of a double-stranded break in the region of interest, the transgene is integrated into the region of interest in a targeted manner via non-homology dependent methods (*e.g.*, non-homologous end joining (NHEJ)) following linearization of a double-stranded donor molecule as described herein. The double-stranded donor is preferably linearized *in vivo* with a nuclease, for example one or more of the same or different nucleases that are used to introduce the double-stranded break in the genome. Synchronized cleavage of the chromosome and the donor in the cell may limit donor DNA degradation (as compared to linearization of the donor molecule prior to introduction into the cell). The nuclease target site(s) used for linearization of the donor preferably do not disrupt the transgene(s) sequence(s).

[0161] The transgene may be integrated into the genome in the direction expected by simple ligation of the nuclease overhangs (designated “forward” or “AB” orientation) or in the alternate direction (designated “reverse” or “BA” orientation). In certain embodiments, the transgene is integrated following accurate ligation of the 5 donor and chromosome overhangs. In other embodiments, integration of the transgene in either the BA or AB orientation results in deletion of several nucleotides.

Delivery

[0162] The nucleases, polynucleotides encoding these nucleases, donor 10 polynucleotides and compositions comprising the proteins and/or polynucleotides described herein may be delivered *in vivo* or *ex vivo* by any suitable means into any cell type.

[0163] Suitable cells include eukaryotic (*e.g.*, animal or plant) and prokaryotic 15 cells and/or cell lines. Non-limiting examples of such cells or cell lines generated from such cells include COS, CHO (*e.g.*, CHO-S, CHO-K1, CHO-DG44, CHO-DUXB11, CHO-DUKX, CHOK1SV), VERO, MDCK, WI38, V79, B14AF28-G3, BHK, HaK, NS0, SP2/0-Ag14, HeLa, HEK293 (*e.g.*, HEK293-F, HEK293-H, HEK293-T), and perC6 cells as well as insect cells such as *Spodoptera frugiperda* (Sf), or fungal cells such as *Saccharomyces*, *Pichia* and *Schizosaccharomyces* as well as 20 plant cells from monocotyledonous or dicotyledonous plants including but not limited to maize, soybean, cotton, *Arabidopsis*, wheat, , barley, oats, sugar cane, sorghum, forage grasses, alfalfa, tomato, tobacco potato, rice, sunflower and *Brassica*. In certain embodiments, the cell line is a CHO, MDCK or HEK293 cell line. Suitable cells also include stem cells such as, by way of example, embryonic stem cells, induced 25 pluripotent stem cells, hematopoietic stem cells, neuronal stem cells and mesenchymal stem cells. In certain embodiments, the plant cells are but not limited to suspension culture, protoplasts, or organized tissues such as embryos, immature-embryos, leaf discs, cotyledons, hypotcols, and microspores. Methods of delivering nucleases as *described* herein are described, for example, in U.S. Patent Nos. 30 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824, the disclosures of all of which are incorporated by reference herein in their entireties.

[0164] Nucleases and/or donor constructs as described herein may also be delivered using vectors containing sequences encoding one or more of the zinc finger protein(s). Any vector systems may be used including, but not limited to, plasmid vectors, retroviral vectors, lentiviral vectors, adenovirus vectors, poxvirus vectors; 5 herpesvirus vectors and adeno-associated virus vectors, etc. *See, also*, U.S. Patent Nos. 6,534,261; 6,607,882; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824, incorporated by reference herein in their entireties. Furthermore, it will be apparent that any of these vectors may comprise one or more of the sequences needed for treatment. Thus, when one or more nucleases and a donor construct are introduced 10 into the cell, the nucleases and/or donor polynucleotide may be carried on the same vector or on different vectors. When multiple vectors are used, each vector may comprise a sequence encoding one or multiple nucleases and/or donor constructs.

[0165] Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids encoding nucleases and donor constructs in cells (*e.g.*, 15 mammalian cells) and target tissues. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome or poloxamer. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of *in vivo* delivery of engineered DNA-binding proteins and fusion 20 proteins comprising these binding proteins, see, *e.g.*, Rebar (2004) *Expert Opinion Invest. Drugs* 13(7):829-839; Rossi *et al.* (2007) *Nature Biotech.* 25(12):1444-1454 as well as general gene delivery references such as Anderson, *Science* 256:808-813 (1992); Nabel & Felgner, *TIBTECH* 11:211-217 (1993); Mitani & Caskey, *TIBTECH* 11:162-166 (1993); Dillon, *TIBTECH* 11:167-175 (1993); Miller, *Nature* 357:455-460 (1992); Van Brunt, *Biotechnology* 6(10):1149-1154 (1988); Vigne, *Restorative 25 Neurology and Neuroscience* 8:35-36 (1995); Kremer & Perricaudet, *British Medical Bulletin* 51(1):31-44 (1995); Haddada *et al.*, in *Current Topics in Microbiology and Immunology* Doerfler and Böhm (eds.) (1995); and Yu *et al.*, *Gene Therapy* 1:13-26 (1994).

30 **[0166]** Methods of non-viral delivery of nucleic acids include electroporation, lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-

enhanced uptake of DNA. Sonoporation using, *e.g.*, the Sonitron 2000 system (Rich-Mar) can also be used for delivery of nucleic acids.

[0167] Additional exemplary nucleic acid delivery systems include those provided by Amaxa Biosystems (Cologne, Germany), Maxcyte, Inc. (Rockville, Maryland), BTX Molecular Delivery Systems (Holliston, MA) and Copernicus Therapeutics Inc, (*see* for example US6008336). Lipofection is described in *e.g.*, U.S. Patent Nos. 5,049,386; 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (*e.g.*, Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424, WO 91/16024.

[0168] The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (*see*, *e.g.*, Crystal, *Science* 270:404-410 (1995); Blaese *et al.*, *Cancer Gene Ther.* 2:291-297 (1995); Behr *et al.*, *Bioconjugate Chem.* 5:382-389 (1994); Remy *et al.*, *Bioconjugate Chem.* 5:647-654 (1994); Gao *et al.*, *Gene Therapy* 2:710-722 (1995); Ahmad *et al.*, *Cancer Res.* 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

[0169] Additional methods of delivery include the use of packaging the nucleic acids to be delivered into EnGeneIC delivery vehicles (EDVs). These EDVs are specifically delivered to target tissues using bispecific antibodies where one arm of the antibody has specificity for the target tissue and the other has specificity for the EDV. The antibody brings the EDVs to the target cell surface and then the EDV is brought into the cell by endocytosis. Once in the cell, the contents are released (*see* MacDiarmid *et al* (2009) *Nature Biotechnology* 27(7):643).

[0170] The use of RNA or DNA viral based systems for the delivery of nucleic acids encoding engineered ZFPs take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (*in vivo*) or they can be used to treat cells *in vitro* and the modified cells are administered to patients (*ex vivo*). Conventional viral based systems for the delivery of ZFPs include, but are not limited to, retroviral, lentivirus, adenoviral, adeno-associated, vaccinia and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often

resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[0171] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system depends on the target tissue. Retroviral vectors are comprised of *cis*-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum *cis*-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immunodeficiency virus (SIV), human immunodeficiency virus (HIV), and combinations thereof (see, e.g., Buchscher *et al.*, *J. Virol.* 66:2731-2739 (1992); Johann *et al.*, *J. Virol.* 66:1635-1640 (1992); Sommerfelt *et al.*, *Virol.* 176:58-59 (1990); Wilson *et al.*, *J. Virol.* 63:2374-2378 (1989); Miller *et al.*, *J. Virol.* 65:2220-2224 (1991); PCT/US94/05700).

[0172] In applications in which transient expression is preferred, adenoviral based systems can be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and high levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, e.g., in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures (see, e.g., West *et al.*, *Virology* 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin *et al.*, *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, *et al.*, *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *PNAS* 81:6466-6470 (1984); and Samulski *et al.*, *J. Virol.* 63:03822-3828 (1989).

[0173] At least six viral vector approaches are currently available for gene transfer in clinical trials, which utilize approaches that involve complementation of defective vectors by genes inserted into helper cell lines to generate the transducing agent.

5 **[0174]** pLASN and MFG-S are examples of retroviral vectors that have been used in clinical trials (Dunbar *et al.*, *Blood* 85:3048-305 (1995); Kohn *et al.*, *Nat. Med.* 1:1017-102 (1995); Malech *et al.*, *PNAS* 94:22 12133-12138 (1997)). PA317/pLASN was the first therapeutic vector used in a gene therapy trial. (Blaese *et al.*, *Science* 270:475-480 (1995)). Transduction efficiencies of 50% or greater have 10 been observed for MFG-S packaged vectors. (Ellem *et al.*, *Immunol Immunother.* 44(1):10-20 (1997); Dranoff *et al.*, *Hum. Gene Ther.* 1:111-2 (1997)).

15 **[0175]** Recombinant adeno-associated virus vectors (rAAV) are a promising alternative gene delivery systems based on the defective and nonpathogenic parvovirus adeno-associated type 2 virus. All vectors are derived from a plasmid that retains only the AAV 145 bp inverted terminal repeats flanking the transgene expression cassette. Efficient gene transfer and stable transgene delivery due to integration into the genomes of the transduced cell are key features for this vector system. (Wagner *et al.*, *Lancet* 351:9117 1702-3 (1998), Kearns *et al.*, *Gene Ther.* 9:748-55 (1996)). Other AAV serotypes, including AAV1, AAV2, AAV3, AAV4, 20 AAV5, AAV6, AAV7, AAV8, AAV9 and AAVrh.10 and any novel AAV serotype can also be used in accordance with the present invention.

25 **[0176]** Replication-deficient recombinant adenoviral vectors (Ad) can be produced at high titer and readily infect a number of different cell types. Most adenovirus vectors are engineered such that a transgene replaces the Ad E1a, E1b, and/or E3 genes; subsequently the replication defective vector is propagated in human 293 cells that supply deleted gene function in *trans*. Ad vectors can transduce multiple types of tissues *in vivo*, including nondividing, differentiated cells such as those found in liver, kidney and muscle. Conventional Ad vectors have a large carrying capacity. An example of the use of an Ad vector in a clinical trial involved 30 polynucleotide therapy for antitumor immunization with intramuscular injection (Sterman *et al.*, *Hum. Gene Ther.* 7:1083-9 (1998)). Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include Rosenecker *et al.*, *Infection* 24:1 5-10 (1996); Sterman *et al.*, *Hum. Gene Ther.* 9:7 1083-1089 (1998);

Welsh *et al.*, *Hum. Gene Ther.* 2:205-18 (1995); Alvarez *et al.*, *Hum. Gene Ther.* 5:597-613 (1997); Topf *et al.*, *Gene Ther.* 5:507-513 (1998); Sterman *et al.*, *Hum. Gene Ther.* 7:1083-1089 (1998).

[0177] Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by a producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host (if applicable), other viral sequences being replaced by an expression cassette encoding the protein to be expressed. The missing viral functions are supplied in *trans* by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess inverted terminal repeat (ITR) sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely *rep* and *cap*, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, *e.g.*, heat treatment to which adenovirus is more sensitive than AAV.

[0178] In many gene therapy applications, it is desirable that the gene therapy vector be delivered with a high degree of specificity to a particular tissue type. Accordingly, a viral vector can be modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the outer surface of the virus. The ligand is chosen to have affinity for a receptor known to be present on the cell type of interest. For example, Han *et al.*, *Proc. Natl. Acad. Sci. USA* 92:9747-9751 (1995), reported that Moloney murine leukemia virus can be modified to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other virus-target cell pairs, in which the target cell expresses a receptor and the virus expresses a fusion protein comprising a ligand for the cell-surface receptor. For example, filamentous phage can be engineered to display antibody fragments (*e.g.*, FAB or Fv) having specific binding affinity for virtually any

chosen cellular receptor. Although the above description applies primarily to viral vectors, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences which favor uptake by specific target cells.

5 [0179] Gene therapy vectors can be delivered *in vivo* by administration to an individual patient, typically by systemic administration (*e.g.*, intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application, as described below. Alternatively, vectors can be delivered to cells *ex vivo*, such as cells explanted from an individual patient (*e.g.*, lymphocytes, bone 10 marrow aspirates, tissue biopsy) or universal donor hematopoietic stem cells, followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

[0180] Vectors (*e.g.*, retroviruses, adenoviruses, liposomes, etc.) containing nucleases and/or donor constructs can also be administered directly to an organism for 15 transduction of cells *in vivo*. Alternatively, naked DNA can be administered. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells including, but not limited to, injection, infusion, topical application and electroporation. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, 20 although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

[0181] Vectors suitable for introduction of polynucleotides (*e.g.* nuclease-encoding and/or double-stranded donors) described herein include non-integrating 25 lentivirus vectors (IDLV). *See*, for example, Ory *et al.* (1996) *Proc. Natl. Acad. Sci. USA* **93**:11382-11388; Dull *et al.* (1998) *J. Virol.* **72**:8463-8471; Zuffery *et al.* (1998) *J. Virol.* **72**:9873-9880; Follenzi *et al.* (2000) *Nature Genetics* **25**:217-222; U.S. Patent Publication No 2009/054985.

[0182] Pharmaceutically acceptable carriers are determined in part by the 30 particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions available, as described below (*see, e.g.*, *Remington's Pharmaceutical Sciences*, 17th ed., 1989).

[0183] It will be apparent that the nuclease-encoding sequences and donor constructs can be delivered using the same or different systems. For example, the nucleases and donors can be carried by the same vector. Alternatively, a donor polynucleotide can be carried by a plasmid, while the one or more nucleases can be carried by a AAV vector. Furthermore, the different vectors can be administered by the same or different routes (intramuscular injection, tail vein injection, other intravenous injection, intraperitoneal administration and/or intramuscular injection. The vectors can be delivered simultaneously or in any sequential order.

[0184] Thus, the instant disclosure includes *in vivo* or *ex vivo* treatment of diseases and conditions that are amenable to insertion of a transgenes encoding a therapeutic protein, for example treatment of hemophilias via nuclease-mediated integration of clotting factors such as Factor VIII (F8). The compositions are administered to a human patient in an amount effective to obtain the desired concentration of the therapeutic polypeptide in the serum or the target organ or cells. Administration can be by any means in which the polynucleotides are delivered to the desired target cells. For example, both *in vivo* and *ex vivo* methods are contemplated. Intravenous injection to the portal vein is a preferred method of administration. Other *in vivo* administration modes include, for example, direct injection into the lobes of the liver or the biliary duct and intravenous injection distal to the liver, including through the hepatic artery, direct injection in to the liver parenchyma, injection via the hepatic artery, and/or retrograde injection through the biliary tree. *Ex vivo* modes of administration include transduction *in vitro* of resected hepatocytes or other cells of the liver, followed by infusion of the transduced, resected hepatocytes back into the portal vasculature, liver parenchyma or biliary tree of the human patient, see *e.g.*, Grossman *et al.*, (1994) *Nature Genetics*, 6:335-341.

[0185] The effective amount of nuclease(s) and donor to be administered will vary from patient to patient and according to the therapeutic polypeptide of interest. Accordingly, effective amounts are best determined by the physician administering the compositions and appropriate dosages can be determined readily by one of ordinary skill in the art. After allowing sufficient time for integration and expression (typically 4-15 days, for example), analysis of the serum or other tissue levels of the therapeutic polypeptide and comparison to the initial level prior to administration will determine whether the amount being administered is too low, within the right range or

too high. Suitable regimes for initial and subsequent administrations are also variable, but are typified by an initial administration followed by subsequent administrations if necessary. Subsequent administrations may be administered at variable intervals, ranging from daily to annually to every several years. One of skill in the art will appreciate that appropriate immunosuppressive techniques may be recommended to avoid inhibition or blockage of transduction by immunosuppression of the delivery vectors, see *e.g.*, Vilquin *et al.*, (1995) *Human Gene Ther.*, 6:1391-1401.

5 [0186] Formulations for both *ex vivo* and *in vivo* administrations include suspensions in liquid or emulsified liquids. The active ingredients often are mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients include, for example, water, saline, dextrose, glycerol, ethanol or the like, and combinations thereof. In addition, the composition may contain minor amounts of auxiliary substances, such as, wetting or emulsifying agents, pH buffering agents, stabilizing agents or other reagents that enhance the 10 effectiveness of the pharmaceutical composition.

15 [0187] The delivery of nucleic acids may be introduced into a plant cell in embodiments of the invention by any method known to those of skill in the art, including, for example and without limitation: by transformation of protoplasts (*See, e.g.*, U.S. Patent 5,508,184); by desiccation/inhibition-mediated DNA uptake (*See, e.g.*, 20 Potrykus *et al.* (1985) *Mol. Gen. Genet.* 199:183-8); by electroporation (*See, e.g.*, U.S. Patent 5,384,253); by agitation with silicon carbide fibers (*See, e.g.*, U.S. Patents 5,302,523 and 5,464,765); by *Agrobacterium*-mediated transformation (*See, e.g.*, U.S. Patents 5,563,055, 5,591,616, 5,693,512, 5,824,877, 5,981,840, and 6,384,301);, by acceleration of DNA-coated particles (*See, e.g.*, U.S. Patents 5,015,580, 5,550,318, 25 5,538,880, 6,160,208, 6,399,861, and 6,403,865) and by Nanoparticles, nanocarriers and cell penetrating peptides (WO201126644A2; WO2009046384A1; WO2008148223A1) in the methods to deliver DNA, RNA, Peptides and/or proteins or combinations of nucleic acids and peptides into plant cells.

30 [0188] Through the application of techniques such as these, the cells of virtually any species may be stably transformed. In some embodiments, transforming DNA is integrated into the genome of the host cell. In the case of multicellular species, transgenic cells may be regenerated into a transgenic organism. Any of these techniques

may be used to produce a transgenic plant, for example, comprising one or more nucleic acid sequences of the invention in the genome of the transgenic plant.

[0189] The most widely-utilized method for introducing an expression vector into plants is based on the natural transformation system of *Agrobacterium*. *A.*

5 *tumefaciens* and *A. rhizogenes* are plant pathogenic soil bacteria that genetically transform plant cells. The T_i and R_i plasmids of *A. tumefaciens* and *A. rhizogenes*, respectively, carry genes responsible for genetic transformation of the plant. The T_i (tumor-inducing)-plasmids contain a large segment, known as T-DNA, which is transferred to transformed plants. Another segment of the T_i plasmid, the vir region, is 10 responsible for T-DNA transfer. The T-DNA region is bordered by left-hand and right-hand borders that are each composed of terminal repeated nucleotide sequences. In some modified binary vectors, the tumor-inducing genes have been deleted, and the functions of the vir region are utilized to transfer foreign DNA bordered by the T-DNA border sequences. The T-region may also contain, for example, a selectable marker for 15 efficient recovery of transgenic plants and cells, and a multiple cloning site for inserting sequences for transfer such as a nucleic acid encoding a fusion protein of the invention.

[0190] Thus, in some embodiments, a plant transformation vector is derived 20 from a T_i plasmid of *A. tumefaciens* (See, e.g., U.S. Patent Nos. 4,536,475, 4,693,977, 4,886,937, and 5,501,967; and European Patent EP 0 122 791) or a R_i plasmid of *A. rhizogenes*. Additional plant transformation vectors include, for example and without 25 limitation, those described by Herrera-Estrella *et al.* (1983) *Nature* 303:209-13; Bevan *et al.* (1983), *supra*; Klee *et al.* (1985) *Bio/Technol.* 3:637-42; and in European Patent EP 0 120 516, and those derived from any of the foregoing. Other bacteria, such as *Sinorhizobium*, *Rhizobium*, and *Mesorhizobium* that naturally interact with plants can be 30 modified to mediate gene transfer to a number of diverse plants. These plant-associated symbiotic bacteria can be made competent for gene transfer by acquisition of both a disarmed T_i plasmid and a suitable binary vector.

[0191] The following Examples relate to exemplary embodiments of the present disclosure in which the nuclease comprises a zinc finger nuclease (ZFN) or a 30 TALEN. It will be appreciated that this is for purposes of exemplification only and that other nucleases can be used, for instance a CRISPR/Cas nuclease system or homing endonucleases (meganucleases) with engineered DNA-binding domains

and/or fusions of naturally occurring or engineered homing endonucleases (meganucleases) DNA-binding domains and heterologous cleavage domains.

EXAMPLES

5 **Example 1: Materials and Methods**

Cell growth, transfection, and ZFN/TALEN assay.

[0192] Transfection of K562 (ATCC CCL-243) used Amaxa Solution V and program T-016; CHO-K1 (ATCC CCL-61), Amaxa Solution T and program U-023.

10 All transfections contained 10^6 cells and the following plasmids: *AAVS1*, 3 μ g of 2A-linked ZFNs and donor; *IL2R γ* and *CCR5*, 2 μ g of each unlinked ZFN plasmid and 10 μ g of the donor plasmid; *GS*, *FUT8*, 3 μ g of 2A-linked ZFNs and 10 μ g of donor plasmid. *See*, U.S. Patent No. 8,110,379 and U.S. Patent Publication Nos.

20100129869 and 20090042250 and DeKelver *et al.* (2010) *Genome Res.* 20(8):1133-1142; Liu *et al.* (2009) *Biotechnol. Bioeng.* 106(1):97-105; Malphettes *et al.* (2010)

15 *Biotechnology and Bioengineering* 106(5):774-83; Perez *et al.* (2008) *Nature Biotech.* 26(7):808-816; Urnov *et al.* (2005) *Nature* 435(7042):646-651 for further details including ZFN designs.

[0193] The *FUT8* TALE nuclease pair (SBS 101082 and SBS 101086) directly overlaps the ZFN binding site in exon 10 of *FUT8* and was constructed using the Δ 152/+63 N- and C-terminal truncation points (Miller *et al.* 2011). The binding site of SBS 101082 *FUT8* TALE is 5'-tgt atc tgg cca ctg at-3' (SEQ ID NO:1); SBS 101086, 5'- ttt gtc ttt gcc tcc tt-3' (SEQ ID NO:2).

Donor plasmid design and construction

25 [0194] Oligos containing ZFN target sites for the *AAVS1* (5'-tgt ccc ctc cAC CCC ACA GTG Ggg cca cTA GGG ACA GGA Ttg gtg aca ga-3', SEQ ID NO:3), spaced-flipped *AAVS1* (5'-tgt ccc ctc cAC CCC ACA GTG Ggt ggc cTA GGG ACA GGA Ttg gtg aca ga-3', SEQ ID NO:541), *GS* (5'-gac cCC AAG CCC ATT CCT GGG Aac tgg aAT GGT GCA GGC Tgc cat acc aa-3', SEQ ID NO:4), and *IL2R γ* (5'-gtt tcg tgt tCG GAG CCG CTT Taa ccc ACT CTG TGG AAG tgc tca gca tt-3', SEQ ID NO:5) ZFN pairs were annealed to their reverse complements in 50 mM NaCl, 10 mM Tris pH 7.5, and 1 mM EDTA. *See, also*, U.S. Patent Publication Nos.

20100129869 and U.S. Patent Nos. 7,951,925 and 8,110,379. Capital letters denote the ZFN binding sites while lowercase letters denote flanking and spacer sequence.

[0195] The double-stranded products were then cloned into the EcoRV site of the pBluescript II KS- vector (Agilent). The *CCR5* donor plasmid resulted from

5 insertion of the *CCR5* target site oligonucleotides (5'-GTC ATC CTC ATC CTG ATA AAC TGC AAA AGa-3', SEQ ID NO:6); 5-CTT TTG CAG TTT ATC AGG ATG AGG ATG ACa-3' SEQ ID NO:7, *see, also*, U.S. Patent No. 7,951,925) into pCR2.1 (Invitrogen). The second *AAVS1* donor plasmid (*see, Figure 1E*) was made by insertion of the above *AAVS1* target site oligos into the EcoRV site of a pCR2.1-based

10 plasmid also containing the GFP open reading frame driven by the pGK promoter.

[0196] The *FUT8* donor plasmid was made via insertion of the ZFN/TALEN binding site (5'- ggc CGT GTA TCT GGC CAC TGA TGA CCC TTC TTt gtt aAA GGA GGC AAA GAC AAA Gta a -3', SEQ ID NO:8) into a donor plasmid containing IgG and puromycin resistance transgenes (Moehle *et al.* (2007) *Proc. Nat'l Acad. Sci. USA* 104(9):3055-3060).

Assay of targeted integration

[0197] All PCR reactions were performed with 100 ng genomic DNA as a template, using Accuprime HiFiTM polymerase (Invitrogen). Genomic DNA was purified with the MasterpureTM kit (Epicentre).

[0198] Targeted integration of the *AAVS1* GFP donor plasmid at *AAVS1* (also known as *PPP1R12C*) was assayed at all four possible chromosome-donor junctions via PCR amplification. PCR reactions used a 60° annealing temperature, a 30 second extension time, 30 cycles of amplification, and the following primers: AB left, *AAVS1* CEL-I F (5'-ccc ctt acc tct cta gtc tgt gc-3', SEQ ID NO:9) and *AAVS1* Junction R (5'-ggc gat taa gtt ggg taa cg-3', SEQ ID NO:10); AB right, *AAVS1* Junction F (5'-ggc ctc ttg gtc aag ttg tt-3', SEQ ID NO:11) and *AAVS1* CEL-I R (5'-ctc agg ttc tgg gag agg gta g-3', SEQ ID NO:12); BA left, *AAVS1* CEL-I F and *AAVS1* Junction F; BA right, *AAVS1* Junction R and *AAVS1* CEL-I R. For the sequences in Figure 8A, the following primers were used: AB left, M13F (5'-gta aaa cga cgg cca gt-3', SEQ ID NO:13) and *AAVS1* CEL-F; BA right, M13F and *AAVS1* CEL-I R.

[0199] Targeted integration at *IL2R γ* , *CCR5*, and GS was assayed via PCR amplification. PCR reactions used a 58° annealing temperature, a 30 second extension

time, 5% DMSO, 26 cycles of amplification, and the M13F primer in combination with the following primers: *IL2R γ* BA left, *IL2R γ* CEL-I F (5'-acc agt gag ttt tca tta gg-3', SEQ ID NO:14); *IL2R γ* AB right, *IL2R γ* CEL-I R (5'-tgg agc aaa aga cag tgg tg-3', SEQ ID NO:15); *CCR5* BA left, R5F (5'-aag atg gat tat caa gtg tca agt cc-3', SEQ ID NO:16); *CCR5* AB right, R5R (5'-caa agt ccc act ggg cg-3', SEQ ID NO:17); *GS* BA left, GJC 172F (5'-atc cgc atg gga gat cat ct-3', SEQ ID NO:18); *GS* AB right, GJC 173R (5'-gtg tat gtt cgt tca ccc ac-3', SEQ ID NO:19).

5 [0200] Targeted integration of the *AAVS1* donor at *GS* (Figure 2B) was assayed via PCR amplification. PCR reactions used a 58°C annealing temperature, a 30 second extension time, 5% DMSO, 26 cycles of amplification, and the following primers: AB left, Jcn1F (5'-caa ata gga ccc tgt gaa gga-3', SEQ ID NO:20) and Jcn1R (5'-gat taa gtt ggg taa cgc cag-3', SEQ ID NO:21); BA left, Jcn3F (5'-aat agg acc ctg tga agg a-3', SEQ ID NO:22) and Jcn3R (5'-gtg tgg aat tgt gag cgg ata-3', SEQ ID NO:23).

10 15 [0201] Targeted integration of the IgG donor at *FUT8* (Figure 4) was assayed via PCR amplification. PCR reaction used a 60° annealing temperature, a 30 second extension time, 30 cycles of amplification (35 cycles for screening of crude lysates), and the following primers: AB left junction, GJC 75F (5'- agt cca tgt cag acg cac tg-3', SEQ ID NO:24) and SC seqpfzR (5'- aga gtg agg ctc tgt ctc aa-3', SEQ ID NO:25); AB right junction, *FUT8* donor CELIF2 (5'- tac gta tag gct gcg caa ct-3', SEQ ID NO:26) and GJC 115R (5'- gca cat gta gtc ttt gat ttt g-3', SEQ ID NO:27); BA left junction, GJC75F and *FUT8* donor CELIF2; BA right junction, SC seqpfzR and GJC115R.

20 25 [0202] The Southern blot of *AAVS1* GFP donor integration at *AAVS1* was probed as previously described (DeKelver et al. 2010, *ibid*). Expected results from this Southern blot are as follows: an *AAVS1* probe will hybridize to either a 2092 or 6592 bp band for the AB and BA orientations, respectively. The wild-type, triploid *AAVS1* locus will be seen as a 3287 bp band. The Southern blot of *AAVS1* GFP donor integration elsewhere in the genome was probed with the complete open reading frame of GFP. Integration at *AAVS1* will produce either 3323 or 4482 bp bands for the AB and BA orientations, respectively; non-targeted integrations elsewhere in the genome will produce secondary bands of indeterminable size. The Southern blot of *GS* donor integration at *GS* was probed with a 424 bp fragment of the *GS* gene

bounded by 5'-ctg cag gtg aag aca gga tg-3' and 5'-ccc act aga aag aac atg tt-3'. Integration at *GS* will be revealed as a hybridizing band at 2933 bp; the wild-type *GS* locus will produce a 1977 bp band. The Southern blot of *GS* donor integration elsewhere in the genome was probed with a *Bsa*I - *Scal*I fragment of the *E. coli bla* gene. Correctly integrated transgenes will give a 2055 bp band; integrations into the *GS* pseudogenes will give bands of 4878, 4214, 10080, and 9416 bp depending on the pseudogene and insert orientation; other non-targeted integrations will produce a single band of unpredictable size. The Southern blot of integration at *FUT8* was probed with a 407 bp *Hind*III-*Xmn*I fragment of the *FUT8* locus. For both *FUT8* Southerns, the genomic DNA was cut with *Hind*III.

10 [0203] Contigs containing *FUT8*, *GS*, and *GS* pseudogenes were extracted from the whole-CHO genome sequencing data using a custom Python script (Xu *et al.* (2011) *Nat Biotechnol* 29(8):735-41). *FUT8* is present on contig AFTD01065932.1; *GS* on contig AFTD01107178.1. One *GS* pseudogene (contained in AFTD01043599.1) has perfect conservation of the ZFN binding sites, 120/128 (94%) bp of homology to the exon 5 portion of the probe, and is expected to be present in a 6333 bp *Scal*I fragment. The second *GS* pseudogene (contained in AFTD01154859.1) has one mismatch in the ZFN binding sites, 116/128 (91%) bp of homology to the exon 5 portion of the probe and is expected to be present in a 13320 bp *Scal*I fragment.

15 [0204] Antibody concentrations were measured using the Pierce Easy-Titer IgG Assay Kit (23310) according to the manufacturer's instructions. Clones with at least two-fold higher than background were classified as positive.

20 **Example 2: Targeted integration following *in vivo* cleavage of a double-stranded donor**

A. *AAVS1*

25 [0205] To test whether transgene cleavage *in vivo* using the same ZFN that cuts the genomic target site would synchronize donor and chromosome cleavage, minimizing the vulnerability of the transgene to degradation, K562 cells were transfected with *AAVS1*-targeted ZFNs and a donor plasmid that includes the *AAVS1* ZFN target sites for cleavage of the donor plasmid *in vivo*. Briefly, as described in Example 1, we cloned the recognition site for the well-characterized and highly active *AAVS1* ZFNs into a donor plasmid containing an autonomous GFP expression

cassette but lacking homology to the *AAVS1* locus. *See*, U.S. Patent No. 8,110,379 and DeKelver *et al.* (2010) *Genome Res.* 20(8):1133-1142. The donor plasmid (with or without the ZFN target site) was co-transfected into K562 cells along with a second plasmid encoding the *AAVS1* ZFNs.

5 [0206] Insertion into the chromosomal *AAVS1* site assayed by PCR amplification of the unique junctions formed by targeted donor integration from genomic DNA isolated 3 days post-transfection as described above in Example 1.

[0207] As shown in Figure 1, when co-transfected with the cognate ZFNs, simultaneous cleavage of both a ZFN site-containing donor plasmid and the 10 chromosome will occur, allowing insertion of the plasmid into the chromosome. Insertion of the donor plasmid in the direction expected by simple ligation of the ZFN overhangs was designated as the AB orientation, the alternate direction was designated as the BA orientation. As shown in Figure 1D, the BA (reverse) orientation is favored when the nucleotides between the target sites (spacer) is the 15 reverse complement of the genomic (wild-type) sequence.

[0208] Furthermore, as shown in Figures 1D and 1E (see lanes 6, 8, 10, 17, 19 and 21 of 1E), consistent with successful capture of the cleaved donor DNA, we detected the expected 5' and 3' junctions formed by donor integration in both the AB and the BA orientations. The BA orientation was favored with the reverse 20 complement spacers (Figure 1D). Donor integration required ZFN-mediated cleavage as both (i) donor without an *AAVS1* ZFN site was not integrated despite efficient cleavage of the *AAVS1* locus (*see*, Figure 1E, lanes 5, 7, 9, and lanes 16, 18, and 20; Figure 5A) and (ii) transfection of a donor without co-transfection of the corresponding ZFN also failed to yield targeted integration.

25 [0209] Cell clones were obtained by limiting dilution from the pool transfected with both ZFN and donor (lane 8/19). Three GFP-positive and junction PCR-positive clones were analyzed in duplicate by Southern blot to confirm integration of the donor plasmid. The clones fall into three classes: clone one contains one AB insertion; clone two contains one BA insertion; clone three contains both AB 30 and BA insertions in addition to a non-inserted allele (Figure 1F).

[0210] The three clones were also analyzed for off-target integration by Southern blotting with a GFP-specific probe. Clone one contains only the expected

insertion at *AAVS1* whereas clones 2 and 3 contain a transgene insertion elsewhere in the genome in addition to the *AAVS1* insertions (Figure 1F).

[0211] PCR amplicons of the chromosome-donor integration junctions from these three cell lines were cloned and sequenced. As shown in Figure 7, clones 1 and 5 3 contained AB insertions with perfect ligation of the donor and chromosomal overhangs at both the 5' and 3' junctions. Clones 2 and 3 contained BA insertions with alleles produced by microhomology-driven repair at the left, 5' junction.

B. *IL2R γ , GS and CCR*

10 [0212] To demonstrate that capture of a cleaved donor was not restricted to integration into *AAVS1* in K562 cells, we performed analogous experiments at three other loci (*IL2R γ* , *CCR* and *GS*) in K562 and CHO cells. Successful targeted integration was monitored at one chromosome-donor junction for each orientation (AB and BA) as described above.

15 [0213] Site-specific integration targeted to the site of ZFN cleavage was observed for the *IL2R γ* , and *CCR5* loci in K562 cells and for the *GS* locus in CHO-K1 cells. (Figure 2A, lanes 3, 7, 11, 15, 19, and 23). As with integration into *AAVS1*, integration at *IL2R γ* , *CCR5*, and *GS* was dependent upon inclusion of the ZFN cleavage site in the donor plasmid (Figure 2A, lanes 1, 5, 9, 13, 17, and 21) and the co-delivery of the ZFNs themselves. ZFN activity, both at the chromosomal target and on the donor plasmid was essentially uniform across all samples (Figure 5B).

20 [0214] Sequencing of chromosome-donor junction PCR products from these loci, as well as from an analogous pool of *AAVS1* integrants, revealed a spectrum of insertion events consistent with correct integration at the targeted locus (Figure 8).

25 [0215] Thus, the ability to capture an *in vivo* cleaved transgene donor at a DSB is a general property of the mammalian DNA repair machinery and is independent of the specific target site or cell type.

Example 3: *In vivo* and *in vitro* cleavage

30 [0216] To confirm that *in vivo* cleavage was necessary to support the observed levels of targeted gene insertion, we performed a direct comparison of targeted integration using *in vivo* cleaved donors and donors cleaved *in vitro* using EcoRV, as described in Example 1.

[0217] As shown in Figure 2, while integration of pre-cleaved donor plasmids was occasionally detectable, it was markedly less efficient compared to the *in vivo*-cleaved donors (Figure 2A, compare lanes 2/3, 6/7, 10/11, 14/15, 18/19, and 22/23). Moreover, the use of pre-cleaved donor DNAs showed an increased range of junction 5 PCR sizes consistent with an increased level of donor DNA degradation prior to chromosomal capture (see, e.g., Figure 2A, lane 22).

[0218] To confirm targeted integration could be stimulated via the use of two different nucleases (ZFNs), we used the *GS* ZFN pair (Example 1) to cut the chromosome of CHO-K1 cells and the *AAVS1* ZFN pair to cleave a donor plasmid in 10 the same cell. Integration at *GS* was detected at a similar frequency both when the *GS* ZFN pair cut the chromosome and the donor (as in Figure 2A) and when the *GS* ZFNs cut the chromosome while the *AAVS1* ZFNs cut the donor (Figure 2B, lanes 6 and 8, lanes 15 and 17). Cleavage efficiency at *GS* was again uniform over all *GS* ZFN-transfected samples (Figure 5C).

15 **[0219]** Thus, *in vivo* cleavage is more efficient than pre-cleavage of the donor molecule.

Example 4: Homology-independent targeted integration into CHO cells

[0220] Targeted integration in CHO cells has particularly important 20 applications in biotechnology yet CHO cells perform HDR-based targeted integration of several kilobase transgenes very poorly. To highlight this point, we compared HDR-mediated targeted integration in both HEK-293 cells and CHO-K1 cells using a system designed to deliver a promoterless GFP gene into a promoter-containing acceptor locus, essentially as described in Moehle *et al.* (2007) *Proc. Nat'l Acad. Sci. USA* 104(9):3055-3060). Targeted integration results in expression of GFP and allows quantitation by flow cytometry.

25 **[0221]** When transfected with ZFNs and a homology-containing donor plasmid (for integration via HDR), between 0.5 and 3% of HEK-293 cells became GFP-positive (Figure 6). In contrast, none of the CHO-K1 cells became GFP-positive 30 when similarly transfected. Given that CHO cells perform HDR-based targeted integration poorly, and yet have proven their utility for recombinant protein production, we next asked whether *in vivo* cleavage of donor DNA could be exploited to drive targeted integration in CHO cells.

[0222] CHO-K1 cells from the pool bearing targeted integration at GS (Figure 2A, lanes 19/23) were cloned by limiting dilution and single-cell derived clones screened by PCR for site-specific integration. In contrast to the negative results obtained with the HDR-based approach, homology-independent capture of the *in vivo* cleaved donor DNA yielded 10% (17/157) single-cell derived clones that were PCR-positive for the left chromosome-donor BA junction, 8% (13) positive for the right BA junction, and 6% (10) positive for both BA junctions. Eight of these ten clones were chosen randomly for analysis by Southern blotting. All 8 clones contained the expected targeted transgene insertion at the GS target site, a wild-type GS allele, and two GS pseudogenes (Figure 3). Only the wild-type GS allele and the pseudogenes are present in wild-type CHO-K1 cells (Figure 3, lane 9). Furthermore, when probed with a transgene-specific sequence, five of the 8 clones were shown to contain only one copy of the transgene at GS, whereas 3 contained a transgene copy at GS along with one or more randomly integrated copies at other sites in the CHO genome, one of which corresponded to integration into a GS pseudogene (Figure 3, lanes 10, 12, 13, 16, 17 and lanes 11, 14, 15, respectively). The chromosome-transgene junctions were sequenced and are shown in Figure 9.

Example 5: Targeted integration into and disruption of FUT8

[0223] Transgenes are routinely inserted into the CHO cell genome to produce biopharmaceutical proteins, notably antibodies. CHO cells with a deletion of the *FUT8* gene yield fucosylated antibodies with 100-fold higher antibody-dependent cellular cytotoxicity (Malphettes *et al.* (2010) *Biotechnology and Bioengineering* 106(5):774-83; Yamane-Ohnuki *et al.* (2004) *Biotech. Bioeng.* 87(5):614-22).

Moreover, knockout of *FUT8* expression can be selected for, thus potentially coupling targeted integration with this selectable trait. We therefore used the previously described *FUT8*-specific ZFNs to disrupt the *FUT8* gene via insertion of an *in vivo*-cleaved antibody production cassette (Moehle *et al.* (2007) *Proc. Nat'l Acad. Sci. USA* 104(9):3055-3060). Furthermore, we wished to determine whether capture of an *in vivo* cleaved donor could occur at a double-strand break produced by a TALE nuclease (TALEN) specific for *FUT8*.

[0224] Briefly, ZFNs or TALENs that cleave *FUT8* were cotransfected with an antibody expression plasmid containing a *FUT8* nuclease cleavage site as

described in Example 1. The transfected pool was selected for biallelic *FUT8* knockout using *Lens culinaris* agglutinin and cells cloned by limiting dilution (see, Malphettes *et al.* (2010) *Biotechnology and Bioengineering* 106(5):774-83). Clones were screened for secretion of IgG and for insertion of the IgG transgene by PCR of

5 both left and right transgene/chromosome junctions.

[0225] As shown in Figure 4, ZFN-treated clones, 25/96 (26%) of clones expressed IgG and 14/96 (15%) of clones were positive for insertion of the complete IgG transgene. All but one clone positive for insertion by PCR expressed IgG. Similar results were obtained with the *FUT8* TALENs: 35/171 (20%) of clones expressed IgG and all 16 (9%) of clones with complete transgene insertion expressed IgG. Clones with one (but not both) transgene integration junctions detectable by PCR accounted for a significant fraction of the remaining IgG-expressing clones (Figure 4A).

[0226] These experiments were also performed using TALE-nucleases targeted to *FUT8*. Transgene integration at *FUT8* was confirmed by Southern blot analysis for ten PCR- and IgG-positive clones derived from TALEN-mediated transgene insertion (Figures 4B and 4C). In addition, Figure 12 shows sequences obtained by sequencing of PCR junctions of *FUT8*-integrated donors using ZFNs (Figure 10A) or TALENs (Figure 10B).

[0227] In sum, the methods and compositions described herein provide for the facile and targeted integration of large transgenes via homology-independent methods, including in cell lines (*e.g.*, CHO cells) that are resistant to homology-driven integration.

25 **Example 6: Targeted integration into and disruption of wheat AHAS loci**
Characterization and identification of AHAS genomic target sequences

[0228] The transcribed regions for three homoeologous AHAS genes were identified and determined, zinc finger nucleases were designed to bind and cleave the sites for NHEJ-mediated targeting of a donor sequence as described in U.S.

30 Provisional Patent Filing No. 61/809,097, incorporated herein by reference. These novel sequences are listed as SEQ ID NO:116, SEQ ID NO:117, and SEQ ID NO:118. Previous sequencing efforts identified and genetically mapped homoeologous copies of AHAS genes from *Triticum aestivum* to the long arms of

chromosomes 6A, 6B and 6D (Anderson *et al.*, (2004) Weed Science 52:83-90; and, Li *et al.*, (2008) Molecular Breeding 22:217-225). Sequence analysis of Expressed Sequence Tags (EST) and genomic sequences available in Genbank (Accession Numbers: AY210405.1, AY210407.1, AY210406.1, AY210408.1, FJ997628.1, FJ997629.1, FJ997631.1, FJ997630.1, FJ997627.1, and AY273827.1) were used to determine the transcribed region for the homoeologous copies of the AHAS gene (SEQ ID NOs: 116-118).

[0229] The novel, non-coding AHAS gene sequences located upstream and downstream of the transcribed region were characterized for the first time. To completely characterize these non-coding sequences, the transcribed sequences for each of the three homoeologous copies of the AHAS gene were used as BLASTN™ queries to screen unassembled ROCHE 454™ sequence reads that had been generated from whole genome shotgun sequencing of *Triticum aestivum* cv. Chinese Spring. The ROCHE 454™ sequence reads of *Triticum aestivum* cv. Chinese Spring had been generated to 5-fold sequence coverage. Sequence assembly was completed using the SEQUENCHER SOFTWARE™ (GeneCodes, Ann Arbor, MI) of the ROCHE 454™. Sequence reads with a significant BLASTN™ hit (E-value <0.0001) were used to characterize these non-transcribed region. Iterative rounds of BLASTN™ analysis and sequence assembly were performed. Each iteration incorporated the assembled AHAS sequence from the previous iteration so that all of the sequences were compiled as a single contiguous sequence. Overall, 4,384, 7,590 and 6,205 of genomic sequences for the homoeologous AHAS genes located on chromosomes 6A, 6B and 6D, respectively, were characterized (SEQ ID NOs:119-121).

25 **Sequence analysis of AHAS genes isolated from *Triticum aestivum* cv. Bobwhite MPB26RH**

[0230] The homoeologous copies of the AHAS gene were cloned and sequenced from *Triticum aestivum* cv. Bobwhite MPB26RH to obtain nucleotide sequence suitable for designing specific zinc finger proteins that could bind the sequences with a high degree of specificity. The sequence analysis of the AHAS nucleotide sequences obtained from *Triticum aestivum* cv. Bobwhite MPB26RH was required to confirm the annotation of nucleotides present in Genbank and the ROCHE 454™ AHAS gene sequences and due to allelic variation between cv.

Bobwhite MPB26RH and the other wheat varieties from which the Genbank and ROCHE 454™ sequences were obtained.

[0231] A cohort of PCR primers were designed for amplification of the AHAS genes (Table 1). The primers were designed from a consensus sequence which was produced from multiple sequence alignments generated using CLUSTALW™ (Thompson *et al.*, (1994) Nucleic Acids Research 22:4673-80). The sequence alignments were assembled from the cv. Chinese Spring sequencing data generated from ROCHE 454™ sequencing which was completed at a 5-fold coverage.

[0232] As indicated in Table 1, the PCR primers were designed to amplify all three homoeologous sequences or to amplify only a single homoeologous sequence. For example, the PCR primers used to amplify the transcribed region of the AHAS gene were designed to simultaneously amplify all three homoeologous copies in a single multiplex PCR reaction. The PCR primers used to amplify the non-transcribed region were either designed to amplify all three homoeologous copies or to amplify only a single homoeologous copy. All of the PCR primers were designed to be between 18 and 27 nucleotides in length and to have a melting temperature of 60 to 65°C, optimal 63°C. In addition, several primers were designed to position the penultimate base (which contained a phosphorothioate linkage and is indicated in Table 1 as an asterisk [*]) over a nucleotide sequence variation that distinguished the gene copies from each wheat sub-genome. Table 1 lists the PCR primers that were designed and synthesized.

Table 1: Primer sequences used for PCR amplification of AHAS sequences

Primer Name	Region	Genome Amplified	SEQ ID NO.	Sequence (5' → 3')
AHAS-p_Fwd5	5' UTR	D	122	TCTGTAAGTTATCGCCT GAATTGCTT
AHAS-p_Rvs6	5' UTR	D	123	CATTGTGACATCAGCA TGACACAA
AHAS-p_Fwd4	5' UTR	D	124	AAGCAYGGCTTGCCTA CAGC
AHAS-p_Rvs3	5' UTR	D	125	AACCAAATRCCCCAT GTCTCTCC
AHAS-p_Fwd1	5' UTR	A, B, and D	126	CGTCGCCCGTAGACC ATTC
AHAS-p_Rvs1	5' UTR	A, B, and D	127	GGAGGGGTGATGKTTT TGTCTTT
AHAS_1F1_transcribed	Coding	A, B, and D	128	TCG CCC AAA CCC TCG CC

AHAS_1R1_transcribed	Coding	A, B, and D	129	GGG TCG TCR CTG GGG AAG TT
AHAS_2F2_transcribed	Coding	A, B, and D	130	GCC TTC TTC CTY GCR TCC TCT GG
AHAS_2R2_transcribed	Coding	A, B, and D	131	GCC CGR TTG GCC TTG TAA AAC CT
AHAS_3F1_transcribed	Coding	A, B, and D	132	AYC AGA TGT GGG CGG CTC AGT AT
AHAS_3R1_transcribed	Coding	A, B, and D	133	GGG ATA TGT AGG ACA AGA AAC TTG CAT GA
AHAS-6A.PS.3'.F1	3'UTR	A	134	AGGGCCATACTTGTG GATATCAT*C
AHAS-6A.PS.3'.R2	3'UTR	A	135	GCCAACACCCTACACT GCCTA*T
AHAS-6B.PS.3'.F1	3'UTR	B	136	TGCGCAATCAGCATGA TACC*T
AHAS-6B.PS.3'.R1	3'UTR	B	137	ACGTATCCGCAGTCGA GCAA*T
AHAS-6D.PS.3'.F1	3'UTR	D	138	GTAGGGATGTGCTGTC ATAAGAT*G
AHAS-6D.PS.3'.R3	3'UTR	D	139	TTGGAGGCTCAGCCGA TCA*C

UTR = untranslated region

Coding = primers designed for the transcribed regions

asterisk (*) indicates the incorporation of a phosphorothioate sequence

5 [0233] Sub-genome-specific amplification was achieved using on-off PCR (Yang *et al.*, (2005) Biochemical and Biophysical Research Communications 328:265-72) with primers that were designed to position the penultimate base (which contained a phosphorothioate linkage) over a nucleotide sequence variation that distinguished the gene copies from each wheat sub-genome. Two different sets of

10 PCR conditions were used to amplify the homoeologous copies of the AHAS gene from cv. Bobwhite MPB26RH. For the transcribed regions, the PCR reaction contained 0.2 mM dNTPs, 1X IMMOLASE PCR™ buffer (Bioline, Taunton, MA), 1.5 mM MgCl₂, 0.25 units IMMOLASE DNA POLYMERASE™ (Bioline, Taunton, MA), 0.2 µM each of forward and reverse primer, and about 50 ng genomic DNA.

15 Reactions containing the AHAS_1F1 and AHAS_1R1 primers were supplemented with 8% (v/v) DMSO. For the non-transcribed regions, the PCR reactions contained 0.2 mM dNTP, 1X PHUSION GC BUFFER™ (New England Biolabs Ipswich, MA), 0.5 units HOT-START PHUSION DNA™ polymerase (New England Biolabs), 0.2 µM each of forward and reverse primer, and about 50 ng genomic DNA. PCR was

performed in a final 25 μ l reaction volume using an MJ PTC200® thermocycler (BioRad, Hercules, CA). Following PCR cycling, the reaction products were purified and cloned using PGEM-T EASY VECTOR™ (Promega, Madison, WI) into *E. coli* JM109 cells. Plasmid DNA was extracted using a DNAEASY PLASMID DNA PURIFICATION KIT™ (Qiagen, Valencia, CA) and Sanger sequenced using BIGDYE® v3.1 chemistry (Applied Biosystems, Carlsbad, CA) on an ABI3730XL® automated capillary electrophoresis platform. Sequence analysis performed using SEQUENCHER SOFTWARE™ (GeneCodes, Ann Arbor, MI) was used to generate a consensus sequence for each homoeologous gene copy (SEQ ID NO:140, SEQ ID 5 NO:141, and SEQ ID NO:142) from cv. Bobwhite MPB26RH. CLUSTALW™ was used to produce a multiple consensus sequence alignment from which homoeologous 10 sequence variation distinguishing between the AHAS gene copies was confirmed.

Design of Zinc Finger Binding domains specific to AHAS gene sequences

15 [0234] Zinc finger proteins directed against the identified DNA sequences of the homoeologous copies of the AHAS genes were designed as previously described. See, e.g., Urnov *et al.*, (2005) Nature 435:646-551. Exemplary target sequence and recognition helices are shown in Table 2 (recognition helix regions designs) and Table 20 3 (target sites). In Table 3, nucleotides in the target site that are contacted by the ZFP recognition helices are indicated in uppercase letters; non-contacted nucleotides are indicated in lowercase. Zinc Finger Nuclease (ZFN) target sites were designed for 4 regions in the AHAS gene: a region about 500-bp upstream of the serine 653 amino acid residue, an upstream region adjacent (within 30-bp) to the serine 653 amino acid residue, a downstream region adjacent (within 80-bp) to the serine 653 amino acid 25 residue, and a region about 400-bp downstream of the serine 653 amino acid residue. Numerous ZFP designs were developed and tested to identify the fingers which bound with the highest level of efficiency with 22 different AHAS target sites which were identified in wheat as described in U.S. Provisional Patent Filing No. 61809097, incorporated herein by reference. The specific ZFP recognition helices (Table 2) 30 which bound with the highest level of efficiency to the zinc finger recognition sequences were used for NHEJ-mediated targeting and integration of a donor sequence (homology-independent targeted integration) within the AHAS locus of the wheat genome.

Table 2: AHAS zinc finger designs (N/A indicates “not applicable”)

ZFP#	F1	F2	F3	F4	F5	F6
29730	TSGNLTR SEQ ID NO:143	HRTSLTD SEQ ID NO:144	QSSDLSR SEQ ID NO:145	HKYHLRS SEQ ID NO:146	QSSDLSR SEQ ID NO:145	QWSTRKR SEQ ID NO:147
29732	RSDSLSA_ SEQ ID NO:148	RSDALAR_ SEQ ID NO:149	RSDDLTR_ SEQ ID NO:150	QKSNLSS_ SEQ ID NO:151	DSSDRKK SEQ ID NO:152	N/A
30012	HSNARKT SEQ ID NO:153	QSGNLAR SEQ ID NO:154	DRSALAR SEQ ID NO:155	RSDNLST SEQ ID NO:156	AQWGRTS SEQ ID NO:157	N/A
30018	QSGDLTR SEQ ID NO:158	MRNRLNR SEQ ID NO:159	DRSNLSR SEQ ID NO:160	WRSCRSA SEQ ID NO:161	QRSNLDS SEQ ID NO:162	N/A

Table 3: Target site of AHAS zinc fingers

ZFP	AHAS Region	Target Site (5'→3')	SEQ ID NO:
29730	Within 30-bp upstream of S653	agGCAGCACGTGCTCCTGATgcgggact	163
29732	Within 30-bp upstream of S653	gaTCCCAAGCGGTGGTGcttcaaggac	164
30012	Within 80-bp downstream of S653N	tcTTGTAGGTCGAAATTcagtacgagg	165
30018	Within 80-bp downstream of S653N	taCAAgTGTGACaTGCGCAatcagcatg	166

5

[0235] The AHAS zinc finger designs were incorporated into zinc finger expression vectors encoding a protein having at least one finger with a CCHC structure. *See*, U.S. Patent Publication No. 2008/0182332. In particular, the last finger in each protein had a CCHC backbone for the recognition helix. The non-canonical zinc finger-encoding sequences were fused to the nuclease domain of the type IIS restriction enzyme *FokI* (amino acids 384-579 of the sequence of Wah *et al.*, (1998) Proc. Natl. Acad. Sci. USA 95:10564-10569) via a four amino acid ZC linker and an *opaque-2* nuclear localization signal derived from *Zea mays* to form AHAS zinc-finger nucleases (ZFNs). *See*, U.S. Patent No. 7,888,121.

10 [0236] The optimal zinc fingers were verified for cleavage activity using a budding yeast based system previously shown to identify active nucleases. *See*, e.g., U.S. Patent Publication No. 2009/0111119; Doyon *et al.*, (2008) Nat Biotechnology

26:702-708; Geurts *et al.*, (2009) Science 325:433. Zinc fingers for the various functional domains were selected for *in vivo* use. Of the numerous ZFNs that were designed, produced and tested to bind to the putative AHAS genomic polynucleotide target site. The ZFNs described in Table 2 above, were identified as having *in vivo* activity at high levels, and were characterized as being capable of efficiently binding and cleaving the unique AHAS genomic polynucleotide target sites *in planta*.

Evaluation of zinc finger nuclease cleavage of AHAS genes using transient assays

[0237] ZFN construct assembly: Plasmid vectors containing ZFN gene

10 expression constructs, which were identified using the yeast assay as previously described, were designed and completed using skills and techniques commonly known in the art. (see, for example, Ausubel or Maniatis). Each ZFN-encoding sequence was fused to a sequence encoding an *opaque-2* nuclear localization signal (Maddaloni *et al.*, (1989) Nuc. Acids Res. 17:7532), that was positioned upstream of 15 the zinc finger nuclease.

[0238] Expression of the fusion proteins was driven by the strong constitutive promoter from the Zea mays Ubiquitin gene, (which includes the 5' untranslated region (UTR) (Toki *et al.*, (1992) Plant Physiology 100:1503-07). The expression cassette also included the 3' UTR (comprising the transcriptional terminator and 20 polyadenylation site) from the Zea mays peroxidase 5 gene (Per5) gene(US Patent Publication No. 2004/0158887). The self-hydrolyzing 2A encoding the nucleotide sequence from *Thosea asigna* virus (Szymczak *et al.*, (2004) Nat Biotechnol. 22:760-760) was added between the two Zinc Finger Nuclease fusion proteins that were cloned into the construct.

25 **[0239]** The plasmid vectors were assembled using the IN-FUSION™

Advantage Technology (Clontech, Mountain View, CA). Restriction endonucleases were obtained from New England BioLabs (Ipswich, MA) and T4 DNA Ligase (Invitrogen, Carlsbad, CA) was used for DNA ligation. Plasmid preparations were performed using NUCLEOSPIN® Plasmid Kit (Macherey-Nagel Inc., Bethlehem, 30 PA) or the Plasmid Midi Kit (Qiagen) following the instructions of the suppliers.

DNA fragments were isolated using QIAQUICK GEL EXTRACTION KIT™ (Qiagen) after agarose tris-acetate gel electrophoresis. Colonies of all ligation reactions were initially screened by restriction digestion of miniprep DNA. Plasmid

DNA of selected clones was sequenced by a commercial sequencing vendor (Eurofins MWG Operon, Huntsville, AL). Sequence data were assembled and analyzed using the SEQUENCHER™ software (Gene Codes Corp., Ann Arbor, MI).

[0240] Representative plasmids pDAB109350 and pDAB109360 are shown in

5 Figure 11 and Figure 12 and were confirmed via restriction enzyme digestion and via DNA sequencing.

Preparation of ZFN constructs DNA for transfection

[0241] Before delivery to *Triticum aestivum* protoplasts, plasmid DNA for each ZFN construct was prepared from cultures of *E. coli* using the PURE YIELD PLASMID MAXIPREP SYSTEM® (Promega Corporation, Madison, WI) or PLASMID MAXI KIT® (Qiagen, Valencia, CA) following the instructions of the suppliers.

15 Isolation of wheat mesophyll protoplasts

[0242] Mesophyll protoplasts from the wheat line cv. Bobwhite MPB26RH were prepared for transfection using polyethylene glycol (PEG)-mediated DNA delivery as follows.

[0243] Mature seed was surface sterilized by immersing in 80% (v/v) ethanol for 30 secs, rinsing twice with tap water, followed by washing in 20% DOMESTOS® (0.8% v/v available chlorine) on a gyratory shaker at 140 rpm for 20 mins. The DOMESTOS® was removed by decanting and the seeds were rinsed four times with sterile water. Excess water was removed by placing the seed on WHATMAN™ filter paper. The seeds were placed in a sterile PETRI™ dish on several sheets of dampened sterile WHATMAN™ filter paper and incubated for 24 h at 24°C. Following incubation, the seeds were surface sterilized a second time in 15% DOMESTOS® with 15 min shaking, followed by rinsing with sterile water as described previously. The seeds were placed on Murashige and Skoog (MS) solidified media for 24 hr at 24°C. Finally, the seeds were surface sterilized a third time in 10% DOMESTOS® with 10 min shaking, followed by rinsing in sterile water as previously described. The seeds were placed, crease side down, onto MS solidified media with 10 seeds per PETRI™ dish and germinated in the dark at 24°C for 14-21 days.

[0244] About 2-3 grams of leaf material from the germinated seeds was cut into 2-3 cm lengths and placed in a pre-weighed PETRITTM dish. Leaf sheath and yellowing leaf material was discarded. Approximately 10 mL of leaf enzyme digest mix (0.6 M mannitol, 10 mM MES, 1.5% w/v cellulase R10, 0.3% w/v macerozyme, 5 mM CaCl₂, 0.1% bovine serum albumin, 0.025% v/v pluronic acid, 5 mM β -mercaptoethanol, pH 5.7) was pipetted into the PETRITTM dish and the leaf material was chopped transversely into 1-2 mm segments using a sharp scalpel blade. The leaf material was chopped in the presence of the leaf digest mix to prevent cell damage resulting from the leaf material drying out. Additional leaf enzyme digest mix was 10 added to the PETRITTM dish to a volume of 10 mL per gram fresh weight of leaf material and subject to vacuum (20" Hg) pressure for 30 min. The PETRITTM dish was sealed with PARAFILM[®] and incubated at 28°C with gentle rotational shaking for 4-5 hours.

[0245] Mesophyll protoplasts released from the leaf segments into the enzyme digest mix were isolated from the plant debris by passing the digestion suspension through a 100 micron mesh and into a 50 mL collection tube. To maximize the yield of protoplasts, the digested leaf material was washed three times. Each wash was performed by adding 10 mL wash buffer (20 mM KCl, 4 mM MES, 0.6 M mannitol, pH 5.6) to the PETRITTM dish, swirling gently for 1 min, followed by passing of the 20 wash buffer through the 100 micron sieve into the same 50 mL collection tube. Next, the filtered protoplast suspension was passed through a 70 micron sieve, followed by a 40 micron sieve. Next, 6 mL aliquots of the filtered protoplast suspension were transferred to 12 mL round bottomed centrifugation tubes with lids and centrifuged at 70 g and 12°C for 10 min. Following centrifugation, the supernatant was removed and 25 the protoplast pellets were each resuspended in 7 mL wash buffer. The protoplasts were pelleted a second time by centrifugation, as described above. The protoplasts were each resuspended in 1 mL wash buffer and pooled to two centrifugation tubes. The wash buffer volume was adjusted to a final volume of 7 mL in each tube before centrifugation was performed, as described above. Following removal of the 30 supernatant, the protoplast pellets were resuspended in 1 mL wash buffer and pooled to a single tube. The yield of mesophyll protoplasts was estimated using a Neubauer haemocytometer. Evans Blue stain was used to determine the proportion of live cells recovered.

PEG-mediated transfection of mesophyll protoplasts

[0246] About 10^6 mesophyll protoplasts were added to a 12 mL round bottomed tube and pelleted by centrifugation at 70 g before removing the supernatant. The protoplasts were gently resuspended in 600 μ l wash buffer containing 70 μ g of plasmid DNA. The plasmid DNA consisted of the Zinc Finger Nuclease constructs described above. Next, an equal volume of 40% PEG solution (40% w/v PEG 4,000, 0.8 M mannitol, 1M Ca(NO₃)₂, pH 5.6) was slowly added to the protoplast suspension with simultaneous mixing by gentle rotation of the tube. The protoplast suspension was allowed to incubate for 15 min at room temperature without any agitation.

[0247] An additional 6 mL volume of wash buffer was slowly added to the protoplast suspension in sequential aliquots of 1 mL, 2mL and 3 mL. Simultaneous gentle mixing was used to maintain a homogenous suspension with each sequential aliquot. Half of the protoplast suspension was transferred to a second 12 mL round bottomed tube and an additional 3 mL volume of wash buffer was slowly added to each tube with simultaneous gentle mixing. The protoplasts were pelleted by centrifugation at 70 g for 10 min and the supernatant was removed. The protoplast pellets were each resuspended in 1 mL wash buffer before protoplasts from the paired round bottomed tubes were pooled to a single 12 mL tube. An additional 7 mL wash buffer was added to the pooled protoplasts before centrifugation as described above. The supernatant was completely removed and the protoplast pellet was resuspended in 2 mL Qiao's media (0.44% w/v MS plus vitamins, 3 mM MES, 0.0001% w/v 2,4-D, 0.6 M glucose, pH 5.7). The protoplast suspension was transferred to a sterile 3 cm PETRITTM dish and incubated in the dark for 24°C for 72 h.

25 **Genomic DNA isolation from mesophyll protoplasts**

[0248] Transfected protoplasts were transferred from the 3 cm PETRITTM dish to a 2 mL microfuge tube. The cells were pelleted by centrifugation at 70 g and the supernatant was removed. To maximize the recovery of transfected protoplasts, the PETRITTM dish was rinsed three times with 1 mL of wash buffer. Each rinse was performed by swirling the wash buffer in the PETRITTM dish for 1 min, followed by transfer of the liquid to the same 2 ml microfuge tube. At the end of each rinse, the cells were pelleted by centrifugation at 70 g and the supernatant was removed. The pelleted protoplasts were snap frozen in liquid nitrogen before freeze drying for 24 h

in a LABCONCO FREEZONE 4.5® (Labconco, Kansas City, MO) at -40°C and 133 x 10⁻³ mBar pressure. The lyophilized cells were subjected to DNA extraction using the DNEASY® PLANT DNA EXTRACTION MINI kit (Qiagen) following the manufacturer's instructions, with the exception that tissue disruption was not required 5 and the protoplast cells were added directly to the lysis buffer.

PCR assay of protoplast genomic DNA for ZFN sequence cleavage

[0249] To enable the cleavage efficacy and target site specificity of ZFNs designed for the AHAS gene locus to be investigated, PCR primers were designed to 10 amplify up to a 300-bp fragment within which one or more ZFN target sites were captured. One of the primers was designed to be within a 100-bp window of the captured ZFN target site(s). This design strategy enabled Illumina short read

technology to be used to assess the integrity of the target ZFN site in the transfected protoplasts. In addition, the PCR primers were designed to amplify the three

15 homoeologous copies of the AHAS gene and to capture nucleotide sequence variation that differentiated between the homoeologs such that the Illumina sequence reads could be unequivocally attributed to the wheat sub-genome from which they were derived.

[0250] A total of four sets of PCR primers were designed to amplify the ZFN 20 target site loci (Table 4). Each primer set was synthesized with the Illumina SP1 and SP2 sequences at the 5' end of the forward and reverse primer, respectively, to provide compatibility with Illumina short read sequencing chemistry. The synthesized primers also contained a phosphorothioate linkage at the penultimate 5' and 3'

nucleotides (indicated in Table 4 as an asterisk [*]). The 5' phosphorothioate linkage 25 afforded protection against exonuclease degradation of the Illumina SP1 and SP2 sequences, while the 3' phosphorothioate linkage improved PCR specificity for amplification of the target AHAS sequences using on-off PCR (Yang *et al.*, (2005)). All PCR primers were designed to be between 18 and 27 nucleotides in length and to have a melting temperature of 60 to 65°C, optimal 63°C.

30 **[0251]** In Table 4, nucleotides specific for the AHAS gene are indicated in uppercase type; nucleotides corresponding to the Illumina SP1 and SP2 sequences are indicated in lowercase type. Each primer set was empirically tested for amplification

of the three homoeologous AHAS gene copies through Sanger-based sequencing of the PCR amplification products.

Table 4: Primer sequences used to assess AHAS ZFN cleavage efficacy and target site specificity

Primer Name	AHAS Region	Primer Sequence (5'→3')	SEQ ID NO:
AHAS-500ZFN.F3	500-bp upstream of S653	a*cactttccctacacgacgtctccgatct CCTCTAGGATTCAAGACTTT G*G	167
AHAS-500ZFN.R1	500-bp upstream of S653	g*tgactggagttcagacgtgtgcttccgatct CGTGGCCGCTTGTAAGTGTAA	168
AHASs653Z FN.F1	Within 30-bp upstream of S653	a*cactttccctacacgacgtctccgatctG AGACCCCAGGGCCATACTT*G	169
AHASs653Z FN.R3	Within 30-bp upstream of S653	g*tgactggagttcagacgtgtgcttccgatct CAAGCAAACATAGAAAACGCA TG*G	170
AHASs653Z FN.F5	Within 80-bp downstream of S653N	a*cactttccctacacgacgtctccgatctA TGGAGGGTGATGGCAGGA*C	171
AHASs653Z FN.R1	Within 80-bp downstream of S653N	g*tgactggagttcagacgtgtgcttccgatct ATGACAGCACATCCCTACAAA AG*A	172
AHAS+400Z FN.F1	400-bp downstream of S653	a*cactttccctacacgacgtctccgatctA ACAGTGTGCTGGTTCCCTTCT*G	173
AHAS+400Z FN.R3	400-bp downstream of S653	g*tgactggagttcagacgtgtgcttccgatct TYTYYCCTCCCAACTGTATTCA AG*A	174

5 asterisk (*) is used to indicate a phosphorothioate

[0252] PCR amplification of ZFN target site loci from the genomic DNA extracted from transfected wheat mesophyll protoplasts was used to generate the requisite loci specific DNA molecules in the correct format for Illumina-based sequencing-by-synthesis technology. Each PCR assay was optimized to work on 200 ng starting DNA (about 12,500 cell equivalents of the *Triticum aestivum* genome). Multiple reactions were performed per transfected sample to ensure sufficient copies of the *Triticum aestivum* genome were assayed for reliable assessment of ZFN efficiency and target site specificity. About sixteen PCR assays, equivalent to 200,000 copies of the *Triticum aestivum* genome taken from individual protoplasts, were performed per transfected sample. A single PCR master-mix was prepared for each transfected sample. To ensure optimal PCR amplification of the ZFN target site (i.e.

to prevent PCR reagents from becoming limiting and to ensure that PCR remained in the exponential amplification stage) an initial assay was performed using a quantitative PCR method to determine the optimal number of cycles to perform on the target tissue. The initial PCR was performed with the necessary negative control 5 reactions on a MX3000P THERMOCYCLER™ (Stratagene). From the data output gathered from the quantitative PCR instrument, the relative increase in fluorescence was plotted from cycle-to-cycle and the cycle number was determined per assay that would deliver sufficient amplification, while not allowing the reaction to become reagent limited, in an attempt to reduce over-cycling and biased amplification of 10 common molecules. The unused master mix remained on ice until the quantitative PCR analysis was concluded and the optimal cycle number determined. The remaining master mix was then aliquoted into the desired number of reaction tubes (about 16 per ZFN assay) and PCR amplification was performed for the optimal cycle number. Following amplification, samples for the same ZFN target site were pooled 15 together and 200 µl of pooled product per ZFN was purified using a QIAQUICK MINIELUTE PCR PURIFICATION KIT™ (Qiagen) following the manufacturer's instructions.

[0253] To enable the sample to be sequenced using Illumina short read 20 technology, an additional round of PCR was performed to introduce the Illumina P5 and P7 sequences onto the amplified DNA fragments, as well as a sequence barcode index that could be used to unequivocally attribute sequence reads to the sample from which they originated. This was achieved using primers that were in part complementary to the SP1 and SP2 sequences added in the first round of amplification, but also contained the sample index and P5 and P7 sequences. The 25 optimal number of PCR cycles required to add the additional sequences to the template without over-amplifying common fragments was determined by quantitative PCR cycle analysis, as described above. Following amplification, the generated product was purified using AMPURE MAGNETIC BEADS® (Beckman-Coulter) with a DNA-to-bead ratio of 1:1.7. The purified DNA fragment were titrated for 30 sequencing by Illumina short read technology using a PCR-based library quantification kit (KAPA) according the manufacturer's instructions. The samples were prepared for sequencing using a cBot cluster generation kit (Illumina) and were sequenced on an ILLUMINA GAII_X™ or HISEQ2000™ instrument (Illumina) to

generate 100-bp paired end sequence reads, according to the manufacturer's instructions.

Data analysis for detecting NHEJ at target ZFN sites

5 [0254] Following generation of Illumina short read sequence data for sample libraries prepared for transfected mesophyll protoplasts, bioinformatics analysis was performed to identify deleted nucleotides at the target ZFN sites. Such deletions are known to be indicators of *in planta* ZFN activity that result from non-homologous end joining (NHEJ) DNA repair.

10 [0255] To identify sequence reads with NHEJ deletions, the manufacturer's supplied scripts for processing sequence data generated on the HISEQ2000™ instrument (Illumina) was used to first computationally assign the short sequence reads to the protoplast sample from which they originated. Sample assignment was based on the barcode index sequence that was introduced during library preparation, 15 as described previously. Correct sample assignment was assured as the 6-bp barcode indexes used to prepare the libraries were differentiated from each other by at least a two-step sequence difference.

[0256] Following sample assignment, a quality filter was passed across all sequences. The quality filter was implemented in custom developed PERL script.

20 Sequence reads were excluded if there were more than three ambiguous bases, or if the median Phred score was less than 20, or if there were three or more consecutive bases with a Phred score less than 20, or if the sequence read was shorter than 40 nucleotides in length.

[0257] Next, the quality trimmed sequences were attributed to the wheat sub- 25 genome from which they originated. This was achieved using a second custom developed PERL script in which sub-genome assignment was determined from the haplotype of the nucleotide sequence variants that were captured by the PCR primers used to amplify the three homoeologous copies of the AHAS gene, as described above.

30 [0258] Finally, the frequency of NHEJ deletions at the ZFN cleavage site in the sub-genome-assigned sequence reads was determined for each sample using a third custom developed PERL script and manual data manipulation in Microsoft

Excel 2010 (Microsoft Corporation). This was achieved by counting the frequency of unique NHEJ deletions on each sub-genome within each sample.

[0259] Two approaches were used to assess the cleavage efficiency and specificity of the ZFNs tested. Cleavage efficiency was expressed (in parts per million reads) as the proportion of sub-genome assigned sequences that contained a NHEJ deletion at the ZFN target site. Rank ordering of the ZFNs by their observed cleavage efficiency was used to identify ZFNs with the best cleavage activity for each of the four target regions of the AHAS genes in a sub-genome-specific manner.

[0260] All of the ZFNs tested showed NHEJ deletion size distributions consistent with that expected for *in planta* ZFN activity. Cleavage specificity was expressed as the ratio of cleavage efficiencies observed across the three sub-genomes. The inclusion of biological replicates in the data analyses did not substantially affect the rank order for cleavage activity and specificity of the ZFNs tested.

[0261] From these results, the ZFNs encoded on plasmid pDAB109350 (i.e. ZFN 29732 and 29730) and pDAB109360 (i.e. ZFN 30012 and 30018) were selected for *in planta* targeting in subsequent experiments, given their characteristics of significant genomic DNA cleavage activity in each of the three wheat sub-genomes.

Evaluation of donor designs for ZFN-mediated AHAS gene editing using transient assays

[0262] To investigate ZFN-mediated genomic editing at the endogenous AHAS gene locus in wheat, a series of experiments were undertaken to assess the effect of donor design on the efficiency of non-homologous end joining (NHEJ)-directed DNA repair. These experiments used transient assays to monitor the efficiency for ZFN-mediated addition of the previously described S653N mutation conferring tolerance to imidazolinone class herbicides (Li *et al.*, (2008) Molecular Breeding 22:217-225) at the endogenous AHAS gene locus in wheat, or alternatively for ZFN-mediated introduction of an *Eco*RI restriction endonuclease sequence site at the double strand DNA break created in the endogenous AHAS genes by targeted ZFN cleavage.

Donor designs for NHEJ-directed DNA repair

[0263] Two types of donor DNA designs were used for NHEJ-directed DNA repair.

[0264] The first type of donor design was a linear, double stranded DNA molecule comprising 41-bp of sequence that shared no homology with the endogenous AHAS genes in wheat. Two donor DNA molecules were designed, each to target the three homoeologous copies of the AHAS gene. Both donor DNA molecules had protruding 5' and 3' ends to provide ligation overhangs to facilitate ZFN-mediated NHEJ-directed DNA repair. The two donor DNA molecules differed by the sequence at their protruding 3' end. The first donor DNA molecule, pDAS000152 (SEQ ID NO:175 and SEQ ID NO:176), was designed to provide ligation overhangs that were compatible with those generated by cleavage of the endogenous AHAS genes by ZFNs 29732 and 29730 (encoded on plasmid pDAB109350) and to result in the insertion of the 41-bp donor molecule into the endogenous AHAS gene at the site of the double strand DNA break via NHEJ-directed DNA repair. The second donor DNA molecule pDAS000149 (SEQ ID NO:177 and SEQ ID NO:178) was designed to provide ligation overhangs that were compatible with those generated by the dual cleavage of the endogenous AHAS genes by ZFNs 29732 and 29730 (encoded on plasmid pDAB109350) and ZFNs 30012 and 30018 (encoded on plasmid pDAB109360) and to result in the replacement of the endogenous AHAS sequence contained between the two double strand DNA breaks created by the ZFNs with the 41-bp donor molecule via NHEJ-directed DNA repair.

[0265] The second type of donor was a plasmid DNA vector containing 41-bp of sequence that shared no homology with the endogenous AHAS genes in wheat and that was flanked on either side by sequence that was recognized by the ZFN(s) used to create double strand DNA breaks in the endogenous AHAS genes. This donor design allowed *in planta* release of the unique 41-bp sequence from the plasmid DNA molecule by the same ZFN(s) used to cleave target sites in the endogenous AHAS genes, and simultaneous generation of protruding ends that were suitable for overhang ligation of the released 41-bp sequence into the endogenous AHAS genes via NHEJ-directed DNA repair. Two plasmid donor DNA molecules were designed, each to target the three homoeologous copies of the AHAS gene. The first plasmid donor molecule, pDAS000153 (SED ID NO:179 and SEQ ID NO:180) (Figure 13), was

designed to provide ligation overhangs on the released 41-bp DNA fragment that were compatible with those generated by cleavage of the endogenous AHAS genes by ZFNs 29732 and 29730 (encoded on plasmid pDAB109350). The second plasmid donor molecule, pDAS000150 (SEQ ID NO:181 and SEQ ID NO:182) (Figure 14),
5 was designed to provide ligation overhangs on the released 41-bp DNA fragment that were at one end compatible with those generated by ZFNs 29732 and 29730 (encoded on plasmid pDAB109350) and at the other end compatible with those generated by ZFNs 30012 and 30018 (encoded on plasmid pDAB109360). This design allowed the replacement of the endogenous AHAS sequence contained between the two double
10 strand DNA breaks created by ZFNs 29732 and 29730 and ZFNs 30012 and 30018 with the 41-bp donor molecule sequence.

Synthesis of donor DNA for NHEJ-directed DNA repair

[0266] Standard cloning methods commonly known by one skilled in the art
15 were used to build the plasmid vectors. Before delivery to *Triticum aestivum*, plasmid DNA for each donor construct was prepared from cultures of *E. coli* using the PURE YIELD PLASMID MAXIPREP SYSTEM® (Promega Corporation, Madison, WI) or PLASMID MAXI KIT® (Qiagen, Valencia, CA) following the instructions of the suppliers.
20 [0267] Standard phosphoramidite chemistry was used to synthetically synthesize the double stranded DNA donor molecules (Integrated DNA Technologies, Coralville, IA). For each donor molecule, a pair of complementary single stranded DNA oligomers was synthesized, each with two phosphorothioate linkages at their 5' ends to provide protection against *in planta* endonuclease degradation. The single
25 stranded DNA oligomers were purified by high performance liquid chromatography to enrich for full-length molecules and purified of chemical carryover from the synthesis steps using Na⁺ exchange. The double stranded donor molecule was formed by annealing equimolar amounts of the two complementary single-stranded DNA oligomers using standard methods commonly known by one skilled in the art. Before
30 delivery to *Triticum aestivum*, the double stranded DNA molecules were diluted to the required concentration in sterile water.

Isolation of wheat protoplasts derived from somatic embryogenic callus

[0268] Protoplasts derived from somatic embryogenic callus (SEC) from the donor wheat line cv. Bobwhite MPB26RH were prepared for transfection using polyethylene glycol (PEG)-mediated DNA delivery as follows:

5 [0269] Seedlings of the donor wheat line were grown in an environment controlled growth room maintained at 18/16°C (day/night) and a 16/8 hour (day/night) photoperiod with lighting provided at 800 mmol m² per sec. Wheat spikes were collected at 12-14 days post-anthesis and were surface sterilized by soaking for 1 min in 70% (v/v) ethanol. The spikes were threshed and the immature seeds were
10 sterilized for 15 min in 17% (v/v) bleach with gentle shaking, followed by rinsing at least three times with sterile distilled water. The embryos were aseptically isolated from the immature seeds under a dissecting microscope. The embryonic axis was removed using a sharp scalpel and discarded. The scutella were placed into a 9 cm PETRITM dish containing 2-4 medium without TIMENTINTM, with the uncut
15 scutellum oriented upwards. A total of 25 scutella were plated onto each 9 cm PETRITM dish. Somatic embryogenic callus (SEC) formation was initiated by incubating in the dark at 24°C for 3 weeks. After 3 weeks, SEC was separated from non-embryogenic callus, placed onto fresh 2-4 medium without TIMENTINTM and incubated for a further 3 weeks in the dark at 24°C. Sub-culturing of SEC was
20 repeated for a total of three times before being used for protoplast preparation.

[0270] About one gram of SEC was chopped into 1-2 mm pieces using a sharp scalpel blade in a 10 cm PETRITM dish contained approximately 10 mL of wheat callus digest mix (2.5% w/v Cellulase RS, 0.2% w/v pectolyase Y23, 0.1% w/v DRISELASE®, 14 mM CaCl₂, 0.8 mM MgSO₄, 0.7 mM KH₂PO₄, 0.6 M Mannitol, pH 5.8) to prevent the callus from dehydrating. Additional callus digest mix was added to the PETRITM dish to a volume of 10 mL per gram fresh weight of callus and subject to vacuum (20" Hg) pressure for 30 min. The PETRITM dish was sealed with PARAFILM® and incubated at 28°C with gentle rotational shaking at 30-40 rpm for 4-5 hours.

30 [0271] SEC protoplasts released from the callus were isolated by passing the digestion suspension through a 100 micron mesh and into a 50 mL collection tube. To maximize the yield of protoplasts, the digested callus material was washed three times. Each wash was performed by adding 10 mL SEC wash buffer (0.6 M Mannitol,

0.44% w/v MS, pH 5.8) to the PETRITM dish, swirling gently for 1 min, followed by passing of the SEC wash buffer through the 100 micron sieve into the same 50 mL collection tube. Next, the filtered protoplast suspension was passed through a 70 micron sieve, followed by a 40 micron sieve. Next, 6 mL aliquots of the filtered 5 protoplast suspension were transferred to 12 mL round bottomed centrifugation tubes with lids and centrifuged in at 70 g and 12°C for 10 min. Following centrifugation, the supernatant was removed, leaving approximately 0.5 mL supernatant behind, and the protoplast pellets were each resuspended in 7mL of 22% sucrose solution. The sucrose/protoplast mixture was carefully overlaid with 2 mL SEC wash buffer, 10 ensuring that there was no mixing of the two solutions. The protoplasts were centrifuged a second time by centrifugation, as described above. The band of protoplasts visible between the SEC wash buffer and sucrose solution was collected using a pipette and placed into a clean 12 mL round bottom tube. Seven mL of SEC wash buffer was added to the protoplasts and the tubes were centrifuged, as described 15 above. The supernatant was removed and the SEC protoplasts were combined to a single tube and resuspended in a final volume 1-2 mL of SEC wash buffer. The yield of SEC protoplasts was estimated using a Neubauer haemocytometer. Evans Blue stain was used to determine the proportion of live cells recovered.

20 **PEG-mediated transfection of SEC protoplasts**

[0272] About two million SEC protoplasts were added to a 12 mL round bottomed tube and pelleted by centrifugation at 70 g before removing the supernatant. The protoplasts were gently resuspended in 480 µl SEC wash buffer containing 70 µg 25 of DNA. The DNA consisted of the Zinc Finger Nuclease and donor DNA constructs described above, with each construct present at the molar ratio required for the experiment being undertaken. Next, 720 µl of 50% PEG solution (50% w/v PEG 4000, 0.8 M mannitol, 1M Ca(NO₃)₂, pH 5.6) was slowly added to the protoplast suspension with simultaneous mixing by gentle rotation of the tube. The protoplast suspension was allowed to incubate for 15 min at room temperature without any 30 agitation.

[0273] An additional 7 mL volume of SEC wash buffer was slowly added to the protoplast suspension in sequential aliquots of 1 mL, 2 mL and 3 mL. Simultaneous gentle mixing was used to maintain a homogenous suspension with

each sequential aliquot. Half of the protoplast suspension was transferred to a second 12 mL round bottomed tube and an additional 3 mL volume of SEC wash buffer was slowly added to each tube with simultaneous gentle mixing. The protoplasts were pelleted by centrifugation at 70 g for 10 min and the supernatant was removed. The 5 protoplast pellets were each resuspended in 1 mL SEC wash buffer before protoplasts from the paired round bottomed tubes were pooled to a single 12 mL tube. An additional 7 mL SEC wash buffer was added to the pooled protoplasts before centrifugation as described above. The supernatant was completely removed and the protoplast pellet was resuspended in 2 mL Qiao's media. The protoplast suspension 10 was transferred to a sterile 3 cm PETRITM dish and incubated in the dark for 24°C for 72 h.

Isolation of scutella from immature zygotic wheat embryos

[0274] Scutella of immature zygotic wheat embryos from the donor wheat line cv. Bobwhite MPB26RH were prepared for transfection using biolistics-mediated 15 DNA delivery as follows.

[0275] Seedlings of the donor wheat line were grown in an environment controlled growth room maintained at 18/16°C (day/night) and a 16/8 hour (day/night) photoperiod with lighting provided at 800 mmol m² per sec. Wheat spikes were collected at 12-14 days post-anthesis and were surface sterilized by soaking for 20 1 min in 70% (v/v) ethanol. The spikes were threshed and the immature seeds were sterilized for 15 min in 17% (v/v) bleach with gentle shaking, followed by rinsing at least three times with sterile distilled water. The embryos were aseptically isolated from the immature seeds under a dissecting microscope. The embryonic axis was removed using a sharp scalpel and discarded. The scutella were placed into a 9 cm 25 PETRITM dish containing osmotic MS (E3 maltose) medium, with the uncut scutellum oriented upwards. A total of 20 scutella were plated onto each 9 cm PETRITM dish. The prepared embryos were pre-cultured in the dark at 26°C for a minimum of 4 h before transfection using biolistics-mediated DNA delivery.

30 **Transfection of scutella of immature zygotic wheat embryos by biolistic-mediated DNA delivery**

[0276] Gold particles for biolistic-mediated DNA delivery were prepared by adding 40 mg of 0.6 micron colloidal gold particles (BioRad) to 1 mL of sterile water

in a 1.5 mL microtube. The gold particles were resuspended by vortexing for 5 min. To prepare sufficient material for 10 bombardments, a 50 μ L aliquot of the gold particle suspension was transferred to a 1.5 mL microtube containing 5 μ g of DNA resuspended in 5 μ L of sterile water. Following thorough mixing by vortexing, 50 μ L 5 of 2.5 M CaCl₂ and 20 μ L of 0.1 M spermidine were added to the microtube, with thorough mixing after the addition of each reagent. The DNA-coated gold particles were pelleted by centrifugation for 1 min at maximum speed in a bench top microfuge. The supernatant was removed and 1 mL of 100% ethanol was added to wash and resuspend the gold particles. The gold particles were pelleted by 10 centrifugation, as described above, and the supernatant discarded. The DNA-coated gold particles were resuspended in 110 μ L of 100% ethanol and maintained on ice. Following a brief vortex, 10 μ L of the gold particle solution was placed centrally onto a macro-carrier membrane and allowed to air dry.

[0277] The PDS-1000/HE PARTICLE GUN DELIVERY SYSTEMTM 15 (BioRad) was used to transfect the scutella of immature zygotic wheat embryos by biolistic-mediated DNA delivery. Delivery of the DNA-coated gold particles was performed using the following settings: gap 2.5 cm, stopping plate aperture 0.8 cm, target distance 6.0 cm, vacuum 91.4 - 94.8 kPa, vacuum flow rate 5.0 and vent flow rate 4.5. The scutella of immature zygotic wheat embryos were bombarded using a 20 900 psi rupture disc. Each PETRITTM dish containing 20 scutella was bombarded once. The bombarded scutella were incubated at 26°C in the dark for 16 h before being transferred onto medium for callus induction. The scutella were cultured on callus induction medium in the dark at 26°C for 7 d.

25 **Genomic DNA isolation from SEC protoplasts**

[0278] Genomic DNA was extracted from SEC protoplasts using the 30 procedure previously described for mesophyll protoplasts. An additional purification step was performed to reduce the presence of the donor DNA used for transfection. This was achieved using gel electrophoresis to separate the genomic DNA from the SEC protoplasts from the donor DNA used for transfection. The extracted DNA was electrophoresed for 3 h in a 0.5% agarose gel using 0.5X TBE. The DNA was visualized by SYBR[®] SAFE staining and the band corresponding to genomic DNA from the SEC protoplasts was excised. The genomic DNA was purified from the

agarose gel using a QIAQUICK DNA PURIFICATION KIT™ (Qiagen), following the manufacturer's instructions, except that the QIAQUICK™ DNA purification column was replaced with a DNA binding column from the DNEASY PLANT DNA EXTRACTION MINI KIT™ (Qiagen).

5

Genomic DNA isolation from scutella of immature zygotic embryos

[0279] The 20 scutella of immature zygotic wheat embryos transfected for each biolistic-mediated DNA delivery were transferred to a 15 ml tube and snap frozen in liquid nitrogen before freeze drying for 24 h in a LABCONCO FREEZONE 4.5® (Labconco, Kansas City, MO) at -40°C and 133 x 10⁻³ mBar pressure. The lyophilized calli were subjected to DNA extraction using the DNEASY® PLANT DNA EXTRACTION MAXIT™ KIT (Qiagen) following the manufacturer's instructions.

[0280] An additional purification step was performed to reduce the presence of the donor DNA used for transfection. This was achieved using gel electrophoresis to separate the genomic DNA from the calli from the donor DNA used for transfection. The extracted DNA was electrophoresed for 3 h in a 0.5% agarose gel using 0.5X TBE. The DNA was visualized by SYBR® SAFE staining and the band corresponding to genomic DNA from the calli was excised. The genomic DNA was purified from the agarose gel using a QIAQUICK™ DNA PURIFICATION kit (Qiagen), following the manufacturer's instructions, except that the QIAQUICK™ DNA purification column was replaced with a DNA binding column from the DNEASY® PLANT DNA EXTRACTION MAXIT™ KIT (Qiagen).

25 **PCR assay of genomic DNA for ZFN-mediated AHAS editing**

[0281] To investigate ZFN-mediated genomic editing at the endogenous AHAS genes in wheat using NHEJ-directed DNA repair, and assess the effect of donor DNA design on the efficacy of each DNA repair pathway, PCR assays were used to amplify the target AHAS regions from genomic DNA of transfected wheat cells. PCR assays were performed as described previously to generate requisite loci specific DNA molecules in the correct format for Illumina-based sequencing-by-synthesis technology. Each assay was performed using the previously described primer pair (SEQ ID NO: 160 and SEQ ID NO: 170) that were designed to amplify

the region targeted by ZFNs 29732 and 29730 (encoded on plasmid pDAB109350) and ZFNs 30012 and 30018 (encoded on plasmid pDAB109360) for each of the three homoeologous copies of the AHAS genes. Multiple reactions were performed per transfected sample to ensure that sufficient copies of the *Triticum aestivum* genome were assayed for reliable assessment of ZFN-mediated gene editing. For transfected SEC protoplasts, up to sixteen PCR assays, equivalent to 200,000 copies of the *Triticum aestivum* genome taken from individual protoplasts, were performed per transfected sample. For transfected scutella of immature zygotic embryos, about forty eight (48) PCR assays, equivalent to 600,000 copies of the *Triticum aestivum* genome taken from individual protoplasts, were performed per transfected sample. Each transfected sample was prepared for sequencing using a CBOT CLUSTER GENERATION KIT™ (Illumina) and was sequenced on an ILLUMINA GAII_X™ or HISEQ2000™ instrument (Illumina) to generate 100-bp paired end sequence reads, as described previously.

Data analysis for detecting ZFN-mediated NHEJ-directed editing at AHAS genes

[0282] Following generation of Illumina short read sequence data for sample libraries prepared for transfected SEC protoplasts and scutella of immature zygotic wheat embryos, analyses were performed to identify molecular evidence for ZFN-mediated NHEJ-directed editing at the target ZFN sites.

[0283] To identify sequence reads with molecular evidence for NHEJ-directed gene editing, the short sequence reads were first computationally processed, as previously described, to assign each read to the sample and sub-genome from which they originated, and to perform quality filtering to ensure that only high quality sequences were used for subsequent analyses. Next, custom developed PERL scripts and manual data manipulation in Microsoft Excel 2010 (Microsoft Corporation) was used to identify reads that contained sequence for both the donor DNA molecule used for transfection and the endogenous AHAS locus. The editing frequency (expressed in parts per million reads) was calculated as the proportion of sub-genome-assigned sequence reads that showed evidence for ZFN-mediated NHEJ-directed gene editing.

[0284] From the results of three biological replicates performed for each linear double stranded DNA donor design, molecular evidence was obtained for the enrichment of sequence reads showing ZFN-mediated NHEJ-directed editing at the

three homoeologous copies of the endogenous AHAS genes in wheat (Table 7 and Table 8). Strong molecular evidence was obtained for the integration of the linear, double-stranded 41-bp donor molecule at the position of the double strand DNA break created by cleavage of the homoeologous copies of the AHAS gene by ZFNs 29732 and 29730 in samples of both SEC protoplasts and scutella of immature zygotic embryos that were transfected with pDAB109350 and pDAS000152. Similar editing efficiency was observed across the three wheat sub-genomes in these samples. In contrast, samples of SEC protoplasts and scutella of immature zygotic embryos transfected with pDAB109350 and pDAS000153 showed poor evidence for ZFN-mediated NHEJ-directed gene editing, presumably due to the prerequisite requirement for *in planta* release of the 41-bp donor sequence from the plasmid backbone.

10 Molecular evidence for the replacement of endogenous AHAS sequence with the 41-bp donor molecule was observed in both SEC protoplasts and scutella of immature zygotic embryos that were transfected with pDAB109350, pDAB109360 and pDAS000149. However, the frequency of editing was significantly lower than that observed for transfections performed using pDAB109350 and pDAS000152, presumably due to the requirement for dual ZFN cleavage of the endogenous AHAS sequence. Limited evidence was obtained for the replacement of endogenous AHAS sequence with the 41-bp donor molecule that required *in planta* release from plasmid backbone in samples of SEC protoplast and scutella of immature zygotic embryos that were transfected with pDAB109350, pDAB109360 and pDAS000150.

15

20

Table 7: Average NHEJ editing frequency in parts per million (ppm) across three biological replicates of scutella transfected with linear double-stranded donor DNA designs. “na” indicates “not applicable.”

Donor	ZFN	Donor-to-ZFN molar ratio	Editing Frequency in Wheat Sub-Genome	Editing Frequency (ppm)
pDAS000152	n/a	n/a	A	0
pDAS000152	29732-2A-29730	5:1	A	0
pDAS000152	29732-2A-29730	10:1	A	131
pDAS000152	n/a	n/a	B	0
pDAS000152	29732-2A-29730	5:1	B	0
pDAS000152	29732-2A-29730	10:1	B	47
pDAS000152	n/a	n/a	D	0

pDAS000152	29732-2A-29730	5:1	D	0
pDAS000152	29732-2A-29730	10:1	D	75
pDAS000153	n/a	n/a	A	0
pDAS000153	29732-2A-29730	5:1	A	4
pDAS000153	29732-2A-29730	10:1	A	0
pDAS000153	n/a	n/a	B	0
pDAS000153	29732-2A-29730	5:1	B	0
pDAS000153	29732-2A-29730	10:1	B	0
pDAS000153	n/a	n/a	D	0
pDAS000153	29732-2A-29730	5:1	D	0
pDAS000153	29732-2A-29730	10:1	D	0
pDAS000149	n/a	n/a	A	0
pDAS000149	29732-2A-29730	5:1	A	23
pDAS000149	29732-2A-29730	10:1	A	9
pDAS000149	n/a	n/a	B	0
pDAS000149	29732-2A-29730	5:1	B	7
pDAS000149	29732-2A-29730	10:1	B	3
pDAS000149	n/a	n/a	D	0
pDAS000149	29732-2A-29730	5:1	D	7
pDAS000149	29732-2A-29730	10:1	D	0
pDAS000150	n/a	n/a	A	0
pDAS000150	29732-2A-29730	5:1	A	1
pDAS000150	29732-2A-29730	10:1	A	0
pDAS000150	n/a	n/a	B	0
pDAS000150	29732-2A-29730	5:1	B	0
pDAS000150	29732-2A-29730	10:1	B	0
pDAS000150	n/a	n/a	D	0
pDAS000150	29732-2A-	5:1	D	4

	29730			
pDAS000150	29732-2A-29730	10:1	D	0

Table 8: Average NHEJ editing frequency in parts per million (ppm) across three biological replicates of SEC protoplast transfected with linear double-stranded donor DNA designs. “na” indicates “not applicable.”

Donor	ZFN	Donor-to-ZFN molar ratio	Editing Frequency in Wheat Sub-Genome	Editing Frequency (ppm)
pDAS000152	n/a	n/a	A	0
pDAS000152	29732-2A-29730	5:1	A	0
pDAS000152	29732-2A-29730	10:1	A	6717
pDAS000152	29732-2A-29730	20:1	A	5404
pDAS000152	n/a	n/a	B	0
pDAS000152	29732-2A-29730	5:1	B	0
pDAS000152	29732-2A-29730	10:1	B	6306
pDAS000152	29732-2A-29730	20:1	B	4106
pDAS000152	n/a	n/a	D	0
pDAS000152	29732-2A-29730	5:1	D	0
pDAS000152	29732-2A-29730	10:1	D	7911
pDAS000152	29732-2A-29730	20:1	D	4059
pDAS000153	n/a	n/a	A	0
pDAS000153	29732-2A-29730	5:1	A	0
pDAS000153	29732-2A-29730	10:1	A	0
pDAS000153	29732-2A-29730	20:1	A	0
pDAS000153	n/a	n/a	B	0
pDAS000153	29732-2A-29730	5:1	B	0
pDAS000153	29732-2A-29730	10:1	B	0
pDAS000153	29732-2A-29730	20:1	B	0
pDAS000153	n/a	n/a	D	0
pDAS000153	29732-2A-	5:1	D	0

	29730			
pDAS000153	29732-2A-29730	10:1	D	0
pDAS000153	29732-2A-29730	20:1	D	0
pDAS000149	n/a	n/a	A	0
pDAS000149	29732-2A-29730	5:1	A	0
pDAS000149	29732-2A-29730	10:1	A	0
pDAS000149	29732-2A-29730	20:1	A	344
pDAS000149	n/a	n/a	B	0
pDAS000149	29732-2A-29730	5:1	B	0
pDAS000149	29732-2A-29730	10:1	B	0
pDAS000149	29732-2A-29730	20:1	B	210
pDAS000149	n/a	n/a	D	0
pDAS000149	29732-2A-29730	5:1	D	4
pDAS000149	29732-2A-29730	10:1	D	0
pDAS000149	29732-2A-29730	20:1	D	24
pDAS000150	n/a	n/a	A	0
pDAS000150	29732-2A-29730	5:1	A	0
pDAS000150	29732-2A-29730	10:1	A	0
pDAS000150	29732-2A-29730	20:1	A	0
pDAS000150	n/a	n/a	B	0
pDAS000150	29732-2A-29730	5:1	B	0
pDAS000150	29732-2A-29730	10:1	B	0
pDAS000150	29732-2A-29730	20:1	B	0
pDAS000150	n/a	n/a	D	0
pDAS000150	29732-2A-29730	5:1	D	0
pDAS000150	29732-2A-29730	10:1	D	0
pDAS000150	29732-2A-29730	20:1	D	0

[0285] Collectively, the results provide strong molecular evidence for precise ZFN-mediated NHEJ-directed editing at the endogenous AHAS gene locus in wheat. These results show that all three sub-genomes can be targeted with a single ZFN and 5 donor. The results clearly demonstrate a higher frequency of editing for linear donor DNA designs as compared to plasmid donor DNA designs. Presumably, these results are due to the prerequisite requirement for *in planta* linearization of the plasmid donor molecules before they can participate in NHEJ-directed DNA repair. The results also indicate that sub-genome-specific mediated NHEJ-directed gene editing is facilitated 10 by a double strand break. The ZFNs that were designed to induce the double strand DNA breaks resulted in a sub-genome-specific mediated NHEJ-directed gene editing when delivered with the donor DNA to the *Triticum aestivum* plant cells.

Development of a transformation system for producing AHAS edited plants

[0286] The endogenous AHAS gene locus in wheat was selected as a model 15 locus to develop a transformation system for generating plants with precise genome modifications induced by ZFN-mediated gene editing. The endogenous AHAS gene was selected as a model locus due to its ability to produce a selectable phenotype (*i.e.*, tolerance to group B herbicides – ALS inhibitors), knowledge of prerequisite 20 information of sub-genome-specific gene coding sequence, and knowledge of specific mutations conferring tolerance to group B herbicides from the characterization of wheat with chemically induced mutations in the AHAS genes. The S653N mutation conferring tolerance to imidazolinone class herbicide was chosen as a target for ZFN-mediated gene editing due to the availability of commercially released wheat varieties 25 carrying the S653N mutation that could be used as positive controls to develop a chemical selection system to enrich for precisely edited events.

Molecular characterization of *Triticum aestivum* cv. Clearfield Janz

[0287] *Triticum aestivum* cv. Clearfield Janz, a commercially released bread 30 wheat variety carrying the S653N mutation in the D-genome, was selected for use as a positive control to develop a chemical selection strategy to enrich for AHAS edited wheat plants produced by ZFN-mediated gene editing. To generate a pure genetic seed stock, 48 seedlings were screened with 96 microsatellite (SSR) markers using

Multiplex-Ready PCR technology (Hayden *et al.*, (2008) BMC Genomics 9;80).

Seedlings with identical SSR haplotypes were used to produce seed that was used in subsequent experiments.

[0288] To ensure that the wheat plants used to produce seed carried the S653N mutation, a PCR assay was developed to amplify the region of the AHAS gene carrying the mutation from the D-genome of wheat. Sub-genome-specific amplification was achieved using on-off PCR (Yang *et al.*, (2005) Biochemical and Biophysical Research Communications 328:265-72) with primers AHAS-PS-6DF2 and AHAS-PS-6DR2 (SEQ ID NO:183 and SEQ ID NO:184) designed to position the penultimate base (which contained a phosphorothioate linkage) over nucleotide sequence variation that distinguished between the homoeologous copies of the AHAS genes. The PCR primers were designed to be between 18 and 27 nucleotides in length and to have a melting temperature of 60 to 65°C, optimal 63°C. The amplified PCR products were purified using a QIAQUICK MINIELUTE PCR PURIFICATION KIT™ (Qiagen) and sequenced using a direct Sanger sequencing method. The sequencing products were purified with ethanol, sodium acetate and EDTA following the BIGDYE® v3.1 protocol (Applied Biosystems) and electrophoresis was performed on an ABI3730XL® automated capillary electrophoresis platform.

[0289] Analysis of the amplified AHAS gene sequences using SEQUENCHER v3.7™ (GeneCodes, Ann Arbor, MI) revealed segregation for the S653N mutation and enabled the identification of plants that were homozygous (N653/N653) and heterozygous (N653/S653) for the S653N mutation or homozygous (S653/S653) for the herbicide-susceptible allele. The harvest of seed from individual plants provided a seed source having different levels of zygosity for the S653N mutation in the cv. Clearfield Janz genetic background.

Optimization of chemical selection conditions based on IMAZAMOX™

[0290] A series of experiments were performed to determine optimal selection conditions for regenerating AHAS edited wheat plants. These experiments were based on testing the basal tolerance to IMAZAMOX™ of the donor wheat line cv. Bobwhite MPB26RH (S653/S653 genotype) at the callus induction, plant regeneration and rooting stages of an established wheat transformation system. Similar experiments were performed to determine the basal tolerance and resistance of cv. Clearfield Janz

genotypes carrying the different doses of the S653N mutation; *i.e.*, plants with N653/N653 and S653/S653 genotypes.

[0291] The basal tolerance of the donor wheat line cv. Bobwhite MPB26RH and basal resistance of cv. Clearfield Janz (N653/N653) genotype to IMAZAMOX® at the callus induction stage was determined as follows: Scutella of immature zygotic embryos from each wheat line were isolated as described previously and placed in 10 cm PETRI™ dishes containing CIM medium supplemented with 0, 50, 100, 200, 300, 400 and 500 nM IMAZAMOX® respectively. Twenty scutella were placed in each PETRI™ dish. A total of 60 scutella from each of the donor wheat line cv. Bobwhite MPB26RH and cv. Clearfield Janz genotype were tested for basal tolerance and basal resistance response, respectively, at each IMAZAMOX® concentration. After incubation at 24°C in the dark for 4 weeks, the amount of somatic embryogenic callus formation (SEC) at each IMAZAMOX® concentration was recorded. The results showed that SEC formation for cv. Bobwhite MPB26RH was reduced by about 70% at 100 nM IMAZAMOX®, compared to untreated samples. Callus formation for the cv. Clearfield Janz genotype was unaffected, relative to the untreated control, at any IMAZAMOX® concentrations tested.

[0292] The basal tolerance of the donor wheat line cv. Bobwhite MPB26RH to IMAZAMOX® at the plant regeneration stage was determined as follows: Scutella of immature zygotic embryos from the donor wheat line were isolated as described previously and placed in 10 cm PETRI™ dishes containing CIM medium. Somatic embryogenic callus was allowed to form by incubating at 24°C in the dark for 4 weeks. The SEC was transferred to 10 cm PETRI™ dishes containing DRM medium supplemented with 0, 100, 200, 300, 400, 500 and 1000 nM IMAZAMOX® respectively. Twenty CIM were placed in each PETRI™ dish. A total of 60 CIM were tested for basal tolerance response at each IMAZAMOX® concentration. After incubation for 2 weeks at 24°C under a 16/8 (light/dark) hour photoperiod in a growth room, the regeneration response was recorded. The results showed that plant regeneration was reduced by about 80% at 200 nM IMAZAMOX®, compared to untreated samples.

[0293] The basal tolerance of the cv. Clearfield Janz (S653/S653) genotype and basal resistance of the cv. Clearfield Janz (N653/N653) genotype to IMAZAMOX® at the plant regeneration stage was determined using a modified

approach, as cv. Clearfield Janz was observed to have poor plant regeneration response (*i.e.*, poor embryogenesis) in tissue culture. Seed for each cv. Clearfield Janz genotype was germinated using the aseptic approach described above for producing wheat mesophyll protoplasts. The germinated seedlings were multiplied *in vitro* by 5 sub-culturing on multiplication medium. Following multiplication, plants for each genotype were transferred to 10 cm PETRITM dishes containing plant growth medium (MS +10 µM BA +0.8% agar) supplemented with 0, 100, 300, 600, 900, 1200, 1500 and 3000 nM IMAZAMOX®, respectively. Ten plants were placed in each PETRITM dish. A total of 30 plants per genotype were tested for basal response at each 10 IMAZAMOX® concentration. After incubation for 3 weeks at 24°C under a 16/8 (light/dark) hour photoperiod in a growth room, the growth response was recorded. The results showed that plant growth for the cv. Clearfield Janz (S653/S653) genotype was severely reduced in medium containing at least 200 nM IMAZAMOX®, compared to untreated samples. This response was similar to that 15 observed for the cv. Bobwhite MPB26RH (S653/S653) genotype. In contrast, plant growth for the cv. Clearfield Janz (N653/N653) genotype was not strongly suppressed, relative to untreated samples, until the IMAZAMOX® concentration exceeded 2,000 nM.

[0294] The basal tolerance of the donor wheat line cv. Bobwhite MPB26RH 20 to IMAZAMOX® at the plant rooting stage was determined as follows: Scutella of immature zygotic embryos from the donor wheat line were isolated as described previously and placed in 10 cm PETRITM dishes containing CIM medium. Somatic embryogenic callus was allowed to form by incubating at 24°C in the dark for 4 weeks. The SEC was transferred to 10 cm PETRITM dishes containing DRM medium 25 and incubated for 2 weeks at 24°C under a 16/8 (light/dark) hour photoperiod to allow plant regeneration to take place. Regenerated plants were transferred to 10 cm PETRITM dishes containing RM medium supplemented with 0, 100, 200, 300, 400, 500 nM IMAZAMOX®, respectively. Twenty regenerated plants were placed in each PETRITM dish. A total of 60 regenerated plants were tested for basal tolerance 30 response at each IMAZAMOX® concentration. After incubation for 3 weeks at 24°C under a 16/8 (light/dark) hour photoperiod in a growth room, the root formation response was recorded. The results showed that root formation was severely restricted at all concentrations of IMAZAMOX® tested, compared to untreated samples.

[0295] The basal tolerance of the cv. Clearfield Janz (S653/S653) genotype and basal resistance of the cv. Clearfield Janz (N653/N653) genotype to IMAZAMOX® at the plant rooting stage was determined using a modified approach, as cv. Clearfield Janz was observed to have poor plant regeneration response (*i.e.*, poor embryogenesis) in tissue culture. Seed for each cv. Clearfield Janz genotype was germinated using the aseptic approach described above for producing wheat mesophyll protoplasts. The germinated seedlings were multiplied *in vitro* by subculturing on multiplication medium. Following multiplication, plants for each genotype were transferred to 10 cm PETRI™ dishes containing plant rooting medium (1/2 MS, 0.5 mg/L NAA, 0.8% agar) supplemented with 0, 50, 100, 200 and 250 nM IMAZAMOX®, respectively. Three plants were placed in each PETRI™ dish. A total of 6 plants per genotype were tested for basal response at each IMAZAMOX® concentration. After incubation for 2 weeks at 24°C under a 16/8 (light/dark) hour photoperiod in a growth room, the root formation response was recorded.

[0296] The results showed that root formation for the cv. Clearfield Janz (N653/N653) genotype was restricted, compared to untreated samples, at 250 nM IMAZAMOX®. Root formation was severely restricted in the cv. Clearfield Janz (S653/S653) genotype at all concentrations of IMAZAMOX® tested, compared to untreated samples.

20

Design and Synthesis of Donor DNA for ZFN-Mediated AHAS Gene Editing

[0297] Donor DNA molecules were designed to promote precise ZFN-mediated NHEJ-directed gene editing at the endogenous AHAS genes in wheat. The donor designs allowed for the introduction of the S653N mutation known to confer tolerance to imidazolinone class herbicides (Li *et al.*, (2008) Molecular Breeding 22:217-225).

[0298] The first design was based on the integration of a 95-bp double stranded donor molecule at the position of the double strand DNA break created by cleavage of a homoeologous copy of the endogenous AHAS gene by ZFNs 29732 and 29730 (encoded on plasmid pDAB109350). The donor DNA molecule, pDAS000267 (SEQ ID NO:423 and SEQ ID NO:424), comprised two portions of the integrating donor polynucleotide. The 5' end contained sequence near identical to the endogenous AHAS gene encoded in the D-genome, starting from the target ZFN cleavage site and

finishing at the AHAS stop codon. Six intentional mutations were introduced into this sequence: two mutations encoded the S653N mutation (AGC→AAT), and four mutations were synonymous (in which a silent mutation was incorporated into the donor sequence). The 3' end of the donor molecule contained a unique sequence that

5 could be used for diagnostic PCR to detect ZFN-mediated NHEJ-directed gene editing events. The donor molecule was designed with protruding 5' and 3' ends to provide ligation overhangs to facilitate ZFN-mediated NHEJ-directed DNA repair.

[0299] The second design was based on replacement of the endogenous AHAS sequence located between a pair of ZFN target sites with a 79-bp double stranded donor molecule. Specifically, the donor was designed to replace the endogenous AHAS sequence released from chromatin upon dual cleavage of a homoeologous copy of the AHAS gene by ZFNs 29732 and 29730 (encoded on plasmid pDAB109350) and ZFNs 30012 and 30018 (encoded on plasmid pDAB109360). The donor molecule, pDAS000268 (SEQ ID NO:425 and SEQ ID NO:426), comprised sequence near identical to the endogenous AHAS gene encoded in the D-genome, starting from the cleavage site for ZFNs 29732 and 29730, and finishing at the cleavage site for ZFNs 30012 and 30018. Ten deliberate mutations were introduced into this sequence. Six mutations were located at the 5' end of the donor: two mutations encoded the S653N mutation (AGC→AAT) and four mutations were synonymous. Four mutations were located at the 3' end of the donor and were located in non-coding sequence. The donor molecule was designed with protruding 5' and 3' ends to provide ligation overhangs to facilitate ZFN-mediated NHEJ-directed DNA repair.

[0300] Standard phosphoramidite chemistry was used to synthetically synthesize the double stranded DNA donor molecules (Integrated DNA Technologies). For each donor molecule, a pair of complementary single stranded DNA oligomers was synthesized, each with two phosphorothioate linkages at their 5' ends to provide protection against *in planta* endonuclease degradation. The single stranded DNA oligomers were purified by high performance liquid chromatography to enrich for full-length molecules and purified of chemical carryover from the synthesis steps using Na⁺ exchange. The double stranded donor molecule was formed by annealing equimolar amounts of the two complementary single-stranded DNA oligomers using standard methods commonly known by one skilled in the art. Before

delivery to *Triticum aestivum*, the double stranded DNA molecules were diluted to the required concentration in sterile water.

Design and Production of Binary Vector Encoding AHAS (S653N)

5 [0301] Standard cloning methods were used in the construction of binary vector pDAS000143 (SEQ ID:185) (Figure 15). The AHAS (S653N) gene expression cassette consists of the promoter, 5' untranslated region and intron from the Ubiquitin (*Ubi*) gene from *Zea mays* (Toki *et al.*, (1992) *Plant Physiology* 100; 1503-07) followed by the coding sequence (1935 bp) of the AHAS gene from *T. aestivum* with 10 base-pairs 1880 and 1181 mutated from CG to AT in order to induce an amino acid change from serine (S) to asparagine (N) at amino acid residue 653. The AHAS expression cassette included the 3' untranslated region (UTR) of the *nopaline synthase* gene (*nos*) from *A. tumefaciens* pTi15955 (Fraley *et al.*, (1983) *Proceedings of the National Academy of Sciences U.S.A.* 80(15); 4803-4807). The selection cassette 15 was comprised the promoter, 5' untranslated region and intron from the actin 1 (*Act1*) gene from *Oryza sativa* (McElroy *et al.*, (1990) *The Plant Cell* 2(2); 163-171) followed by a synthetic, plant-optimized version of *phosphinothricin acetyl transferase* (PAT) gene, isolated from *Streptomyces viridochromogenes*, which encodes a protein that confers resistance to inhibitors of glutamine synthetase 20 comprising phosphinothricin, glufosinate, and bialaphos (Wohlleben *et al.*, (1988) *Gene* 70(1); 25-37). This cassette was terminated with the 3' UTR from the 35S gene of cauliflower mosaic virus (CaMV) (Chenault *et al.*, (1993) *Plant Physiology* 101 (4); 1395-1396).

25 [0302] The selection cassette was synthesized by a commercial gene synthesis vendor (GeneArt, Life Technologies) and cloned into a Gateway-enabled binary vector with the RfA Gateway cassette located between the Ubiquitin (*Ubi*) gene from *Zea mays* and the 3' untranslated region (UTR) comprising the transcriptional terminator and polyadenylation site of the nopaline synthase gene (*nos*) from *A. tumefaciens* pTi15955. The AHAS(S653N) coding sequence was amplified with 30 flanking attB sites and sub-cloned into pDONR221. The resulting ENTRY clone was used in a LR CLONASE II™ (Invitrogen, Life Technologies) reaction with the Gateway-enabled binary vector encoding the *phosphinothricin acetyl transferase* (PAT) expression cassette. Colonies of *E. coli* cells transformed with all ligation

reactions were initially screened by restriction digestion of miniprep DNA. Restriction endonucleases were obtained from New England BioLabs and Promega. Plasmid preparations were performed using the QIAPREP SPIN MINIPREP KIT™ or the PURE YIELD PLASMID MAXIPREP SYSTEM™ (Promega Corporation, WI) following the manufacturer's instructions. Plasmid DNA of selected clones was sequenced using ABI Sanger Sequencing and BIG DYE TERMINATOR v3.1™ cycle sequencing protocol (Applied Biosystems, Life Technologies). Sequence data were assembled and analyzed using the SEQUENCHER SOFTWARE™ (Gene Codes Corporation, Ann Arbor, MI).

10

Biostatic-mediated transformation system for generating AHAS edited wheat Plants

[0303] About 23,000 scutella of immature zygotic embryos from the donor wheat line cv. Bobwhite MPB26RH were prepared for biolistics-mediated DNA delivery, as described previously. DNA-coated gold particles were prepared as described above with the following formulations. For transfections performed using pDAS000267, the donor DNA was mixed at a 5:1 molar ratio with plasmid DNA for pDAB109350 (encoding ZFNs 29732 and 29730). For transfections performed using pDAS000268, the donor DNA was mixed at a 10:1:1 molar ratio with plasmid DNA for pDAB109350 (encoding ZFNs 29732 and 29730) and pDAB109360 (encoding ZFNs 30012 and 30018). Transfections performed using pDAS000143 were performed using gold particles that were coated only with plasmid DNA for pDAS000143.

15

20

25

30

[0304] Biostatic-mediated transfections were performed as described previously. A total of 15,620 scutella were bombarded with gold particles coated with DNA containing pDAS000267, a total of 7,310 scutella were bombarded with gold particles coated with DNA containing pDAS000268, and a total of 2,120 scutella were bombarded with gold particles coated with pDAS000143. Following bombardment, the transfected scutella were incubated at 26°C in the dark for 16 h before being transferred onto medium for callus induction.

[0305] Four different chemical selection strategies based on IMAZAMOX® were used to enrich for regenerated wheat plants that had the S653N mutation precisely integrated into one or more homoeologous copies of the endogenous AHAS

gene by ZFN-mediated NHEJ-directed gene editing. The four chemical selection strategies are described in Table 9. For each strategy, scutella were cultured in the dark on callus induction medium at 24°C for 2 weeks. The resultant calli were subcultured once onto fresh callus induction medium and kept in the same conditions for 5 a further two weeks. Somatic embryogenic callus (SEC) was transferred onto plant regeneration medium and cultured for 2 weeks at 24°C under a 16/8 (light/dark) hour photoperiod in a growth room. Regenerated plantlets were transferred onto rooting medium and cultured under the same conditions for 2-3 weeks. To increase stringency for the selection of regenerated plants having the S653N mutation, the roots of 10 regenerated plants were removed and the plants were again sub-cultured on rooting media under the same conditions. Plantlets rooting a second time were transferred to soil and grown under glasshouse containment conditions. T₁ seed was harvested from individual plants, following bagging of individual spikes to prevent out-crossing.

15 [0306] The scutella explants bombarded with gold particles coated with pDAS000143 were used to monitor the selection stringency across the four chemical selection strategies for regenerating wheat plants carrying the AHAS S653N mutation. Plants transformed with pDAS000143 were regenerated using process described above.

20 **Table 9: Chemical selection strategies used to regenerate wheat plants that had the S653N mutation precisely integrated into one or more homoeologous copies of the endogenous AHAS gene by ZFN-mediated NHEJ-directed gene editing. (IMI = IMAZAMOX™)**

Plant Regeneration Stage	Strategy 1	Strategy 2	Strategy 3	Strategy 4
Callus induction (CIM)	150 nM IMI	250 nM IMI	150 nM IMI	250 nM IMI
Plant Regeneration (DRM)	150 nM IMI	0 nM IMI	250 nM IMI	250 nM IMI
Rooting (RM)	200 nM IMI	200 nM IMI	200 nM IMI	200 nM IMI

25 [0307] Overall, 14 putatively ZFN-mediated NHEJ-directed AHAS edited wheat plants were recovered from the transfection of 22,930 scutella of immature zygotic embryos from the donor wheat line cv. Bobwhite MPB26RH. Putatively edited plants were obtained from all four selection strategies for scutella bombarded with gold particles coated with DNA containing pDAS000267. Two putatively edited plants were obtained from the second selection strategy for scutella bombarded with 30 gold particles coated with DNA containing pDAS000268. A total of 129 putatively

transformed wheat plants carrying at least one randomly integrated copy of the AHAS (S653N) donor polynucleotide were recovered across the four chemical selection strategies.

5 **Molecular Characterization of Edited Wheat Plants**

[0308] The wheat plants resulting from bombardments with a donor polynucleotide encoding the S653N mutation were obtained and molecularly characterized to identify the wheat sub-genomes that comprised an integration of the S653N mutation that occurred as a result of the donor integration at a genomic double strand cleavage site. Two series of bombardments were completed. The first set of experiments was completed with pDAS000143, and the second set of experiments was completed with pDAS000267 and pDAS000268. Individual wheat plants were obtained from both sets of experiments and assayed via a molecular method to identify plants which contained an integrated copy of the AHAS donor polynucleotide encoding the S653N mutation.

[0309] A hydrolysis probe assay (analogous to the TAQMAN® based assay) for quantitative PCR analysis was used to confirm that recovered wheat plants that had been bombarded with pDAS000143 carried at least one randomly integrated copy of the AHAS donor polynucleotide encoding the S653N mutation. Confirmation via 20 Sanger sequence analysis indicated that wheat plants recovered from bombardments performed with pDAS000267 and pDAS000268 comprised the S653N donor polynucleotide in at least one of the homoeologous copies of the AHAS gene at the position expected for ZFN-mediated NHEJ-directed gene editing.

25 **Genomic DNA isolation from regenerated wheat plants**

[0310] Genomic DNA was extracted from freeze-dried leaf tissue harvested from each regenerated wheat plant. Freshly harvested leaf tissue was snap frozen in liquid nitrogen and freeze-dried for 24 h in a LABCONCO FREEZONE 4.5® (Labconco, Kansas City, MO) at -40°C and 133×10^{-3} mBar pressure. The lyophilized 30 material was subjected to DNA extraction using the DNEASY® PLANT DNA EXTRACTION MINI KIT™ (Qiagen) following the manufacturer's instructions.

PCR assay to confirm random integration of AHAS donor polynucleotide encoding S653N mutation

[0311] To confirm that the regenerated wheat plants from bombardments performed with pDAS000143 carried at least one randomly integrated copy of the AHAS donor polynucleotide encoding the S653N mutation, a duplex hydrolysis probe qPCR assay (analogous to TAQMAN®) was used to amplify the endogenous single copy gene, puroindoline-b (Pinb), from the D genome of hexaploid wheat (Gautier *et al.*, (2000) Plant Science 153, 81–91; SEQ ID NO:186, SEQ ID NO:187 and SEQ ID NO:188 for forward and reverse primers and probe sequence, respectively) and a region of the Actin (Act1) promoter present on pDAS000143 (SEQ ID NO:189, SEQ ID NO:190 and SEQ ID NO:191 for forward and reverse primers and probe sequence, respectively). Hydrolysis probe qPCR assays were performed on 24 randomly chosen wheat plants that were recovered from each of the four chemical selection strategies. Assessment for the presence, and estimated copy number of pDAS00143 was performed according to the method described in Livak and Schmittgen (2001) Methods 25(4):402-8.

[0312] From the results, conclusive evidence was obtained for the integration of at least one copy of the AHAS donor polynucleotide encoding the S653N mutation into the genome of each of the wheat plants tested. These results indicate that the four chemical selection strategies provided stringent selection for the recovery of plants expressing the S653N mutation.

PCR assay of genomic DNA for ZFN-mediated AHAS editing

[0313] To characterize the sub-genomic location and outcome of ZFN-mediated NHEJ-directed gene editing in the recovered wheat plants, PCR with primers AHAS_3F1 and AHAS_3R1 (SEQ ID NO:192 and SEQ ID NO:193) was used to amplify the target region from the homoeologous copies of the AHAS genes. The resulting PCR products were cloned into plasmid vector and Sanger sequenced using BIGDYE® v3.1 chemistry (Applied Biosystems) on an ABI3730XL® automated capillary electrophoresis platform. Sanger sequencing of up to 120 independent plasmid clones was performed to ensure that each allele at the endogenous AHAS homoeologs was sequenced. Sequence analysis performed using SEQUENCER SOFTWARE™ was used to generate a consensus sequence for each

allele of the three homoeologous copies of the AHAS gene in each of the recovered wheat plants, and to determine the sub-genomic origin and sequence for each edited allele.

[0314] From the results, conclusive evidence for precise ZFN-mediated NHEJ-directed gene editing at the endogenous AHAS loci was demonstrated for 11 of the 12 recovered wheat plants that were transformed using pDAB109350 and pDAS000267 (Table 10), and both of the recovered wheat plants that were transformed using pDAB109350, pDAB109360 and pDAS000268 (Table 11). Plants with a range of editing outcomes were observed including: (1) independent events with perfect sub-genome-specific allele edits; (2) events with single perfect edits in the A-genome, B-genome and D-genomes; (3) events with simultaneous editing in multiple sub-genomes; and, (4) events demonstrating hemizygous and homozygous sub-genome-specific allele editing. Disclosed for the first time is a method which can be utilized to mutate a gene locus within all three genomes of a wheat plant.

15 Wheat plants comprising an integrated AHAS donor polynucleotide encoding a S653N mutation are exemplified; integration of the polynucleotide sequence provides tolerance to imidazolinone class herbicides. The utilization of ZFN-mediated genomic editing at an endogenous gene locus in wheat allows for the introduction of agronomic traits (via mutation) without time consuming wheat breeding techniques

20 which require backcrossing and introgression steps that can increase the amount of time required for introgressing the trait into all three sub-genomes. Consensus Sanger sequences for the alleles present in each sub-genome for the edited wheat plants are provided as SEQ ID NO:194-277 in Tables 10 and 11.

Table 10: ZFN-mediated NHEJ-directed AHAS editing outcomes for wheat plants transformed using pDAB109350 and pDAS000267

		A-genome		B-genome		D-genome		SEQ ID NO:
		Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	
Plant No.1	Status	PE	NHEJ	IE	UE	IE	UE	194-199
	No. clones ¹	13	20	12	19	14	22	
Plant No.2	Status	NHEJ	UE	UE	nd	IE	UE	200-205
	No. clones ¹	9	3	16	0	75	17	
Plant No.3	Status	PE	UE	UE	nd	UE	nd	206-211
	No. clones ¹	7	11	29	0	35	0	
Plant	Status	PE	UE	IE	UE	PE	IE	212-217

No.4	No. clones ¹	6	11	44	30	6	11	
Plant No.5	Status	PE	UE	NHEJ	UE	UE	nd	218-223
	No. clones ¹	10	9	15	26	21	0	
Plant No.6	Status	UE	nd	PE	UE	UE	nd	224-229
	No. clones ¹	22	0	11	18	43	0	
Plant No.7	Status	PE	UE	UE	nd	UE	nd	230-235
	No. clones ¹	5	12	26	0	22	0	
Plant No.8	Status	UE	nd	UE	nd	UE	nd	236-241
	No. clones ¹	32	0	40	0	26	0	
Plant No.9	Status	PE	nd	IE	UE	UE	nd	242-247
	No. clones ¹	24	0	13	21	33	0	
Plant No.10	Status	PE	UE	UE	nd	UE	nd	248-253
	No. clones ¹	10	19	37	0	29	0	
Plant No.11	Status	UE	nd	UE	nd	PE	UE	254-259
	No. clones ¹	35	0	37	0	15	11	
Plant No.12	Status	UE	nd	UE	nd	IE	NHEJ	260-265
	No. clones ¹	34	0	40	0	14	8	

¹Number of independent plasmid clones sequenced.

PE = perfect edit; *i.e.*, ZFN-mediated NHEJ-directed genome editing produced a predicted outcome.

IE = imperfect edit; *i.e.*, ZFN-mediated NHEJ-directed genome editing produced an unpredicted outcome.

UE = unedited allele; *i.e.*, allele had wild-type sequence.

nd = not detected; *i.e.*, sufficient independent plasmid clones were sequenced to conclude that an alternate allele was not present and that the locus was homozygous for a single allele.

5 NHEJ = Non Homologous End Joining; *i.e.*, evidence for a non-homologous end joining DNA repair outcome that did not result in the integration of a donor molecule at the ZFN cleavage site.

10 **Table 11: ZFN-mediated NHEJ-directed AHAS editing outcomes for wheat plants transformed using pDAB109350, pDAB109360 and pDAS000268.**

		A-genome		B-genome		D-genome		SEQ ID NO:
		Allele 1	Allele 2	Allele 1	Allele 2	Allele 1	Allele 2	
Plant No.12a	Status	IE	UE	UE	nd	IE	nd	266-271
	No. clones ¹	5	14	53	0	1	24	
Plant	Status	IE	UE	UE	nd	UE	nd	272-277

No.13a	No. clones ¹	10	12	49	0	18	0	
--------	-------------------------	----	----	----	---	----	---	--

¹Number of independent plasmid clones sequenced.

IE = imperfect edit; *i.e.*, ZFN-mediated NHEJ-directed genome editing produced unexpected outcome.

UE = unedited allele; *i.e.*, allele had wild-type sequence.

5 nd = not detected; *i.e.*, sufficient independent plasmid clones were sequenced to conclude that an alternate allele was not present and that the locus was homozygous for a single allele.

Design of zinc finger binding domains specific to region in AHAS genes encoding the P197 amino acid residue

[0315] Zinc finger proteins directed against DNA sequence of the homoeologous copies of the AHAS genes were designed as previously described.

Exemplary target sequence and recognition helices are shown in Table 12 (recognition helix regions designs) and Table 13 (target sites). In Table 13,

15 nucleotides in the target site that are contacted by the ZFP recognition helices are indicated in uppercase letters; non-contacted nucleotides are indicated in lowercase. Zinc Finger Nuclease (ZFN) target sites were designed upstream (from 2 to 510 nucleotides upstream) of the region in the AHAS gene encoding the proline 197 (P197) amino acid residue.

20 **Table 12: AHAS zinc finger designs (N/A indicates “Not Applicable”)**

ZFP#	F1	F2	F3	F4	F5	F6
3447 0	SEQ ID NO:278 RSADLSR	SEQ ID NO:279 RSDHLSA	SEQ ID NO:280 QSSDLRR	SEQ ID NO:160 DRSNLSR	SEQ ID NO:281 RSDDRKT	N/A
3447 1	SEQ ID NO:158 QSGDLTR	SEQ ID NO:282 RRADRAK	SEQ ID NO:150 RSDDLTR	SEQ ID NO:283 TSSDRKK	SEQ ID NO:284 RSADLTR	SEQ ID NO:285 RNDDRKK
3447 2	SEQ ID NO:284 RSADLTR	SEQ ID NO:286 DRSNLTR	SEQ ID NO:287 ERGTLAR	SEQ ID NO:150 RSDDLTR	SEQ ID NO:288 DRSDLSR	SEQ ID NO:289 DSSTRRR
3447 3	SEQ ID NO:290 RSDHLSE	SEQ ID NO:291 HSRTRTK	SEQ ID NO:292 RSDTLSE	SEQ ID NO:293 NNRDRTK	SEQ ID NO:287 ERGTLAR	SEQ ID NO:155 DRSALAR
3447 4	SEQ ID NO:287 ERGTLAR	SEQ ID NO:150 RSDDLTR	SEQ ID NO:294 DRSDLSR	SEQ ID NO:289 DSSTRRR	SEQ ID NO:286 DRSNLTR	N/A
3447 5	SEQ ID NO:296 RSDHLSR	SEQ ID NO:297 QQWDRKQ	SEQ ID NO:298 DRSHLTR	SEQ ID NO:152 DSSDRKK	SEQ ID NO:160 DRSNLSR	SEQ ID NO:299 VSSNLTS

Table 13: Target site of AHAS zinc fingers

pDAB#	Approximate Cleavage Site Relative to AHAS Pro-197	ZFP # and Binding Site (5'→3')	SEQ ID NO:
pDAB111850 (34456-2A-34457)	499-bp upstream	34456: cnCGGCCATGGCGGCGGCGagg gtttg	300
		34457: acCTCcCCCGCCGTCGCAttctengcg g	301
pDAB111855 (34470-2A-34471)	109-bp upstream	34470: ggCCGGACGCGCGGGCGtancggac cgc	302
		34471: cgTCGGCGTCTGCGTCGCCAcctcc ggc	303
pDAB111856 (34472-2A-34473)	99-bp upstream	34472: acGCCGACGCGGCCgGACGCGcgg gcgt	304
		34473: gcGTCGCCaCCTCCGGCCCGGggg ccac	305
pDAB111857 (34474-2A-34475)	96-bp upstream	34474: caGACGCCGACGCGGCCggacgcgc ggg	306
		34475: gtGCCACcTCCGGCCCGGGGGgcc acca	307
pDAB111858 (34476-2A-34477)	90-bp upstream	34476: gcGACGCAGACGCCGACgcggccgg acg	308
		34477: ccTCCGGCCCGGGGGGCCaccaacctc gt	309
pDAB111859 (34478-2A-34479)	24-bp upstream	34478: ggGATGGAGTCGAGGAGngcgteng cga	310
		34479: tgGTCGCCATCACGGGCCAGgtccc ccg	311
pDAB111860 (34480-2A-34481)	18-bp upstream	34480: acCATGGGGATGGAGTCGAGgagn gcgt	312
		34481: ccATCACGGGCCAGGTCCCCGCCgc at	313
pDAB111861	16-bp	34482:	314

(34482-2A-34483)	upstream	cgACCATGGGGATGGAGTCGagga gngc 34483: caTCACGGGCCAGGTCCcccgccgca tg	
			315

[0316] The AHAS zinc finger designs were incorporated into zinc finger expression vectors and verified for cleavage activity using a budding yeast system, as described previously. Of the numerous ZFNs that were designed, produced and tested to bind to the putative AHAS genomic polynucleotide target sites, the ZFNs described above were identified as having *in vivo* activity at high levels, and selected for further experimentation. These ZFNs were designed to bind to the three homoeologous AHAS and were characterized as being capable of efficiently binding and cleaving the unique AHAS genomic polynucleotide target sites *in planta*.

10

Evaluation of zinc finger nuclease cleavage of AHAS genes using transient assays

[0317] **ZFN Construct Assembly;** Plasmid vectors containing ZFN expression constructs verified for cleavage activity using the yeast system were designed and completed as previously described. The resulting plasmid constructs; pDAB111855 (ZFNs 34470-2A-34471), pDAB111856 (ZFNs 34472-2A-34473), and pDAB111857 (ZFNs 34474-2A-34475) were confirmed via restriction enzyme digestion and via DNA sequencing.

20 [0318] Before delivery to *Triticum aestivum* protoplasts, plasmid DNA for each ZFN construct was prepared from cultures of *E. coli* using the PURE YIELD PLASMID MAXIPREP SYSTEM® (Promega Corporation, Madison, WI) or PLASMID MAXI KIT® (Qiagen, Valencia, CA) following the instructions of the suppliers.

25

Isolation and transfection of wheat mesophyll protoplasts

[0319] Mesophyll protoplasts from the donor wheat line cv. Bobwhite MPB26RH were prepared and transfected using polyethylene glycol (PEG)-mediated DNA delivery as previously described.

30

PCR assay of protoplast genomic DNA for ZFN sequence cleavage

[0320] Genomic DNA was isolated from transfected protoplasts and used for PCR assays to assess the cleavage efficiency and target site specificity of ZFNs designed to the region of the AHAS gene encoding P197, as previously described.

5 Five sets of PCR primers which contained a phosphorothioate linkage as indicated by the asterisk [*] were used to amplify the ZFN target site loci (Table 14). Each primer set was designed according to criteria previously described.

Table 14: Primer sequences used to assess AHAS ZFN cleavage efficacy and target site specificity.

Primer Name	Primer Set	Primer Sequence (5'→3')	SEQ ID NO:
AHAS-P197ZFN.F2	Set 1	a*cactttccctacacgacgcttccgatctTCC CCAATTCCAACCCCTCT*C	316
AHAS-P197ZFN.R1	Set 1	g*tgactggagttcagacgtgtcttccgatctC GTCAGCGCCTGGTGGATC*T	317
AHASs653ZFN.F5	Set 2	a*cactttccctacacgacgcttccgatctGC CCGTCCGAGCCCCGCA*A	318
AHASs653ZFN.R1	Set 2	g*tgactggagttcagacgtgtcttccgatctC GTCAGCGCCTGGTGGATC*T	319
AHASs653ZFN.F7	Set 3	a*cactttccctacacgacgcttccgatctGC GCTCGCCCGTCATCA*C	320
AHASs653ZFN.R5	Set 3	g*tgactggagttcagacgtgtcttccgatctA TGGGGATGGAGTCGAGGA*G	321
AHAS+400ZFN.F9	Set 4	a*cactttccctacacgacgcttccgatctCTT CCGCCACGAGCAGG*G	322
AHAS+400ZFN.R5	Set 4	g*tgactggagttcagacgtgtcttccgatctA TGGGGATGGAGTCGAGGA*G	323
AHAS+400ZFN.F11	Set 5	a*cactttccctacacgacgcttccgatctTC GTCTCCGCGCTCGCTG*A	324
AHAS+400ZFN.R6	Set 5	g*tgactggagttcagacgtgtcttccgatctTC CACTATGGCGTCTCCT*G	325

10

Data analysis for detecting NHEJ at target ZFN sites

[0321] Following generation of Illumina short read sequence data for sample libraries prepared for transfected mesophyll protoplasts, bioinformatics analysis (as previously described) was performed to identify deleted nucleotides at the target ZFN sites. Such deletions are known to be indicators of *in planta* ZFN activity that result from non-homologous end joining (NHEJ) DNA repair.

[0322] Two approaches were used to assess the cleavage efficiency and specificity of the ZFNs tested. Cleavage efficiency was expressed (in parts per million reads) as the proportion of sub-genome assigned sequences that contained a NHEJ

deletion at the ZFN target site (Table 15). Rank ordering of the ZFNs by their observed cleavage efficiency was used to identify ZFNs with the best cleavage activity for the target region of the AHAS genes in a sub-genome-specific manner. All of the ZFNs tested showed NHEJ deletion size distributions consistent with that expected for *in planta* ZFN activity. Cleavage specificity was expressed as the ratio of cleavage efficiencies observed across the three sub-genomes.

Table 15: ZFN cleavage efficacy (expressed as number of NHEJ events per million reads) and target site specificity.

ZFN	A-genome	B-genome	D-genome
pDAB111855 (34470-2A-34471)	177,866	156,139	134,694
pDAB111856 (34472-2A-34473)	119,857	100,300	87,770
pDAB111857 (34474-2A-34475)	248,115	251,142	202,711

10 [0323] From these results, the ZFNs encoded on plasmids pDAB111855 (34470-2A-34471), pDAB111856 (34472-2A-34473) and pDAB111857 (34474-2A-34475) were selected for *in planta* targeting in subsequent experiments, given their characteristics of significant genomic DNA cleavage activity in each of the three wheat sub-genomes.

15

Generation of molecular evidence for ZFN-mediated, exogenous marker-free sequential transgene stacking at an endogenous AHAS locus using transient assays

[0324] The generation of molecular evidence using transient assays for ZFN-mediated, sequential exogenous marker-free transgene stacking at an endogenous AHAS locus within the genome of *Triticum aestivum* cells via homology directed DNA repair is achieved as follows.

20

[0325] The AHAS (S653N) edited wheat plants, which were produced via transformation with donor pDAS000267 and the Zinc Finger Nuclease encoded on plasmid pDAB109350, demonstrate the first step for sequential, exogenous marker-free transgene stacking at an endogenous AHAS locus in the genome of wheat. These edited plants are used to generate explant material (*e.g.*, protoplasts or scutella of immature zygotic embryos) for transfection using the previously described methods. The explant material is subsequently co-transfected with a donor DNA molecule and a

plasmid encoding a ZFN (e.g., pDAB111855, pDAB111856 or pDAB111857) that is designed to target a Zinc Finger binding site located in the AHAS genes upstream of the region encoding the P197 amino acid residue. The ZFN cleaves an AHAS locus and the donor molecule is integrated within the genome of *Triticum aestivum* cells via 5 homology directed repair. As a result of NHEJ-mediated donor molecule integration, the AHAS(P197S) mutation conferring tolerance to sulfonylurea class herbicides is introduced into the endogenous AHAS sequence and simultaneously, the AHAS(S653N) mutation introduced in the first round of transgene stacking is removed. Consequently, the expression of the endogenous AHAS gene is changed 10 from conferring tolerance to imidazolinones and susceptibility to sulfonylureas (the phenotype of correctly targeted wheat cells in the first round of transgene stacking) to conferring susceptibility for imidazolinones and tolerance for sulfonylureas, thus allowing for the regeneration of correctly targeted cells using a sulfonylurea selection agent. Molecular evidence for the integration of the donor DNA and generation of 15 correctly targeted wheat cells is confirmed using the previously described methods.

15 [0326] It is appreciated by those skilled in the art that co-transformation of wheat cells with a donor DNA molecule that contains one or more transgenes and a plasmid encoding a Zinc Finger Nuclease enables both parallel (simultaneous) or sequential transgene integration (transgene stacking) in plant genomes at precisely the 20 same genomic location, including simultaneous editing of multiple alleles across multiple genomes in polyploid plant species.

Development of a transformation system for sequential, exogenous marker-free transgene stacking at the endogenous AHAS loci in wheat

25 [0327] The endogenous AHAS gene in wheat was selected as a model locus to develop a ZFN-mediated, exogenous marker-free transformation system for generating plants with one or more transgenes precisely positioned at the same genomic location. The transformation system enables parallel (simultaneous integration of one or more transgenes) or sequential stacking (consecutive integration 30 of one or more transgenes) at precisely the same genomic location, including simultaneous parallel or sequential stacking at multiple alleles across multiple sub-genomes, by exploiting known mutations in the AHAS gene that confer tolerance to Group B herbicides. ZFN-mediated integration of a donor DNA into the wild-type

(herbicide susceptible) AHAS locus is used to introduce transgene(s) and a mutation to the endogenous AHAS gene that confers tolerance to imidazolinones, thus allowing the regeneration of correctly targeted plants using an imidazolinone selection agent. Stacking of a second transgene(s) at the AHAS locus is achieved by integration of a 5 donor DNA that introduces one or more additional transgenes and confers susceptibility to imidazolinones but tolerance to sulfonylureas, thus allowing the regeneration of correctly targeted plants using a sulfonylurea selection agent. Stacking of a third transgene can be achieved by integration of a donor molecule that introduces further transgene(s) and confers susceptibility to sulfonylurea and 10 tolerance to imidazolinones, thus allowing the regeneration of correctly targeted plants using an imidazolinone selection agent. As such, continued rounds of sequential transgene stacking are possible by the use of donor DNA that introduce transgene(s) and mutations at the endogenous AHAS genes for differential cycling between imidazolinone and sulfonylurea selection agents. The transgenes can be 15 integrated within the AHAS gene and stacked via an NHEJ pathway. The NHEJ repair and recombination pathway can be determined by the design of the donor transgene. In an embodiment, transgenes that are integrated and stacked within the AHAS gene would be designed to contain single or double cut ZFN sites that flank the payload (e.g., AHAS mutation and gene of interest). Accordingly, such a design 20 would utilize an NHEJ pathway for the integration and stacking of the donor polynucleotide within the chromosome.

Generation of low-copy, randomly integrated T-DNA wheat plants with AHAS(P197S) expression constructs

25 [0328] A binary vector pDAS000164 (SEQ ID NO:326, Figure 16) containing the AHAS(P197S) expression and PAT selection cassettes was designed and assembled using skills and techniques commonly known in the art. The AHAS (P197S) expression cassette consisted of the promoter , 5' untranslated region and intron from the Ubiquitin (*Ubi*) gene from *Zea mays* (Toki *et al.*, (1992) *Plant 30 Physiology*, 100;1503-07) followed by the coding sequence (1935 bp) of the AHAS gene from *T. aestivum* cv. Bobwhite MPB26RH with nucleotide 511 mutated from C to T in order to induce an amino acid change from proline (P) to serine (S). The AHAS expression cassette included the 3' untranslated region (UTR) comprising of

the nopaline synthase gene (*nos*) from *A. tumefaciens* pTi15955 (Fraley *et al.*, (1983) *Proceedings of the National Academy of Sciences U.S.A.* 80(15): 4803-4807). The selection cassette was comprised of the promoter, 5' untranslated region and intron from the actin 1(*Act1*) gene from *Oryza sativa* (McElroy *et al.*, (199) *The Plant Cell* 2(2): 163-171) followed by a synthetic, plant-optimized version of *phosphinothricin acetyl transferase* (PAT) gene, isolated from *Streptomyces viridochromogenes*, which encodes a protein that confers resistance to inhibitors of glutamine synthetase comprising phosphinothricin, glufosinate, and bialaphos (Wohlleben *et al.*, (1988) *Gene*, 70(1): 25-37). This cassette was terminated with the 3' UTR from the 35S 5 gene of cauliflower mosaic virus (CaMV) (Chenault *et al.*, (1993) *Plant Physiology* 101 (4): 1395-1396).

10 [0329] The selection cassette was synthesized by a commercial gene synthesis vendor (e.g., GeneArt, Life Technologies, etc.) and cloned into a GATEWAY®-enabled binary vector with the RfA Gateway cassette located between the Ubiquitin 15 (*Ubi*) gene from *Zea mays* and the 3' untranslated region (UTR) comprising the transcriptional terminator and polyadenylation site of the nopaline synthase gene (*nos*) from *A. tumefaciens* pTi15955. The AHAS (P197S) coding sequence was amplified with flanking attB sites and sub-cloned into pDONR221. The resulting ENTRY clone was used in a LR CLONASE II® (Invitrogen, Life Technologies) reaction with the 20 Gateway-enabled binary vector encoding the *phosphinothricin acetyl transferase* (PAT) expression cassette. Colonies of all assembled plasmids were initially screened by restriction digestion of miniprep DNA. Restriction endonucleases were obtained from New England BioLabs (NEB; Ipswich, MA) and Promega (Promega Corporation, WI). Plasmid preparations were performed using the QIAPREP SPIN 25 MINIPREP KIT® (Qiagen, Hilden) or the PURE YIELD PLASMID MAXIPREP SYSTEM® (Promega Corporation, WI) following the instructions of the suppliers. Plasmid DNA of selected clones was sequenced using ABI Sanger Sequencing and BIG DYE TERMINATOR V3.1® cycle sequencing protocol (Applied Biosystems, Life Technologies). Sequence data were assembled and analyzed using the 30 SEQUENCHER™ software (Gene Codes Corporation, Ann Arbor, MI).

35 [0330] The resulting binary expression clone pDAS000164 was transformed into *Agrobacterium tumefaciens* strain EHA105. Transgenic wheat plants with randomly integrated T-DNA were generated by *Agrobacterium*-mediated

transformation using the donor wheat line cv. Bobwhite MPB26RH, following a protocol similar to Wu *et al.* (2008) *Transgenic Research* 17:425-436. Putative T₀ transgenic events expressing the AHAS (P197) expression constructs were selected for phosphinothricin (PPT) tolerance, the phenotype conferred by the PAT selectable marker, and transferred to soil. The T₀ plants were grown under glasshouse containment conditions and T₁ seed was produced.

5 [0331] Genomic DNA from each T₀ plant was extracted from leaf tissue, as previously described, and tested for the presence of *Agrobacterium tumefaciens* and for the number of integrated copies of the T-DNA encoding AHAS(P197S). The presence of *A. tumefaciens* was performed using a duplex hydrolysis probe qPCR assay (analogous to TAQMAMTM) to amplify the endogenous ubiquitin gene (SEQ ID NO:327, SEQ ID NO:328, and SEQ ID NO:329 for forward and reverse primers and probe sequence, respectively) from the wheat genome, and virC from pTiBo542 (SEQ ID NO:330, SEQ ID NO:331, and SEQ ID NO:332 for forward and reverse primers and probe sequence, respectively). The number of integrated T-DNA copies was estimated using a duplex hydrolysis probe qPCR assay, as previously described, based on the puroindoline-b (Pinb) from the D genome of hexaploid wheat and a region of the Actin (Act1) promoter present on pDAS000164. Overall, 35 independent T₀ events with fewer than three randomly integrated copies of T-DNA were generated.

20

Optimization of chemical selection conditions based on sulfometuron methyl

[0332] A series of experiments were performed to determine optimal selection conditions for regenerating wheat plants expressing the AHAS(P197S) mutation conferring tolerance to sulfonylurea class herbicides. These experiments were based on testing the basal tolerance of the wild-type donor wheat line cv. Bobwhite MPB26RH (P197/P197 genotype, which confers susceptibility to sulfonylureas) at the callus induction, plant regeneration and rooting stages of an established wheat transformation system. Similar experiments were performed to determine the basal tolerance of transgenic cv. Bobwhite MPB26RH events that had randomly integrated T-DNA expressing the AHAS(P197) mutation, which confers tolerance to sulfonylurea selection agents.

[0333] The basal tolerance of the wild-type donor wheat line to sulfometuron methyl at the callus induction stage was determined as follows: Scutella of immature

zygotic embryos were isolated, as previously described, and placed in 10 cm PETRITM dishes containing CIM medium supplemented with 0, 100, 500, 1000, 1500 and 2000 nM sulfometuron methyl, respectively. Twenty scutella were placed in each PETRITM dish. A total of 60 scutella were tested at each sulfometuron methyl concentration. After incubation at 24°C in the dark for 4 weeks, the amount of somatic embryogenic callus formation (SEC) at each sulfometuron methyl concentration was recorded. The results showed that SEC transformation for cv. Bobwhite MPB26RH was reduced by about 70% at 100 nM sulfometuron methyl, compared to untreated samples.

10 [0334] The basal tolerance of the wild-type donor wheat line to sulfometuron methyl at the plant regeneration stage was determined as follows: Scutella of immature zygotic embryos from the donor wheat line were isolated and placed in 10 cm PETRITM dishes containing CIM medium. Somatic embryogenic callus was allowed to form by incubating at 24°C in the dark for 4 weeks. The SEC was transferred to 10 cm PETRITM dishes containing DRM medium supplemented with 0, 100, 500, 1000, 1500, 2000, 2500 and 3000 nM sulfometuron methyl, respectively. Twenty CIM were placed in each PETRITM dish. A total of 60 CIM were tested for basal tolerance response at each sulfometuron methyl concentration. After incubation for 2 weeks at 24°C under a 16/8 (light/dark) hour photoperiod in a growth room, the regeneration response was recorded. The results showed that plant regeneration was reduced by about 80% at 2000 nM sulfometuron methyl, compared to untreated samples.

15 [0335] The basal tolerance of the wild-type donor wheat line to sulfometuron methyl at the plant rooting stage was determined as follows: Scutella of immature zygotic embryos were isolated and placed in 10 cm PETRITM dishes containing CIM medium. Somatic embryogenic callus was allowed to form by incubating at 24°C in the dark for 4 weeks. The SEC was transferred to 10 cm PETRITM dishes containing DRM medium and incubated for 2 weeks at 24°C under a 16/8 (light/dark) hour photoperiod to allow plant regeneration to take place. Regenerated plants were transferred to 10 cm PETRITM dishes containing RM medium supplemented with 0, 100, 200, 250, 300, 400, 500, 1000 and 2000 nM sulfometuron methyl, respectively. Ten regenerated plants were placed in each PETRITM dish. A total of 30 regenerated plants were tested for basal tolerance response at each sulfometuron methyl

concentration. After incubation for 3 weeks at 24°C under a 16/8 (light/dark) hour photoperiod in a growth room, the root formation response was recorded. The results showed that root formation was severely inhibited when concentrations of sulfometuron methyl higher than 400 nM, compared to untreated samples.

5 [0336] The basal tolerance of transgenic wheat events with randomly integrated, low-copy (≤ 3) T-DNA expressing the AHAS(P197S) mutation to sulfometuron methyl at the plant rooting stage was determined as follows: Four independent transgenic events were randomly selected and multiplied *in vitro* by sub-culturing on multiplication medium. Following multiplication, plants for each event
10 were transferred to 10 cm PETRITM dishes containing RM medium supplemented with 0, 400, 450, 500, 550 and 600 nM sulfometuron methyl, respectively. Four plants (one from each of the four events) were placed in each PETRITM dish. A total of 3 plants per event was tested for basal tolerance at each sulfometuron methyl concentration. After incubation for 2 weeks at 24°C under a 16/8 (light/dark) hour
15 photoperiod in a growth room, the root formation response was recorded. The results showed that root formation was not restricted, compared to untreated controls, at any of the concentrations tested, indicating that the AHAS(P197S) mutation conferred high tolerance to sulfometuron methyl.

20 **Design and synthesis of donor DNA for first sequential transgene stacking at an endogenous AHAS locus using NHEJ-directed DNA repair**

[0337] The donor DNA for the first round of transgene stacking is designed to promote precise donor integration at an endogenous AHAS locus via ZFN-mediated, NHEJ-directed repair. The design is based on the integration of a double stranded
25 donor molecule at the position of the double strand DNA break created by cleavage of a homoeologous copy of the endogenous AHAS gene by ZFNs 29732 and 29730 (encoded on plasmid pDAB109350). The donor molecule (pDAS000433; SEQ ID NO:333, Figure 17) several portions of polynucleotide sequences. The 5' end contains sequence near identical to the endogenous AHAS gene encoded in the D-genome,
30 starting from the target ZFN cleavage site and finishing at the AHAS stop codon. Seven deliberate mutations are introduced into this sequence: two mutations encode the S653N mutation and five codon-optimized, synonymous mutations positioned across the binding site of ZFN 29732 to prevent re-cleavage of the integrated donor.

Following the stop codon is 316-bp of non-coding sequence corresponding to the conserved 3' untranslated region (3'UTR) across the AHAS homoeologs. The 3'UTR sequence is followed by Zinc Finger binding sites for ZFNs 34480 and 34481 (encoded on plasmid pDAB111860) and ZFNs 34482 and 34483 (encoded on plasmid pDAB111861). These Zinc Finger binding sites allow for self-excision of donor-derived AHAS (coding and 3'UTR) sequence integrated at the endogenous locus during the next round of transgene stacking. The self-excision Zinc Finger binding sites are followed by several additional Zinc Finger binding sites (each of which is separated by 100-bp of random sequence) that flank two unique restriction

5 endonuclease cleavage sites and which enable the insertion of a transgene expression cassette (e.g. the PAT expression cassette, as described previously) into the donor molecule. The additional Zinc Finger binding sites enable future excision of

10 transgenes integrated at an AHAS locus by sequential marker-free transgene stacking, or continued sequential transgene stacking at the same genomic location using an

15 alternate stacking method.

[0338] The donor cassette is synthesized by a commercial gene service vendor (e.g., GeneArt, Life Sciences, etc.) with a short stretch of additional flanking sequence at the 5' and 3' ends to enable generation of a donor molecule with protruding 5' and 3' ends that are compatible with the ligation overhangs generated by ZFNs 29732 and

20 29730 (encoded on plasmid pDAB109350) upon cleavage of an endogenous AHAS locus. The donor molecule with protruding 5' and 3' ends is generated by digesting plasmid DNA containing the donor molecule with the restriction endonuclease *BbsI* using standard methods known to the person having skill in the art.

25 **Design and synthesis of donor DNA for second sequential transgene stack at an endogenous AHAS locus using NHEJ-directed DNA repair**

[0339] The donor DNA for the second round of transgene stacking is designed to promote precise donor integration at the same AHAS locus targeted in the first transgene stack via ZFN-mediated, NHEJ-directed repair. The design is based on the

30 integration of a double stranded donor molecule at the double strand DNA break created by cleavage of the AHAS gene copy containing the first stacked transgene by ZFNs 34480 and 34481 (encoded on plasmid pDAB111860) or ZFNs 34482 and 34483 (encoded on plasmid pDAB111861). The donor molecule (pDAS000434; SEQ

ID NO:334, Figure 18) comprises several portions of polynucleotide sequences. The 5' end contains sequence near identical to the endogenous AHAS gene encoded in the D-genome, starting from the target ZFN cleavage site and finishing at the AHAS stop codon. Several deliberate mutations are introduced into this sequence: mutations 5 encoding the P197S mutation and codon-optimized, synonymous mutations positioned across the binding site of ZFNs 34481 and 34483 to prevent re-cleavage of the integrated donor. Following the stop codon is 316-bp of non-coding sequence corresponding to the conserved 3' untranslated region (3'UTR) in the AHAS homoeologs. The 3'UTR sequence is followed by Zinc Finger binding sites for ZFNs 10 34474 and 34475 (encoded on plasmid pDAB111857) and ZFNs 34476 and 34477 (encoded on plasmid pDAB111858). These Zinc Finger binding sites allow for self-excision of donor-derived AHAS (coding and 3'UTR) sequence integrated at an endogenous locus in the next round of transgene stacking. The self-excision Zinc 15 Finger binding sites are followed by several additional Zinc Finger binding sites (each of which is separated by 100-bp of random sequence) that flanks unique restriction endonuclease cleavage sites and which enable insertion of a transgene expression cassette (e.g. the DGT-28 expression cassette, as described in Patent Application Number 13757536). The additional Zinc Finger binding sites enable future excision of transgenes which can be integrated at an AHAS locus by sequential marker-free 20 transgene stacking, or continued sequential transgene stacking at the same genomic location using an alternate stacking method. The donor cassette is synthesized by a commercial gene service vendor (e.g., GeneArt, Life Sciences) with a short stretch of additional flanking sequence at the 5' and 3' ends to enable generation of a donor molecule with protruding 5' and 3' ends that are compatible with the ligation 25 overhangs generated by ZFNs 34474 and 34475 (encoded on plasmid pDAB111857) or ZFNs 34476 and 34477 (encoded on plasmid pDAB111858), upon cleavage of an endogenous AHAS locus. The donor molecule with protruding 5' and 3' ends is generated by digesting plasmid DNA containing the donor molecule with the restriction endonuclease *BbsI* using standard methods known to one in the art.

30

Transformation system for exogenous marker-free, sequential transgene stacking at an endogenous AHAS locus in wheat using NHEJ-directed DNA repair

[0340] Transgenic wheat events with multiple transgenes stacked at the same

5 endogenous AHAS locus are produced by exogenous marker-free, sequential transgene stacking via transformation with donor pDAS000433 and ZFNs 29732 and 29730 (encoded on plasmid pDAB109350). Precise ZFN-mediated, NHEJ-directed donor integration introduces the first transgene and S653N mutation conferring tolerance to imidazolinones at an AHAS locus, thus allowing for the regeneration of 10 correctly targeted plants using IMAZAMOX® as a selection agent, as previously described. Figure 19a depicts the integration. Subsequent transformation of wheat cells, derived from first transgene stacked events, with donor pDAS000434 and ZFNs 34480 and 34481 (encoded on plasmid pDAB111860) results in the replacement of the endogenous chromatin located between the ZFN binding sites positioned upstream 15 of P197 and at the self-excision site integrated during the first transgene stack with the donor molecule. This results in integration of the second transgene and a P197S mutation conferring tolerance to sulfonylurea, thus allowing for the regeneration of correctly targeted plants using sulfometuron methyl as a selection agent. At the same time, integration of the second donor removes the S653N mutation, thus restoring 20 susceptibility to imidazolinones (Figure 19B). One skilled in the art will appreciate that stacking of a third transgene can be achieved by transformation with appropriate zinc finger nucleases and a donor that contains an additional transgene and confers susceptibility to sulfonylurea and tolerance to imidazolinones, thus allowing the regeneration of correctly targeted plants using IMAZAMOX® as a selection agent.

25 As such, continued rounds of sequential transgene stacking are possible via transformation with donors that introduce transgenes and mutations in the endogenous AHAS genes for differential cycling between imidazolinone and sulfonylurea selection agents.

30 **Design and synthesis of donor DNA for first sequential transgene stacking at an endogenous AHAS locus using HDR-directed DNA repair**

[0341] The donor DNA for the first round of transgene stacking is designed to promote precise donor integration at an endogenous AHAS locus via ZFN-mediated

repair. The design is based on the integration of a double stranded donor molecule at the position of the double strand DNA break created by cleavage of a homoeologous copy of the endogenous AHAS gene by ZFNs 29732 and 29730 (encoded on plasmid pDAB109350). The donor molecule (pDAS000435; SEQ ID NO: 335, Figure 21) is 5 identical in sequence to pDAS000433 (SEQ ID NO:333).

Transformation system for exogenous marker-free, sequential transgene stacking at an endogenous AHAS locus in wheat using HDR-directed DNA repair

10 [0342] Transgenic wheat events with multiple transgenes stacked at the same endogenous AHAS locus are produced by exogenous marker-free, sequential transgene stacking via transformation with donor pDAS000435 and ZFNs 29732 and 29730 (encoded on plasmid pDAB109350). Precise ZFN-mediated, HDR-directed donor integration introduces the first transgene and S653N mutation conferring tolerance to imidazolinones at an AHAS locus, thus allowing for the regeneration of correctly targeted plants using IMAZAMOX® as a selection agent, as previously described. Figure 20a depicts the integration. Subsequent transformation of wheat cells, derived from first transgene stacked events, with donor pDAS000436 and ZFNs 34480 and 34481 (encoded on plasmid pDAB111860) results in the replacement of 15 the endogenous chromatin located between the ZFN binding sites positioned upstream of P197 and at the self-excision site integrated during the first transgene stack with the donor molecule. This results in integration of the second transgene and a P197S mutation conferring tolerance to sulfonylurea, thus allowing for the regeneration of correctly targeted plants using sulfometuron methyl as a selection agent. At the same 20 time, integration of the second donor removes the S653N mutation, thus restoring susceptibility to imidazolinones (Figure 20b). As will be obvious to one skilled in the art, stacking of a third transgene can be achieved by transformation with appropriate zinc finger nucleases and a donor that contains an additional transgene and confers susceptibility to sulfonylurea and tolerance to imidazolinones, thus allowing the 25 regeneration of correctly targeted plants using IMAZAMOX® as a selection agent. As such, continued rounds of sequential transgene stacking are possible via 30 transformation with donors that introduce transgenes and mutations in the endogenous

AHAS genes for differential cycling between imidazolinone and sulfonylurea selection agents.

Artificial crossing and molecular analysis to recover transgenic plants with

5 specific combinations of precise genome modifications

[0343] The *Triticum aestivum* events which are produced via transformation with donor DNA and zinc finger nuclease constructs result in the integration of donor molecule sequence at one or more copies the target endogenous locus. As shown previously, ZFN-mediated genome modification can include simultaneous editing of multiple alleles across multiple sub-genomes. Artificial crossing of transformation events can be subsequently used to select for specific combinations of precise genome modifications. For example, artificial crossing of transformation events produced that have precisely modified AHAS genes with the S653N mutation can be used to produce wheat plants that have the S653N mutation on a specific sub-genome, on multiple sub-genomes, or on all three sub-genomes. Subsequent artificial crossing of transformation events facilitates the generation of plants that have specific combinations of precise genome modifications. One skilled in the art can deploy molecular assays, such as those previously described, to track the inheritance of specific genome modification during artificial crossing in subsequent generations.

20

Example 7: Targeted integration into and disruption of *Brassica napus* omega-3 fatty acid desaturase (*Fad3*)

Selection of zinc finger binding domains specific to Fad3C and Fad3A

[0344] The transcribed regions for homoeologous *Fad3* genes were identified and characterized, zinc finger nucleases that were designed to bind and cleave these sites for NHEJ-mediated targeting of a donor sequence as described herein were designed and constructed. See, U.S. Provisional Patent Filing No. 61/697,854, herein incorporated by reference. Zinc finger proteins (ZFPs) directed against DNA sequences from homeologues of *Fad3* sequences were designed and tested as previously described in the U.S. Provisional Patent Filing No. 61/697,854. From the ZFNs showing on-target activity, two zinc finger proteins were selected that cut the *Fad3* target at high efficiency: ZFP 28051-2A-28052 recognizes SEQ ID NO:336 5'-GCCCAAGGAACCCTTTCTGGGCCATCTCGTACTCGGCCACGACTGGTAA

TTTAAT-3' and was previously shown to specifically bind and cleave the Fad3C genomic locus. Likewise Zinc finger protein 28053-2A-28054 recognizes SEQ ID NO:337 5'-

AGCGAGAGAAAGCTTATTGCAACTTCAACTACTTGCTGGTCGATCGTGTG

5 GCCACTC-3' and was previously shown to specifically bind and cleave the Fad3A and Fad3C genomic locus. Nucleotides in the target sites that are contacted by the ZFP recognition helices are shown in Table 16.

Table 16: Zinc Finger Protein Binding Sites specific to Fad3C (28051-2A-28052) or Fad3A and Fad3C (28053-2A-28054). Nucleotides in the target site that are contacted by the ZFP recognition helices are indicated in uppercase letters; non-contact nucleotides are indicated in lowercase. Nucleotides in copies of Fad3 that differ from Fad3C are identified by underlining.

28051-2A-28052	SEQ ID NO:	SEQ ID NO:356 gccaaggAACCTTTCTGGGCCAT <u>C</u> t SEQ ID NO:357 cgTACTCGGCCACGactggtaatttaat
Fad3C	338	GCCCAAGGAACCCTTTCTGGGCCATCTTCGTACTCGGC CACGACTGGTAATTAAAT
Fad3A	339	GCCCAAGGAACCCT <u>G</u> TTCTGGGC <u>T</u> ATCTTCGTACTCGGC CACGACTGGTAATTAAAT
Fad3C'	340	GCCCAAGGAACCCTTTCTGGGCCAT <u>CTTCGT</u> <u>CCTCGGC</u> CACGACTGGTAA <u>AGTTTC</u>
Fad3A'	341	GCCCAAGGAACCCTTTCTGGGCCAT <u>CTTCGT</u> <u>CCTCGGC</u> CACGACTGGTAA <u>AGTTTC</u>
Fad3A"	342	GCCCAAGGAACCCTTTCTGGGCCAT <u>CTTCGT</u> <u>TCTTGGC</u> CACGACTGGTAA <u>ATTAAA</u>
Fad3C"	343	GCCCAAGGAACCCTTTCTGGGCCAT <u>CTTCGT</u> <u>TCTTGGC</u> CACGACTGGTAA <u>ATTAAA</u>

28053-2A-28054	SEQ ID NO:	SEQ ID NO:358 agcgagagaaAGCTTAtTGCAACTTCaa SEQ ID NO:359 acTTGCTGGTCGATCGTGTggccactc
Fad3C	344	AGCGAGAGAAAGCTTATTGCAACTTCAACTACTTGCTG GTCGATCGTGTGGCCACTC
Fad3A	345	AGCGAGAGAAAGCTTATTGCAACTTCAACTACTTGCTG GTCGATCATGTTGGCCACTC
Fad3C'	346	AGCGAGAGAAAGCTTATTGCAACTTCAACTACTTGCTG GTCC <u>ATAATGTTGGCCATTC</u>
Fad3A'	347	AGCGAGAGAAAGCTTATTGCAACTTC <u>GACTACTTGCTG</u> GTCC <u>ATAATGTTGGCAATT</u> C
Fad3A"	348	AGCGAGAGGAAGCTTATTGCAACTTCAAC <u>ACAAC</u> <u>ACTTGCTG</u> GTCC <u>ATAATGTTGGCCACTC</u>
Fad3C"	349	AGCGAGAGGAAGCTTATTGCAACTTCAACTACTTGCTG GTCC <u>ATAATGTTGGCCACTC</u>

Design and construction of expression vectors encoding zinc finger nucleases specific to Fad3C and Fad3A

[0345] The Fad3 zinc finger designs were incorporated into zinc finger expression vectors encoding a protein having at least one finger with a CCHC structure (U.S. Patent Publication No. 2008/0182332). In particular, the last finger in each protein had a CCHC backbone for the recognition helix. The non-canonical, zinc finger-encoding-sequences were fused to the nuclease domain of the type IIS restriction enzyme *FokI* (amino acids 384-579 of the sequence of Wah *et al.*, (1998) Proc. Natl. Acad. Sci. USA 95:10564-10569) via a four amino acid ZC linker and a *sop2* nuclear localization signal. The self-hydrolyzing 2A encoding nucleotide sequence from *Thosea asigna* virus (Szymczak *et al.*, 2004) was added between the two Zinc Finger Nuclease fusion proteins. Expression of the ZFNs was driven by the strong constitutive promoter and 5' untranslated region (UTR) from Cassava Vein Mosaic Virus (Verdaguer *et al.*, Plant Molecular Biology 1996, 31(6); 1129-1139) and flanked by the 3' UTR (including the transcriptional terminator and polyadenylation site) from open reading frame 23 (ORF23) of *Agrobacterium tumefaciens* pTi15955 (Barker *et al.*, Plant Molecular Biology 1983, 2(6); 335-50).

[0346] The vectors were assembled using the IN-FUSION™ Advantage Technology (Clontech, Mountain View, CA). Restriction endonucleases were obtained from New England BioLabs (NEB; Ipswich, MA) and T4 DNA Ligase (Invitrogen) was used for DNA ligation. Plasmid preparations were performed using NUCLEOSPIN® Plasmid Kit (Macherey-Nagel Inc., Bethlehem, PA) or the PLASMID MIDI KIT™ (Qiagen) following the instructions of the suppliers. DNA fragments were isolated using QIAQUICK GEL EXTRACTION KIT™ (Qiagen) after agarose Tris-acetate gel electrophoresis. Colonies of assembled plasmids were initially screened by restriction digestion of miniprep DNA. Plasmid DNA of selected clones was sequenced by a commercial sequencing vendor (Eurofins MWG Operon, Huntsville, AL). Sequence data were assembled and analyzed using the SEQUENCER™ software (Gene Codes, Ann Arbor, MI). The resulting plasmid constructs: pDAB107827 (ZFN 28051-2A-28052, Figure 22, SEQ ID NO:350) and pDAB107828 (ZFN 28053-2A-28054, Figure 23, SEQ ID NO:351) were confirmed via restriction enzyme digestion and via DNA sequencing.

Design and construction of “donor” vectors for NHEJ-directed DNA repair

[0347] Two strategies of integration of DNA into Fad3 were undertaken; gene splicing, where an expression cassette was inserted into a single ZFN-induced double-stranded break and gene-editing where a portion of the gene was removed by the use of two ZFN-induced double-stranded breaks and an expression cassette was inserted to repair the gap.

[0348] For each integration method, gene splicing or gene-editing, two vectors were constructed. The first encoded a *turboGFP* (tGFP) gene expression cassette and the second encoded a gene expression cassette to confer resistance to the antibiotic hygromycin. The tGFP expression cassette consisted of the promoter, 5' untranslated region and intron from the *Arabidopsis thaliana* polyubiquitin 10 (UBQ10) gene (Norris *et al*, Plant Molecular Biology 1993, 21(5), 895-906) followed by the tGFP coding sequence (Evrogen, Moscow, Russia). The tGFP coding sequence was codon-optimized for expression in dicot plants and the 3' untranslated region (UTR) comprising the transcriptional terminator and polyadenylation site of open reading frame 23 (ORF23) of *A. tumefaciens* pTi15955 (Barker *et al*, Plant Molecular Biology 1983, 2(6), 335-50). The hygromycin resistance gene expression cassette consisted of the 19S promoter including a 5' UTR from cauliflower mosaic virus (CaMV) (Cook and Penon Plant Molecular Biology 1990 14(3), 391-405) followed by the *hygromycin phosphotransferase (hph)* gene (Kaster *et al* Nucleic Acids Research 1983 11 (19), 6895-6911). The *hph* gene been codon-optimized for expression in dicots and was flanked by a 3'UTR comprising the transcriptional terminator and polyadenylation site of Open Reading Frame 1 (ORF1) of *A. tumefaciens* pTi15955 (Barker *et al*, Plant Molecular Biology 1983, 2(6), 335-50). Both cassettes were synthesized by a commercial gene synthesis vendor (GeneArt, Life Technologies, Regensberg, Germany).

[0349] Vectors for gene splicing were constructed by cloning two tandem copies of the ZFN recognition sequence targeted by the ZFN encoded in the vector pDAB10782. Vectors for gene editing were constructed by cloning one copy of each of the ZFN recognition sequences targeted by the ZFNs encoded in the vectors pDAB107827 and pDAB107828. In both cases the two ZFN recognition sequences were separated by the recognition sequences for *Bam*HI and *Not*I restriction endonucleases. The tGFP and HPH cassettes were cloned into the *Bam*HI and *Not*I

sites of each vector resulting in four “donor” vectors: pDAS000340 (hygromycin-resistant gene-splicing donor: SEQ ID NO:352, Figure 24), pDAS000341 (tGFP reporter gene splicing donor: SEQ ID NO:353, Figure 25) , pDAS00342 (hygromycin-resistant gene-editing donor: SEQ ID NO:354, Figure 26) and 5 pDAS000343 (tGFP reporter gene editing donor: SEQ ID NO:355, Figure 27).

[0350] Colonies of the assembled plasmids were initially screened by restriction endonuclease digestion of DNA purified from overnight cultures of *E. coli*. Restriction endonucleases were obtained from New England BioLabs (NEB, Ipswich, MA) and Promega (Promega Corporation, WI). Plasmid preparations were performed 10 using the QIAPREP SPIN MINIPREP KIT™ (Qiagen, Hilden, Germany) or the PURE YIELD PLASMID MAXIPREP SYSTEM™ (Promega Corporation, WI) following the instructions of the suppliers. After the restriction fragments were confirmed by agarose gel electrophoresis of resulting fragments, plasmid DNA of selected clones were sequenced using ABI Sanger Sequencing and BIG DYE 15 TERMINATOR V3.1™ cycle sequencing protocol (Applied Biosystems, Life Technologies). Sequence data were assembled and analyzed using the Sequencher™ software (Gene Codes, Ann Arbor, MI).

Maintenance of plant material for protoplast isolation

[0351] Mesophyll derived protoplasts were isolated from three-week old 20 sterile shoot cultures of *Brassica napus* (DH10275). The corresponding seeds were germinated following the methods herein described. The seeds were surface-sterilized using 70% ethanol for 1 minute and gently shaken followed by 3-4 rinses in sterile double-distilled water. The seeds were subsequently sterilized using 20% bleach and 25 10 µl of Tween 20. The seeds were further treated with the bleach on a table top shaker at approximately 100 RPM, for 15 minutes followed by 3-4 rinses in sterile double-distilled water, seeds were carefully transferred to a sterile filter paper to remove the excess moisture and plated on seed germination medium (½ strength MS/B5 Vitamins + 1% sucrose + 0.8% Agar; pH 5.8.

[0352] Approximately, 50-60 ml of media was poured into each PETRI™ dish 30 (15 X 100 mm) and placed with a slight angle using a support). Approximately 50 seeds were placed per plate. The plates were incubated upright at 22°C in 16h/d light (20 µmol m⁻² s⁻¹) for 6 days. Hypocotyl segments of 0.5 cm size were dissected from

the six day old seedlings and cultured on shoot induction medium (MS/B5 Vitamins + 3% sucrose + 500 mg/L MES + BAP (13 μ m) + Zeatin (5 μ m) + Silver Nitrate (5mg/L) + 0.8% Agar (pH 5.8). The medium was poured in 100 x 20 mm sterile PETRITM dish; approximately 20 explants were placed per plate. Shoot meristems that appeared 5 after 3-4 weeks were transferred to shoot elongation medium (MS/B5 Vitamins + 2% sucrose + 500mg/L MES + BAP (2 μ m) + GA-3 (0.1 μ m) + 0.8% Agar (pH 5.8) and poured in 250 ml culture vessels) and the cultures were maintained in this medium for 4 weeks with one round of sub-culturing in between. Shoots of 2-3cm height were then transferred to root initiation media (1/2 strength MS/B5 Vitamins + 1% sucrose + 10 500mg/L MES + IBA (2.5 μ m) + 0.6% Agar (pH 5.8) and poured in 700 ml culture vessels) for root development. Rooted shoots were sub-cultured in fresh root initiation media at 3-4 weeks intervals as stem cuttings for two-three rounds before use. The cultures were maintained throughout at 22°C in 16h/d light (30 μ mol m⁻² s⁻¹).

15 **Isolation and purification of mesophyll protoplasts**

[0353] *In vitro* grown DH12075 *Brassica napus* plants were used as the explant source for isolating mesophyll protoplasts. To isolate the protoplasts, the 3rd to 4th upper fully expanded leaves from 3 - 4 weeks old plantlets were cut with a sharp scalpel into small strips (0.5 to 1 mm) for protoplast isolation. Enzymatic digestion 20 was carried out by treating 250-500 mg of leaf material with 25 ml of digestion buffer (1.2% (w/v) Cellulase "ONozukaTM" R10 and 0.2% (w/v) MACEROZYME[®] R10 (Source - Duchefa) dissolved in K4 media (Spangenberg et al., 1998)). The PETRITM dish containing the leaf material and digestion buffer was sealed with PARAFILMTM and incubated at room temperature for 12 to 15 h in darkness. After overnight 25 incubation the digests were filtered through a BD[®] cell strainer (mesh size 70 μ m). Protoplast suspensions (5-6 ml) collected in a 14 ml round bottomed tube was over layered with 1 ml of W5 washing buffer (154 mM NaCl, 125 mM CaCl₂, 5 mM KCl and 5 mM glucose; pH 5.8 Menzel et al. (1981)).

[0354] The protoplast suspensions were further centrifuged at 400 RPM for 10 30 min. After centrifugation, protoplasts that floated in the interphase were withdrawn and washed by centrifugation using 10 ml of W5 buffer at 400 RPM for 10 min. After the final wash, isolated protoplasts were resuspended at a density of 1X 10⁶ protoplasts per mL of W5 buffer and incubated for 1 hour before transfections.

Assessment of protoplast yield and viability

[0355] Protoplasts yield was assessed using a haemocytometer following the method of Sambrook and Russell, (2006). The cell viability was tested using 400mg/L of Evans blue stain dissolved in 0.5 M of Mannitol as described by Huang et al. (1996) with few minor modifications to the protocol.

PEG 4000 mediated DNA delivery

[0356] Before delivery to *B. napus* protoplasts, plasmid DNA of each donor and ZFN construct was prepared from cultures of *E. coli* using the PURE YIELD PLASMID MAXIPREP SYSTEM® (Promega Corporation, Madison, WI) following the instructions of the suppliers. Aliquots of donor and ZFN plasmid DNA were prepared in three molar ratios: 1:1 (30 µg of each plasmid), 5:1 (donor plasmid to ZFN plasmid to a total of 30 µg of plasmid DNA) and 10:1 (donor plasmid to ZFN plasmid to a total of 30 µg of plasmid DNA). Additionally, donor-only and ZFN-only aliquots (30 µg) were prepared as controls. The amounts of DNA delivered to the *B. napus* protoplasts via the PEG4000 mediated transformation are summarized in **Table 17.**

Table 17: Quantities of ZFN and donor DNA delivered to protoplasts

	<i>Molar Ratio of plasmid DNA</i>	<i>Total quantity of DNA (µg) delivered to 1 million protoplasts</i>
Splicing	Donor plasmid only	30
	ZFN plasmid only (pDAB107827)	30
	1:1 Donor:ZFN	60
	5:1 Donor:ZFN	30
	10:1 Donor:ZFN	30
Editing	Donor plasmid only	30
	1:1:1 ZFN plasmids (pDAB107827 and pDAB107828)	30
	1:1:1 Donor:ZFN:ZFN	90
	5:1:1 Donor:ZFN:ZFN	30
	10:1:1 Donor:ZFN:ZFN	30

[0357] Each aliquot of plasmid DNA was applied to one million protoplasts (viability \geq 95) suspended in 100µl of transformation buffer (15mM MgCl₂, 0.1% (w/v) morpholinoethanesulphonic acid (MES) and 0.5 M Mannitol; pH 5.8) followed

by 150 μ l of PEG solution (40% (w/v) PEG 4000 in 0.4 M Mannitol and 0.1 M Ca (NO₃)₂ (pH 6-7) Spangenberg and Potrykus (1995). After 10-15 min of incubation at room temperature, 5 ml of W5 buffer was added in a drop wise manner and the protoplasts were gently mixed. Another 5ml of W5 buffer was added as a slow stream to the protoplasts suspension. Protoplasts were mixed gently and centrifuged at 400 RPM for 10 min and the W5 supernatant was removed carefully leaving behind the protoplasts in the form of a pellet. Transfected protoplasts were then incubated in 1 ml of W5 buffer at room temperature until they were embedded in bead type cultures. The transfected protoplasts were embedded following the sodium alginate method as described below.

Culturing of mesophyll derived protoplasts to recover viable microcalli

[0358] Before embedding the transfected protoplasts were centrifuged at 400 RPM for 10 min and the W5 buffer was carefully removed. The protoplasts were then resuspended in 1.0 ml of 0.5 M Mannitol and incubated on ice. To this equal volume of 1.0 % sodium alginate was added and mixed gently. The protoplasts suspension was incubated in ice until it was embedded. Bead forming solution (0.4 M Mannitol + 50mM CaCl₂ (pH 5.8)) was transferred to a sterile six well plate (3-4 ml per well) using a serological pipette. Exactly 1.0 ml of the protoplasts suspension was added in a drop wise manner using a 1 ml pipette into the bead forming solution and each transfected sample (ca. 5 x 10⁵ protoplasts) was embedded per well. The protoplasts suspension was incubated for 1-2 hours at room temperature to form sodium alginate beads. After the incubation period the bead forming solution was carefully removed and replaced with 4-5ml of 1: 2 mixture of K3+H:A media (Spangenberg et al 1998) supplemented with 1.5 mg/L of Hygromycin. The protoplasts were cultured for 3-4 weeks in darkness at 22°C in a shaker (50 RPM). After 3-4 weeks the resistant microcalli (0.5-1.0mm) were released by treating with depolymerisation buffer (0.3 M Mannitol + 20mM Sodium Citrate (pH 5.8)). After removing the liquid media 3-4 ml of depolymerisation buffer was added to each well containing the bead-type cultures and incubated at room temperature for 2 hours. Using a sterile forceps the beads were gently mixed to enhance the efficient release of the microcalli. Next a sterile 1.0 ml pipette was used to gently mix gelling agent that was released in the depolymerisation buffer and subsequently removed. The microcalli was washed twice using 5 ml of

liquid A media and the microcalli was resuspended in sufficient quantity of liquid A (50 ml of liquid A was used for one ml of the settled cell volume (SCV: this was measured after transferring all the released microcalli to a sterile 50 or 15 ml falcon tube and allowed to settle down for 5 min)). After mixing the microcalli uniformly, 5 0.5 ml of the microcalli suspended in the liquid A media was transferred to B1 media (MS/MS Vitamins + 3.5 % Sucrose + 500mg/L MES + BAP (5 μ m) + NAA (5 μ m) + 2, 4-D (5 μ m) + 1.5 mg/L Hygromycin + 0.7 % Agarose Type I (pH 6.0) and poured in 100 x 20 mm sterile PETRITM dish) and using 1-2 ml of additional liquid A media the microcalli was distributed uniformly in the B1 media and the excess liquid A media 10 was carefully removed from each plate. The plates were sealed using a micropore tape which enhanced the embryo maturation. The cultures were maintained at 22°C in 16h/d light (30 μ mol m⁻² s⁻¹).

Proliferation and regeneration of shoots from mesophyll derived protoplasts

15 [0359] Hygromycin resistant colonies were picked from B1 media (microcalli derived from both SA and SP methods) after 2-3 weeks of incubation and transferred to B2 media (MS/MS Vitamins + 3.0 % Sucrose + 500mg/L MES + 500mg/L PVP + 5mg/L Silver nitrate + 5mg/L 2i P + NAA (0.5 μ m) + GA-3 (0.3 μ m) + 1.5 mg/L Hygromycin + 0.7 % Agarose Type I (pH 5.8) and poured in 100 x 20 mm sterile PETRITM dish). Approximately 25-30 calli were placed per plate and the plates were 20 sealed using PARAFILMTM and incubated at 22°C in 16h/d light (30 μ mol m⁻² s⁻¹). Hygromycin resistant colonies were subsequently recovered after 5-6 rounds of sub-culturing in B2 media at two weeks interval. The number of calli per plate was reduced to 12-15 after a third round of sub-culturing. Shoot primordias that appear 25 after 10-12 weeks were carefully recovered along with the residual calli and transferred to shoot elongation medium (MS/B5 Vitamins + 2% sucrose + 500mg/L MES + BAP (2 μ m) + GA-3 (0.1 μ m) + 300mg/L Timentin + 1.5 mg/L Hygromycin + 0.8% Agar (pH 5.8) and poured in 250 ml culture vessels). The shoots that survive 30 after 2- 3 rounds of Hygromycin selection were transferred to rooting media (1/2 strength MS/B5 Vitamins + 1% sucrose + 500mg/L MES + IBA (2.5 μ m) + 1.5 mg/L Hygromycin + 0.6% Agar (pH 5.8) and poured in 700 ml culture vessels).

Isolation of Genomic DNA from mesophyll protoplasts

[0360] Transfected protoplasts were transferred from the 3 cm PETRITM dish to a 2 mL microfuge tube. The cells were pelleted by centrifugation at 70 g and the supernatant was removed. To maximize the recovery of transfected protoplasts, the PETRITM dish was rinsed three times with 1 mL of wash buffer. Each rinse was performed by swirling the wash buffer in the PETRITM dish for 1 minute, followed by transfer of the liquid to the same 2 ml microfuge tube. At the end of each rinse, the cells were pelleted by centrifugation at 70 g and the supernatant was removed. The pelleted protoplasts were snap frozen in liquid nitrogen before freeze drying for 24 h in a LABCONCO FREEZONE 4.5[®] (Labconco, Kansas City, MO) at -40°C and 133 x 10⁻³ mBar pressure. The lyophilized cells were subjected to DNA extraction using the DNEASY[®] PLANT DNA EXTRACTION MINI KIT (Qiagen) following the manufacturer's instructions, with the exception that tissue disruption was not required and the protoplast cells were added directly to the lysis buffer.

15

Isolation of genomic DNA from callus tissue

[0361] Individual calli was snap frozen in liquid nitrogen before freeze drying for 24 h in a LABCONCO FREEZONE 4.5[®] (Labconco, Kansas City, MO) at -40°C and 133 x 10⁻³ mBar pressure. The lyophilized calli was subjected to DNA extraction using the DNEASY[®] PLANT DNA EXTRACTION MAXI kit (Qiagen, Hilden, Germany) following the manufacturer's instructions.

20

Isolation of genomic DNA from leaf tissue

[0362] Thirty (30) mg of young leaf tissue from regenerated plants was snap frozen in liquid nitrogen before freeze drying for 24 h in a LABCONCO FREEZONE 4.5[®] (Labconco, Kansas City, MO) at -40°C and 133 x 10⁻³ mBar pressure. The lyophilized calli was subjected to DNA extraction using the DNEASY[®] PLANT DNA EXTRACTION MAXI KIT (Qiagen, Hilden, Germany) following the manufacturer's instructions.

30

PCR assays of genomic DNA for NHEJ-mediated splicing and editing of Fad3C

[0363] Detection of integration of donor DNA to the Fad3C gene of *B. napus* was done by a series of PCR where at least one primer was specific to the Fad3C

locus (Table 18) and a second primer specific to either the promoter or terminator of the *gfp* cassette (Table 18 and Figure 28A). Specificity was obtained by designing oligonucleotides where the last base pair aligned to a SNP that differentiated Fad3C genomic sequence from the other copies of Fad3 genes and including a phosphorothioate internucleotide linkage before this base pair as indicated by an asterisk [*]. This design, used in combination with a polymerase having proofreading activity, directed specific amplification of each Fad3C or Fad3A allele and excluded other Fad3 copies as noted. Each primer set was empirically tested for amplification of the correct gene copies through Sanger-based sequencing of the PCR amplification products obtained from wild type *B. napus*.

Table 18: Oligonucleotide sequences used to detect integration of DNA into ZFN-induced double-stranded breaks.

	<i>Primer Name</i>	<i>Primer Sequence</i>	<i>SEQ ID NO:</i>	<i>Specificity</i>
1	FAD3CNHEJ-L4-F2	gattcctaaggcattgtgggt*c	360	Fad3C only
2	FAD3CNHEJ-L4-R2	gaaaatctcatatcgAACgtgcg*t	361	Fad3C only
3	FAD3CNHEJ-L6-F1	cgcttaccctcttatctggta*a	362	Does not amplify Fad3C' or Fad3C"
4	FAD3CNHEJ-L6-R2	ccttgcctctgtaccaaggca*g	363	Fad3C only
5	19SPNHEJ-R2	gtgtgtggaaatcttatcttcgg	364	n/a
6	AtORF1NHEJ-F1	caagtcaggattatagtccaaagca	365	n/a
7	AtUbiNHEJ-R1	caagaatatcctgatccgttgac	366	n/a
8	AtORF23tNHEJ-F1	tggcagttgaaatactcaaacc	367	n/a
9	FAD3aCNHEJ-L4-F1	gtccttgagatccatgagcta*t	368	Fad3A only
10	FAD3aCNHEJ-L4-F2	gattcctaaggcattgtgggt*a	369	Fad3A only
11	FAD3aNHEJ-L4-R1	tgcgttcaagaaatcaaagac*a	370	Fad3A only
12	FAD3aNHEJ-L4-R2	gaaaatctcatatcgAACgtgcg*g	371	Fad3A only
13	FAD3aNHEJ-L6-F1	tctggtaaatcctaattcct*c	372	Fad3A only
14	FAD3aNHEJ-L6-R2	ccttgcctctgtaccaaggca*a	373	Fad3A only
15	FAD3aNHEJ-L6-R1	cttgcctctgtaccaaggcaactt*c	374	Excludes Fad3C

5 * Indicates phosphorothioate internucleotide linkages to direct specific amplification (with proofreading polymerase) of Fad3C or Fad3A to exclusion of other copies of Fad3 as noted. Each primer set was empirically tested for amplification of the correct gene copies by Sanger-based sequencing of the PCR amplification products obtained from wild type *B. napus*.

Detection of gene addition to *Fad3C* by non-homologous end joining in protoplasts

[0364] Genomic DNA was extracted from protoplast pools (one million

10 protoplast per pool) to which donor DNA encoding a functional tGFP reporter cassette (pDAS000341 or pDAS000343), ZFN DNA (pDAB107827 or pDAB107828) or a mixture of donor and ZFN DNA had been delivered twenty-four hours earlier. Quantities of DNA delivered for transformation are described above.

15 PCR products were cloned into plasmid vectors. The genomic editing occurs independently in each cell giving rise to a variety of different insertion events, by cloning into a plasmid vector, each genomic edit can be sequenced without ambiguity. Several clones were sequenced on an ABI3730XL® automated capillary electrophoresis platform. Analysis of gene sequences was done using SEQUENCHER SOFTWARE V5.0™ (GeneCodes, Ann Arbor, MI).

20 [0365] Evidence of gene addition to Fad3C locus by editing or splicing was provided by amplification of both the 5' and 3' Fad3C-cassette junctions from genomic DNA extracted from protoplasts using the primers described in Table 18. Products of PCR amplification with primers "FAD3CNHEJ-L4-F2" and

25 "AtUbiNHEJ-R1" was completed to amplify the 5' junction of tGFP cassette and Fad3C. PCR amplification with primers "FAD3CNHEJ-L4-R2" and "AtORF23tNHEJ-F1" was completed to amplify the 3' junction of tGFP cassette and Fad3C. PCR amplification with primers "FAD3CNHEJ-L4-F2" and "FAD3CNHEJ-L4-R2" was completed to amplify across the double strand breaks induced by ZFN 28051-2A-28052. No amplification was observed from protoplasts to which ZFN

30 plasmid or donor plasmid alone had been delivered. All junction sequences were indicative of insertion of the tGFP cassette at the Fad3C locus via an NHEJ-mediated repair pathway. Deletions of varying lengths from either or both the genome and the cassette were observed as well as the addition of sequences derived from the vector backbones (either from the donor or ZFN) being inserted between the genome and the cassette.

Detection of gene addition to *Fad3C* by non-homologous end joining in callus tissue regenerated from protoplasts

[0366] Further evidence of splicing and editing of the *Fad3C* locus was obtained from callus tissue regenerated from protoplasts on selection (1.5 mg/L hygromycin, as described above) to which donor DNA encoding an *hph* cassette (pDAS000340 or pDAS000342), ZFN DNA only (pDAB107827 or pDAB107828) or donor and ZFN DNA had been delivered (quantities of DNA delivered are given in **Table 17**). DNA was extracted from approximately 80 calli for each ratio, except editing 1:1:1, for which no calli survived, four weeks after protoplast transfection.

[0367] Integration of the *hph* cassette into the *B. napus* genome (fwat *Fad3C* or randomly) was confirmed by TAQMANTM qPCR using primers (SEQ ID NO:402; F - 5'CTTACATGCTTAGGATCGGACTTG 3', SEQ ID NO:403; R - 5'AGTCCAGCACCAGATCTAACG 3') and probe (SEQ ID NO:404; 5' CCCTGAGCCCAAGCAGCATCATCG 3') specific to the *hph* gene. These primer-probe pairs were used in a duplex reaction with primers (SEQ ID NO:405; F - 5' CGGAGAGGGCGTGGAGG 3', SEQ ID NO:406; R - 5' TTCGATTGCTACAGCGTCAAC 3') and probe (SEQ ID NO:407; 5' AGGCACCATCGCAGGCTTCGCT 3') specific to the *B. napus* high mobility group protein I/I (HMG I/Y), which is present as a single copy on the A genome (Weng *et al.*, 2004, Plant Molecular Biology Reporter). Amplification was performed on a C1000 thermal cycler with the CFX96 or CF384 REAL-TIME PCR DETECTION SYSTEMTM (BioRad, Hercules, CA). Results were analyzed using the CFX MANAGERTM (BioRad) software package. Relative quantification was calculated according to the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen, 2001), which provided an estimation of the number of copies of *hph* cassette inserted into the genome. Evidence of NHEJ-mediated splicing and editing of *Fad3C* was obtained by conducting PCR assays with one primer specific to *Fad3C* and a second primer specific to either the promoter or terminator of the *hph* cassette (Table 17 and Figure 28B). Due to limited quantities of DNA obtained from callus tissue, only integration in the sense orientation was assayed. PCR products were gel-purified using QIAQUICK MINIELUTE PCR PURIFICATION KITTM (Qiagen) and sequenced using a direct Sanger sequencing method. The sequencing products were purified with ethanol,

sodium acetate and EDTA following the BIGDYE® v3.1 protocol (Applied Biosystems) and sequenced and analyzed as above.

[0368] The numbers of calli containing the donor cassette in each experiment are given in Table 18. Evidence of donor gene addition to the Fad3C locus by editing and/or splicing was provided by PCR amplification (with primers shown in Table 19) across the ZFN cut sites and both the 5' and 3' Fad3C-*hph* cassette junctions. PCR amplification of the genomic DNA isolated from callus tissue recovered from control protoplasts which were transformed with only the *hph* plasmid (pDAS000340 and pDAS000342) or only the ZFN plasmid (pDAB107827 and pDAB107828) did not result in the production of PCR amplification products.

[0369] The PCR amplicons produced from the amplification of the 5' and 3' Fad3C-*hph* cassette junctions were purified from the agarose gel and sequenced to confirm specificity of the integration within the Fad3C genomic locus. The results of the sequencing analysis of the PCR products indicated that each isolated callus which was generated from an individually transformed protoplast only produced a single PCR amplification product and did not contain cells of mixed genotypes.

[0370] In NHEJ-mediated integration of donor sequences within the Fad3C genomic locus experiments the frequency of addition to the target locus (as defined by any part of the donor DNA vector being amplified from the target locus) was 42%, 46% and 32% for the DNA concentrations of 1:1, 5:1, and 10:1 (Donor DNA: ZFN DNA), respectively. See, Table 20. The frequency of on-target splicing was determined by assaying whether both cassette junctions were amplifiable and from the sequencing of the PCR products. These results verified that the cassette was inserted at the target locus in the correct orientation. The frequency of integration was calculated as 4%, 3% and 3% for the 1:1, 5:1 and 10:1 of Donor plasmid DNA: ZFN plasmid DNA concentrations, respectively. In gene editing experiments the frequency of addition to the target locus defined by any part of the donor DNA vector being amplified from the target locus, was 66% and 65% for the 5:1:1 and 10:1:1 of Donor plasmid DNA: ZFN plasmid DNA concentrations, respectively. See, Table 21. The frequency of on-target editing, was determined by both cassette junctions being amplifiable and producing a sequence of PCR products. These results verified that the cassette was inserted at the target locus in the correct orientation at frequencies of 3% and 6% for the 5:1:1 and 10:1:1 of Donor plasmid DNA: ZFN plasmid DNA

concentrations, respectively. As observed in the protoplast assays, the base pairs were either deleted or additional bases were inserted between the genome and the cassette as a result of the cleavage of the genomic locus by the ZFN (Figures 30-31).

[0371] In certain instances the PCR products resulted in an addition of 5 nucleotide sequences within the target locus, no PCR product, or a larger PCR product than observed in wild-type samples. These results which were produced from the PCR amplification using primers flanking the cut site indicated that the locus had been disrupted in both pairs of chromosomes (Figures 30-31). In some of the instances more than one band was amplified at the splice junctions (Figures 30-31) indicating that 10 different insertions had occurred independently in each copy of the genome.

Table 19: Number of calli positive for presence of *hph* after four weeks on selection

Vectors delivered	Molar Ratio of Donor DNA: ZFN DNA	Number of calli sampled	Number of calli positive for <i>hph</i> after four weeks on selection
pDAS000340 DAB107827	1:1	88	76
	5:1	88	35
	10:1	87	37
pDAS000342 DAB107827 DAB107828	1:1:1	-	-
	5:1:1	80	38
	10:1:1	79	52

Table 20: Number of calli with *hph* inserted by splicing at FadC locus at the DSB induced by ZFN28051-2A-28052

Vectors delivered	Molar Ratio of Donor DNA: ZFN DNA	Number of calli positive for <i>hph</i> after four weeks on selection	Number of calli from which at least one splicing border amplified	Number of calli from which at least one perfect* border amplified	Number of calli from which both splicing borders amplified
pDAS000340 + DAB107827	1:1	76	32	0	3
	5:1	35	16	0	1
	10:1	37	12	0	1

* no base pairs deleted or additional base pairs inserted at cut site

Table 21: Number of calli with *hph* inserted by editing at FadC locus at the cut sites induced by ZFN28051-2A-28052 and ZFN28053-2A-28054

Vectors delivered	Molar Ratio of Donor DNA: ZFN DNA	Number of calli positive for <i>hph</i> after four weeks on selection	Number of calli from which at least one splicing border amplified	Number calli from which at least one perfect* border amplified	Number of calli from which both editing borders amplified
pDAS000342 + DAB107827 + DAB107828	5:1:1	38	25	2	1
	10:1:1	52	34	2	3

* no base pairs deleted or additional base pairs inserted at cut site

Detection of gene addition to *Fad3C* by non-homologous end joining in plants

5 [0372] DNA was extracted from plants that were regenerated from protoplasts and transferred to potting medium (as described above). The majority of plants recovered were estimated to contain only 1-2 copies of the *hph* cassette encoded in the donor DNA. Plants were analyzed with the same suite of assays described for callus tissue as well as with assays to determine if the cassette had inserted in an antisense 10 orientation or donor integration at the Fad3A locus.

Table 22: Estimated copy number of plants regenerated from protoplasts. For each ratio three transfections of one million protoplasts were performed.

Vectors delivered	Molar Ratio of Donor DNA: ZFN DNA	# plants with 1-2 copies <i>hph</i>	# plants with 3-4 copies <i>hph</i>	# plants with 5 or more copies <i>hph</i>
pDAS000340 DAB107827	1:1	28	10	22
	5:1	14	6	17
	10:1	24	12	10
pDAS000342 DAB107827 DAB107828	1:1:1	1	0	0
	5:1:1	22	7	10
	10:1:1	24	13	26
Total	---	109	48	85

15 [0373] The frequency of on-target splicing, where the *hph* cassette was inserted into Fad3C in either direction, was 51%, 32% and 56% for Donor DNA: ZFN

DNA at concentrations of 1:1, 5:1 and 10:1, respectively (Table 23). Of these results, 35% 32% and 50% (1:1, 5:1 and 10:1) were inserted in the forward orientation (Table 23).

[0374] The frequency of on-target editing, where the *hph* cassette was inserted into Fad3C in either direction, replacing the area from locus 4 to locus 6, was 2% and 0% for Donor DNA: ZFN DNA at concentrations of 5:1:1 and 10:1:1, respectively (Table 24). In addition, when both ZFNs were delivered at 5:1:1, 2% and spliced into locus 4 and 10% spliced into locus 6 and when both ZFNs were delivered at 10:1:1 10% and spliced into locus 4 and 15% spliced into locus 6.

[0375] The bands obtained can be sequenced to determine the number of perfect borders. Additionally, plants can be screened for off-target insertions to determine the frequency of integration of *hph* at sites other than Fad3, and the frequency of integration at Fad3A rather than Fad3C.

Table 23: Number of plants with *hph* inserted by splicing at FadC locus at the DSB induced by ZFN28051-2A-28052

Vectors delivered	Molar Ratio of Donor DNA: ZFN DNA	Number of plants analysed (positive for <i>hph</i>)	Number of plants from which at least one splicing border amplified (forward/reverse/either)	Number of plants from which both splicing borders amplified (forward/reverse/either)
pDAS000340 + DAB107827	1:1	60	21/23/31	4/7/8
	5:1	37	12/4/12	3/1/3
	10:1	46	23/12/26	4/4/7

* no base pairs deleted or additional base pairs inserted at cut site

Table 24: Number of plants with *hph* inserted by editing at FadC locus at the cut sites induced by ZFN28051-2A-28052 and ZFN28053-2A-28054

Vectors delivered	Molar Ratio of Donor DNA: ZFN DNA: ZFN DNA	Number of plants analysed (positive for <i>hph</i>)	Number of plants from which at least one splicing border amplified (forward/reverse/either)	Number of plants from which both editing borders amplified (forward/reverse/either)
pDAS000342 + DAB107827 +	5:1:1	39	17/11/24	0/1/1
	10:1:1	63	27/27/34	0/0/0

DAB107828					
-----------	--	--	--	--	--

* no base pairs deleted or additional base pairs inserted at cut site

Example 8: Targeted integration into and disruption of Corn Event DAS-59132

Characterization of an Endogenous Genomic Locus for Gene Targeting

[0376] The genomic locus of Corn Event DAS-59132 was described in

5 International Patent Application No. WO 2009/100188 A2. Corn Event DAS-59132 comprises the Cry34Abl, Cry35Abl, and PAT transgene expression cassettes. These transgene expression cassettes were integrated into chromosome 8 of the B73 maize genome derived region of Hi-II maize germplasm (D. D. Songstad, W. L. Petersen, C. L. Armstrong, *American Journal of Botany*, Vol. 79, pp. 761-764, 1992) as a full length T-strand insert. In addition, the genomic DNA surrounding the transgenic locus lacked any large deletions relative to the native B73 sequence, and was generally devoid of repetitive elements except for a single, small repetitive element.

[0377] The genomic locus in which Corn Event DAS-59132 integrated was

10 selected as an endogenous genomic locus for gene targeting. The selection of this endogenous genomic locus was based on the characterization of Corn Event DAS-59132. This event resulted from the integration of a T-strand into the endogenous genomic locus, and the subsequent expression of three transgene expression cassettes. In addition, there was minimal alteration of normal growth and development of corn plants which comprise Corn Event DAS-59132. The event retained the agronomic and breeding characteristics and was comparable in agronomic performance to non-transformed control plants.

[0378] An embodiment of the disclosure includes polynucleotide sequences

that can be targeted for the integration of a transgene. The full length DNA molecule (PHI17662A) used to transform Corn Event DAS-59132, the 3' end of the genomic flanking sequence, and the PHI17662A / 3' maize genome junction were described in the disclosure of International Patent Application No. WO 2009/100188 A2, and are disclosed in this filing as SEQ ID NO:427, SEQ ID NO:428 and SEQ ID NO:429, respectively. The 5' end of the genomic flanking sequence, and the genomic locus where Corn Event DAS-59132 integrated into the corn genome is disclosed in this filing as SEQ ID NO:430 and SEQ ID NO:431, respectively. The genomic locus listed as SEQ ID NO:431 was used to design zinc finger proteins for gene targeting.

Production of Zinc Finger Proteins Designed to bind the Genomic Locus for Corn Event DAS-59132

[0379] Zinc finger proteins directed against DNA sequences which comprise the genomic locus for Corn Event DAS-59132 (*see, Figure 32*) were designed as previously described. *See, e.g., Urnov et al. (2005) Nature 435:646-651.* Exemplary target sequence and recognition helices are shown in Tables 25A (recognition helix regions designs) and Table 25B (target sites). In Table 25B, nucleotides in the target site that are contacted by the ZFP recognition helices are indicated in uppercase letters.

10 **Table 25A: Genomic locus for Corn Event DAS-59132-binding zinc finger designs**

ZFP#	F1	F2	F3	F4	F5
25716	RSDDLSK SEQ ID NO:533	QSGSLTR SEQ ID NO:534	RSDNLRE SEQ ID NO:535	QSGDLTR SEQ ID NO:158	DTGARLK SEQ ID NO:536
25717	RSADRKT SEQ ID NO:537	DRSHLSR SEQ ID NO:538	TSGNLTR SEQ ID NO:143	RSDDLSSR SEQ ID NO:539	QSANRTK SEQ ID NO:540

Table 25B: Target Sequences for zinc finger proteins

Zinc Finger Number	SEQ ID NO:	Target Sequence
25686	432	caCAACAAGACtGCGGGTtcggggcgc
25687	433	gaTAGGTGGCAGTGGCAgtggcactggc
25688	434	taTCGGCACACAACAAGACtgcgggttcgg
25689	435	tgGCAGTGGCAGTGGCActggcacggca
25692	436	caGCAGATGAGcGGAGCGatcgatcg
25693	437	caGTGGATAGAGCAGCGtggccgttgg
25710	438	agGAAGCCGGCGGTGAAtgtcgccgt
25711	439	cgTCGCCAcTCGGCACAAggctcatcag
25712	440	atCGGGCATCGGCGACTgatgagcctt
25713	441	gaTCAACGGAAGCGGATGGCccgcttct
25716	442	tgATCGCAtCGGGCATCGgcactgatg
25717	443	cgGAAGCGGATGGCCCCGtttttagaa

15 [0380] The Corn Event DAS-59132 zinc finger designs were incorporated into vectors encoding a protein having at least one finger with a CCHC structure. *See,*

U.S. Patent Publication No. 2008/0182332. In particular, the last finger in each protein had a CCHC backbone for the recognition helix. The non-canonical zinc finger-encoding sequences were fused to the nuclease domain of the type IIS restriction enzyme FokI (amino acids 384-579 of the sequence of Wah *et al.* (1998)

5 Proc. Natl. Acad. Sci. USA 95:10564-10569) via a four amino acid ZC linker and an *opaque-2* nuclear localization signal derived from *Zea mays* to form Corn Event DAS-59132 zinc-finger nucleases (ZFNs). Expression of the fusion proteins in a bicistronic expression construct utilizing a 2A ribosomal stuttering signal as described in Shukla *et al.* (2009) *Nature* 459:437-441 was driven by a relatively strong,

10 constitutive and ectopic promoter such as the CsVMV promoter.

[0381] The optimal zinc fingers were verified for cleavage activity using a budding yeast based system previously shown to identify active nucleases. *See, e.g.*, U.S. Patent Publication No. 20090111119; Doyon *et al.* (2008) *Nat Biotechnol.* 26:702-708; Geurts *et al.* (2009) *Science* 325:433. Zinc fingers for the various functional domains were selected for *in-vivo* use. Of the numerous ZFNs that were designed, produced and tested to bind to the putative Corn Event DAS-59132 genomic polynucleotide target sites, four pairs of ZFNs were identified as having *in vivo* activity at high levels, and selected for further experimentation. *See, Table 25A.* These ZFNs were characterized as being capable of efficiently binding and cleaving 20 the four unique Corn Event DAS-59132 genomic polynucleotide target sites *in planta*.

[0382] Figure 1 shows the genomic organization of the Corn Event DAS-59132 locus in relation to the ZFN polynucleotide binding/target sites of the four ZFN pairs. The first three ZFN pairs (E32 ZFN1, E32 ZFN2, and E32 ZFN3) bind upstream of the Corn Event DAS-59132 T-strand insert, the second three ZFN pairs (E32 ZFN4, E32 ZFN5, and E32 ZFN6) bind downstream of the Corn Event DAS-59132 T-strand insert. After testing the ZFN pairs in the budding yeast assay, ZFN pairs which optimally bound the Corn Event DAS-59132 locus were advanced for testing in a transient corn transformation assay.

30 **Zinc finger nuclease constructs for expression in maize**

[0383] Plasmid vectors containing ZFN expression constructs of the four exemplary zinc finger nucleases, which were identified using the yeast assay and described in Example 2, were designed and completed using skills and techniques

commonly known in the art. Each zinc finger-encoding sequence was fused to a sequence encoding an opaque-2 nuclear localization signal (Maddaloni *et al.* (1989) *Nuc. Acids Res.* 17(18):7532), that was positioned upstream of the zinc finger nuclease.

5 [0384] Next, the opaque-2 nuclear localization signal::zinc finger nuclease fusion sequence was paired with the complementary opaque-2 nuclear localization signal::zinc finger nuclease fusion sequence. As such, each construct consisted of a single open reading frame comprised of two opaque-2 nuclear localization signal::zinc finger nuclease fusion sequences separated by the 2A sequence from *Thosea asigna* virus (Mattion *et al.* (1996) *J. Virol.* 70:8124-8127). Expression of the ZFN coding sequence was driven by the highly expressing constitutive *Zea mays* Ubiquitin 1 Promoter (Christensen *et al.* (1992) *Plant Mol. Biol.* 18(4):675-89) and flanked by the *Zea mays* Per 5 3' polyA untranslated region (US PAT NO. 6,699,984). The resulting four plasmid constructs were confirmed via restriction enzyme digestion and 10 via DNA sequencing. Figures 33 and 34 provide a graphical representation of the completed plasmid construct. The ZFN expressed in plasmid construct, pDAB105906 (Figure 33), contains "Fok-Mono" which is a wildtype FokI endonuclease. The ZFN expressed in plasmid construct, pDAB111809 (Figure 34), contains "Fok1-ELD" which is a modified FokI endonuclease. The modified Fok1 endonuclease contains 15 alterations as described in Doyon Y., Vo T., Mendel M., Greenberg S., Wang J., Xia D., Miller J., Urmov F., Gregory P., and Holmes M. (2010) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architecture. *Nature Methods*, 8(1); 74-79.

20 [0385] A donor construct was designed to integrate into the ZFN cleaved genomic DNA of the Corn Event DAS-59132 genomic locus. Figure 35 illustrates the donor construct, pDAB100655, which consists of a single gene expression cassette. This single gene expression cassette is driven by the *Zea mays* Ubiquitin 1 promoter (Zm Ubi1 promoter) :: the *aad-1* coding sequence (AAD1; US Patent No. 7,838,733) :: and is terminated by the *Zea mays* Per 5 3' untranslated region (ZmPer5 3'UTR). 25 The construct contains a pair of repeated E32 ZFN6 binding sequences which were included downstream of the *aad-1* gene expression cassette. The various gene elements were assembled in a high copy number pUC based plasmid.

Transient transformation of maize to determine ZFN efficiency

[0386] Maize Hi-II embryogenic cultures were produced as described in U.S. Pat. No. 7,179,902, and were used to evaluate and test the efficiencies of the different ZFNs. Plasmid DNA consisting of pDAB105901, pDAB105902, pDAB105903,

5 pDAB105904, pDAB105905 and pDAB105906 were transiently transformed into maize callus cells to compare the cutting frequency of different ZFNs against a standard tested ZFN, pDAB7430, which was designed to the inositol polyphosphate 2-kinase gene locus within the maize genome as described in US Patent Application No. 2011/0119786.

10 [0387] From the cultures, 12 mL of packed cell volume (PCV) from a previously cryo-preserved cell line plus 28 mL of conditioned medium was subcultured into 80 mL of GN6 liquid medium (N6 medium (Chu *et al.*, (1975) *Sci Sin.* 18:659–668), 2.0 mg/L 2, 4-D, 30 g/L sucrose, pH 6.0) in a 500 mL Erlenmeyer flask, and placed on a shaker at 125 rpm at 28°C. This step was repeated two times
15 using the same cell line, such that a total of 36 mL PCV was distributed across three flasks. After 24 hours, the contents were poured into sterile a PETRITM dish and the GN6 liquid media was removed. Slightly moistened callus was transferred to a 2.5 cm diameter circle on GN6 S/M solid medium (N6 Medium (Chu *et al.*, (1975) *Sci Sin.* 18:659–668), 2.0 mg/L 2,4-D, 30 g/L sucrose, 45.5 g/L sorbitol, 45.5 g/L
20 mannitol, 100 mg/L myo-inositol, 2.5 g/L Gelrite, pH 6.0) containing filter paper. The plates were incubated in the dark for 4 hours at 28° C.

[0388] Microparticle gold (0.6 micron, BioRad, Hercules, CA,) was prepared for DNA precipitation by weighing out 21 mg into a sterile, siliconized 1.7 mL microcentrifuge tube (Sigma-Aldrich, St. Louis, MO) and 350 µL of ice cold 100%
25 ethanol was added and vortexed for 1 minute. The gold was pelleted by centrifugation at 10,000 rpm for 15 seconds using a MINISPINTM centrifuge (Eppendorf, Hauppauge, NY). After removing the supernatant, 350 µL of ice cold, sterile water was added, mixed up and down with the pipette and centrifuged at 10,000 rpm for 15 seconds. The wash step was repeated one more time prior to
30 suspending the gold in 350 µL of ice cold, sterile water. The washed gold was then stored at -20°C until needed.

[0389] For each DNA precipitation, 3 mg of gold in 50 µL of water was aliquoted into a siliconized 1.7 mL microcentrifuge tube (Sigma-Aldrich, St. Louis,

MO). Plasmid DNA (2.5 µg E32 ZFN in plasmids pDAB105901, pDAB105902, pDAB105903, pDAB105904, pDAB105905 or pDAB105906 and 2.5 µg IPPK2 ZFN in plasmid pDAB7430) was premixed in 0.6 mL microcentrifuge tubes (Fisher Scientific, Nazareth, PA) and added to the gold suspension gently pipetting up and down 5-10 times to mix thoroughly. Twenty microliters (20 µL) of cold 0.1 M spermidine was then added and gently mixed by pipetting up and down 5-10 times. Fifty microliters (50 µL) of ice cold 2.5 M calcium chloride was added slowly and gently mixed by pipetting up and down 5-10 times. The tube was then capped and allowed to incubate at room temperature for 10 minutes. After centrifuging for 15 seconds at 10,000 rpm, the supernatant was carefully removed and 60 µL of ice cold, 100% ethanol was added. The gold DNA mixture was resuspended by gently pipetting up and down 5-10 times.

[0390] For microparticle bombardment, sterilized macrocarriers (BioRad, Hercules, CA) were fit into stainless steel holders (BioRad, Hercules, CA) and autoclaved. Nine microliters (9 µL) of gold/DNA suspension was evenly spread in the center of the macrocarrier being sure to pipette up and down so as to keep suspension well mixed between aliquots. Macrocarriers were then placed onto a piece of sterile 125 mm Whatman #4 filter paper (GE Healthcare, Buckinghamshire, UK) on a bed of 8-mesh DRIERITE™ (W.A Hammond Drierite Co., Xenia, OH) in a 140 x 25 mm glass PETRI™ dish. The gold/DNA was allowed to dry completely for about 5-10 minutes. Rupture discs (1100 psi, BioRad, Hercules, CA) were sterilized by soaking for a few seconds in isopropyl alcohol then loaded into the retaining cap of a microparticle bombardment devise (PDS-1000, BioRad, Hercules, CA). An autoclaved stopping screen (BioRad, Hercules, CA) and a loaded macrocarrier was placed into the launch assembly, the lid was screwed on and slide into the bombardment chamber just under the nozzle. The PETRI™ dish containing target was uncovered and placed in the bombardment chamber 6 cm below the nozzle. A vacuum was pulled (-0.9 bar) and the devise was fired. Steps were repeated for each target blasted. Targets were incubated in dark at a temperature of 28 °C for 24 hours on the same blasting medium. Blasted cells were transferred to recovery GN6 solid recovery medium (N6 medium (Chu *et al.*, (1975) Sci Sin. 18:659–668), 2.0 mg/L 2, 4-D, 30 g/L sucrose, 2.5 g/L Gelrite, pH 6.0) and incubated for additional 48 hours @ 28°C in the dark. Seventy-two hours post bombardment, the cells were harvested into

2 mL EPPENDORF MICROFUGE SAFE LOCK TUBES™ and lyophilized for 48 hours in a VIRTIS MODEL # 50L VIRTUAL XL-70 LYOPHILIZER™ (SP Scientific, Gardiner NY).

5 **Next generation sequencing (NGS) analysis of transiently transformed maize**

[0391] The transiently transformed maize callus tissue was analyzed to determine the cleavage efficiency of the zinc finger nuclease proteins.

Sample Preparation:

[0392] Maize callus tissue transiently transformed with the ZFN constructs and two control vectors, pDAB100664 and pDAB100665, were collected in 2 mL EPPENDORF™ tubes and lyophilized for 48 hrs. Genomic DNA (gDNA) was extracted from lyophilized tissue using the QIAGEN PLANT DNA EXTRACTION KIT™ (Valencia, CA) according to manufacturer's specifications. The isolated gDNA was resuspended in 200 µl of water and the concentration was determined using a NANODROP® spectrophotometer (Invitrogen, Carlsbad, CA). Integrity of the DNA was estimated by running all samples on a 0.8% agarose E-gels (Invitrogen). All gDNA samples were normalized (25 ng/µl) for PCR amplification to generate amplicons which would be analyzed via ILLUMINA™ sequencing (San Diego, CA).

[0393] PCR primers for amplification of the genomic regions which span each tested ZFN cleavage site and the control samples were purchased from Integrated DNA Technologies (Coralville, IA). Optimum amplification conditions for the primers were identified by gradient PCR using 0.2 µM appropriate primers, ACCUPRIME PFX SUPERMIX™ (1.1X, Invitrogen) and 100 ng of template genomic DNA in a 23.5 µL reaction. Cycling parameters were initial denaturation at 95 °C (5 min) followed by 35 cycles of denaturation (95°C, 15 sec), annealing (55-72 °C, 30 sec), extension (68°C, 1 min) and a final extension (72°C, 7 min).

20 Amplification products were analyzed on 3.5% TAE agarose gels. After identifying an optimum annealing temperature, preparative PCR reactions were carried out to validate each set of PCR primers and for generating the ILLUMINA™ sequencing amplicon.

[0394] For preparative PCR, 8-individual small scale PCR reactions were performed for each template using conditions described above and the resulting PCR products were pooled together and gel purified on 3.5% agarose gels using the

5 QIAGEN MINELUTE GEL EXTRACTION/PURIFICATION KIT™ per manufacturer's recommendations. Concentrations of the gel purified amplicons were determined by NANODROP™ and the ILLUMINA™ sequencing samples were prepared by pooling approximately 100 ng of PCR amplicons from ZFN targeted and corresponding wild type controls. Primers used for the PCR amplicon generation are shown in Table 26 below.

Table 26: Oligonucleotides for amplification of ZFN binding sites

Corn Event DAS-59132 Zinc Finger Number	Direction // SEQ ID NO:	Primer Sequence
25686/25687 and 25688/25689	Forward // SEQ ID NO:444	5'-CAGGCAGCGCCACCGAAC-3'
	Reverse // SEQ ID NO:445	5'-CGATCGATCGCGTGCCTG-3'
256892/256893	Forward // SEQ ID NO:446	5'-CTGGCACGGCACGCGATC-3'
	Reverse // SEQ ID NO:447	5'-CGGAGATCCGGCCCCAAC-3'
25710/25711	Forward // SEQ ID NO:448	5'-GACACGGCACACACGGCG-3'
	Reverse // SEQ ID NO:449	5'-TCGGGCATCGCGACTGA-3'
25712/25713 and 25716/25717	Forward // SEQ ID NO:450	5'-ACTCGGCACAAGGCTCAT-3'
	Reverse // SEQ ID NO:451	5'-CCTGTGCCAATTCTAAAG-3'
9149/9215	Forward // SEQ ID NO:452	5'-GCAGTGCATGTTATGAGC-3'
	Reverse // SEQ ID NO:453	5'-CAGGACATAAATGAAGTGAA TC-3'

ILLUMINA™ Sequencing and Analysis:

10 [0395] The ZFNs were designed to recognize, bind and modify specific DNA sequences within the genomic locus of transgenic Corn Event DAS-59132. The efficiency by which the four ZFNs cleaved the genomic locus was assayed to determine which ZFN cleaved most efficiently. ILLUMINA™ sequencing was performed at Cofactor Genomics (St. Louis, MO) and sequences were analyzed using a sequence analysis script. Low quality sequences were filtered out and the remaining sequences were parsed according to unique DNA sequences identifiers. The unique DNA sequences identifiers were then aligned with the reference sequence and scored for insertions/deletions (Indels). To determine the level of cleavage

15

activity, the region surrounding the ZFN cleavage site was scored for the presence of sequence variants which resulted from the INDELS. Cleavage activity for each ZFN in the study was calculated as the number of sequences with indels/1M high quality sequences or as a percentage of high quality sequences with indels. Next, the levels 5 of cleavage efficiency were determined by normalizing the ZFN level of cleavage activity with the activity of a ZFN directed to the IPP2-K gene as described in U.S. Patent Publication No. 2011/0119786. Figure 36 and Table 27 present the cleavage efficiency of the tested ZFNs.

10 [0396] Event 32 ZFN6 which contains the 25716 and 25717 zinc finger binding domains cleaved the genomic locus of transgenic Corn Event DAS-59132 with the highest efficiency. This ZFN functioned at 380 times the efficiency of the control IPPK2 zinc finger nuclease. Given the surprisingly high levels of cleavage activity of Event 32 ZFN6, this ZFN was selected for advancement to test the integration of a donor DNA fragment into a genomic locus via non homologous end-15 joining.

Table 27: Cleavage efficiency of the tested eZFNs

E32 ZFN Number	% IPPK2 ZFN Activity
25686/25687 shown as “686/687”	32
25688/25689 shown as “688/689”	108
25712/25713 shown as “712/713”	69
25716/25717 shown as “716/717”	380

Transformation of ZFN in protoplast

20 [0397] A system for gene targeting was established to target the endogenous genomic loci of Corn Event DAS-59132 and to optimize donor targeting parameters in maize. Double strand breaks were generated within the genome at Corn Event DAS-59132 and repaired by either the non-homologous end joining (NHEJ) or homology dependent repair (HDR).

25 Protoplast Isolation:

[0398] Maize Hi-II embryogenic suspension cultures were obtained and were maintained on a 3.5 day maintenance schedule. In a 50 mL sterile conical tube a 10

mL solution of sterile 6% (w/v) cellulase and a 10 mL solution of sterile 0.6% (w/v) pectolyase enzyme solutions were pipette into the conical tube using a 10 mL pipette tip. Next, 4 pack cell volumes (PCV) of Hi-II suspension cells were added into the 50 mL tube containing the digest solution and wrapped with parafilm. The tubes were 5 placed on a platform rocker overnight at room temperature for ~16-18 hrs. The next morning, the tubes were removed from the shaker. In a sterile 50 mL conical tube the cells and enzyme solution were slowly filtered through a 100 µm cell strainer. Next, the cells were rinsed using a 100 µm cell strainer by pipetting 10 mLs of W5 media through the strainer. In a sterile 50 mL conical tube, the cells and enzyme solution 10 were slowly filtered through a 70 µm cell strainer. This straining step was followed by a second straining step, wherein the cells and enzyme solution were slowly strained into a 50 mL conical tube through a 40 µm cell strainer. Using a 10 mL pipette tip, the 40 µm cell strainer was rinsed with 10 mL of W5 media to give a final volume of 40 mL and the tube was inverted. Very slowly, 8 mL of sucrose cushion was added to 15 the bottom of the protoplast/enzyme solution. Using a centrifuge with a swing arm bucket rotor, the tubes were spun for 15 minutes at 1500 rpm. The protoplast cells were removed using a 5 mL narrow bore pipette tip. These cells (7-8 mLs) which were observed as a protoplast bane were removed very slowly and put into a sterile 50 mL conical tube. Next, 25 mL of W5 media was used to wash the tubes. The W5 20 media was added and the tubes were inverted slowly and centrifuge for 10 minutes at 1500 rpm. The supernatant was removed and 10mL of MMG solution was added with slow inversion of the tube to resuspend protoplast pellet. The density of protoplasts were determined using a haemocytometer, the 4 PCV yields ~30 million protoplasts.

25 Protoplast Transformation:

[0399] The protoplast cells were diluted to 1.6 million protoplasts per ml using an MMG solution. The protoplasts were gently resuspended by slowly inverting the tube. Next, 300 µL of protoplasts (~500k protoplasts) were added to a sterile 2 mL tube, the tubes were inverted to evenly distribute the protoplast cells. Plasmid 30 DNA of a concentration about 40 -80 µg suspended in TE buffer was added to the protoplasts. The different experimental conditions are described in Table 28. The tubes were slowly rolled to suspend the DNA with the protoplasts and the tubes were incubated for 5-10 minutes at room temperature. Next 300 µL of PEG solution was

added to the protoplast/DNA solution. Once all the PEG solution had been added, the PEG solution was mixed with the protoplast solution by gently inverting the tube. The cocktail was incubated at room temperature for 15-20 minutes with periodic inverting of the tube(s). After the incubation, 1 mL of W5 solution was slowly added to the 5 tubes and the tubes were gently inverted. Finally, the solution was centrifuged at 1000 rpm for 15 minutes. The supernatant was carefully removed so as not to disturb the cell pellet. Finally, 1 mL of washing/incubating solution was added. The tubes were gently inverted to resuspend the cell pellet. The tubes were covered with aluminum foil to eliminate any exposure to light, and were laid on a rack on their side 10 to incubate overnight. The cells were harvested 24 hours post-transformation for molecular analysis.

Table 28: Different treatment groups were used for the transformation of the protoplast cells. The differing concentrations of the DNA used for the transformations are described below.

Treatment Groups	Donor DNA pDAB100651 (μg)	E32 ZFN6 pDAB105906 (μg)	pUC19 Filler (μg)	Salmon Sperm DNA Filler (μg)	Total DNA (μg)
E32 Donor alone + No enzyme control (filler-1)	pDAB100651 (40 μg)	N/A (0 μg)	pUC19 (40 μg)	N/A (0 μg)	80
E32 Donor alone + No enzyme control (filler-2)	pDAB100651 (40 μg)	N/A (0 μg)	N/A (0 μg)	ssDNA (40 μg)	80
E32 Donor alone control (no filler)	pDAB100651 (40 μg)	N/A (0 μg)	N/A (0 μg)	N/A (0 μg)	40
E32 ZFN6 alone control (no donor) filler1	N/A (0 μg)	pDAB105906 (4 μg)	pUC19 (76 μg)	N/A (0 μg)	80
E32 ZFN6 alone control (no donor) filler2	N/A (0 μg)	pDAB105906 (4 μg)	N/A (0 μg)	ssDNA (76 μg)	80
E32 ZFN6 wt FokI alone control (no donor) No filler	N/A (0 μg)	pDAB105906 (40 μg)	N/A (0 μg)	N/A (0 μg)	40
E32 ZFN6 wt FokI+ E32 Donor (1:10) filler1	pDAB100651 (40 μg)	pDAB105906 (4 μg)	pUC19 (36 μg)	N/A (0 μg)	80
E32 ZFN6 wt FokI+ E32 Donor (1:10) filler2	pDAB100651 (40 μg)	pDAB105906 (4 μg)	N/A (0 μg)	ssDNA (36 μg)	80

Sequence Validation of Targeting

[0400] The results of the ZFN cleavage activity in maize protoplasts were confirmed using the Next Generation Sequencing protocol described above. The sequenced PCR amplified fragments were scored for the presence of sequence variants resulting from indels. Event 32 ZFN6 cleaved the genomic locus of transgenic Corn Event DAS-59132 at about 1.5% of NHEJ/10 ng of targeted amplicon.

[0401] Targeting of an AAD-1 donor cassette into the genomic locus of transgenic Corn Event DAS-59132 into the Hi-II maize transgenic cell suspensions via Non Homologous End Joining (NHEJ) was confirmed via an in-out PCR reaction. The in-out PCR reaction was completed, wherein a first PCR reaction was designed to amplify the junction of the AAD-1 donor and genomic locus of transgenic Corn Event DAS-59132. The resulting amplicon was subjected to a second PCR reaction, wherein primers were designed to bind internally within the first amplicon. The combination of two independent PCR reactions resulted in the removal of background amplifications which may be false-positives. The in-out PCR results of the protoplast transformation experiments demonstrated that the genomic locus of transgenic Corn Event DAS-59132 could be reproducibly targeted with a 5.3 kb AAD1 plasmid donor and the E32 ZFN6 zinc finger nuclease at a ratio of 1:10 µg of DNA (with and without filler DNA comprised of either pUC19 plasmid DNA or salmon sperm DNA). Targeting via a NHEJ method was evidenced by the insertion of the AAD-1 donor cassette in both orientations. The sequence data produced from the PCR reactions resulted in three instances of perfect integration of the donor DNA. Thus it was possible to demonstrate donor targeting into an endogenous maize locus using ZFNs via a NHEJ-DSB repair mechanism.

WHISKERS™ mediated stable transformation of ZFN and donor for targeted integration

[0402] Transgenic events were targeted to the endogenous genomic locus of Corn Event DAS-59132. Constructs as described in Example 2 include the donor sequence (pDAB100655) and Event 32 ZFN 6 (pDAB105906).

[0403] Maize callus cells, consisting of 12 mL of packed cell volume (PCV) from a previously cryo-preserved cell line plus 28 mL of conditioned medium was

subcultured into 80 mL of GN6 liquid medium (N6 medium (Chu *et al.*, (1975) *Sci Sin.* 18:659–668), 2.0 mg/L of 2, 4-D, 30 g/L sucrose, pH 5.8) in a 500 mL Erlenmeyer flask, and placed on a shaker at 125 rpm at 28°C. This step was repeated two times using the same cell line, such that a total of 36 mL PCV was distributed 5 across three flasks. After 24 hours, the GN6 liquid media was removed and replaced with 72 mL GN6 S/M osmotic medium (N6 Medium, 2.0 mg/L 2,4-D, 30 g/L sucrose, 45.5 g/L sorbitol, 45.5 g/L mannitol, 100 mg/L myo-inositol, pH 6.0). The flask was incubated in the dark for 30-35 minutes at 28° C with moderate agitation (125 rpm). During the incubation period, a 50 mg/mL suspension of silicon carbide

10 WHISKERS™ (Advanced Composite Materials, LLC, Greer, SC) was prepared by adding 8.1 mL of GN6 S/M liquid medium to 405 mg of sterile, silicon carbide WHISKERS™.

[0404] Following incubation in GN6 S/M osmotic medium, the contents of each flask were pooled into a 250 mL centrifuge bottle. After all cells in the flask 15 settled to the bottom, the content volume in excess of approximately 14 mL of GN6 S/M liquid was drawn off and collected in a sterile 1-L flask for future use. The pre-wetted suspension of WHISKERS™ was mixed at maximum speed on a vortex for 60 seconds, and then added to the centrifuge bottle.

[0405] In this example, 159 µg of pDAB100655 (donor sequence) and 11 µg 20 of pDAB10506 (ZFN) plasmid DNA were added to each bottle. Once the plasmid DNA was added, the bottle was immediately placed in a modified RED DEVIL 5400™ commercial paint mixer (Red Devil Equipment Co., Plymouth, MN), and agitated for 10 seconds. Following agitation, the cocktail of cells, media, WHISKERS™ and plasmid DNA were added to the contents of a 1-L flask along 25 with 125 mL fresh GN6 liquid medium to reduce the osmoticant. The cells were allowed to recover on a shaker set at 125 rpm for 2 hours. 6 mL of dispersed suspension was filtered onto Whatman #4 filter paper (5.5 cm) using a glass cell collector unit connected to a house vacuum line such that 60 filters were obtained per bottle. Filters were placed onto 60 x 20 mm plates of GN6 solid medium (same as 30 GN6 liquid medium except with 2.5 g/L Gelrite gelling agent) and cultured at 28° C under dark conditions for 1 week.

Identification and isolation of putative targeted events integrated within the Corn Event DAS-59132 genomic locus

[0406] One week post-DNA delivery, filter papers were transferred to 60x20

mm plates of GN6 (1H) selection medium (N6 Medium, 2.0 mg/L 2, 4-D, 30 g/L

5 sucrose, 100 mg/L myo-inositol, 2.5 g/L Gelrite, pH 5.8) containing a selective agent.

These selection plates were incubated at 28° C for one week in the dark. Following 1 week of selection in the dark, the tissue was embedded onto fresh media by scraping

½ the cells from each plate into a tube containing 3.0 mL of GN6 agarose medium

held at 37-38° C (N6 medium, 2.0 mg/L 2, 4-D, 30 g/L sucrose, 100 mg/L myo-

10 inositol, 7 g/L SEAPLAQUE® agarose, pH 5.8, autoclaved for 10 minutes at 121° C).

[0407] The agarose/tissue mixture was broken up with a spatula and,

subsequently, 3 mL of agarose/tissue mixture was evenly poured onto the surface of a

100 x 25 mm PETRI™ dish containing GN6 (1H) medium. This process was

repeated for both halves of each plate. Once all the tissue was embedded, plates

15 incubated at 28° C under dark conditions for up to 10 weeks. Putatively transformed

isolates that grew under these selection conditions were removed from the embedded

plates and transferred to fresh selection medium in 60 x 20 mm plates. If sustained

growth was evident after approximately 2 weeks, an event was deemed to be resistant

to the applied herbicide (selective agent) and an aliquot of cells was subsequently

20 harvested for genotype analysis. In this example, 24 events were recovered from 6

bottles treated. These events were advance for molecular analysis to confirm the

integration of the AAD-1 gene within a genomic locus of Corn Event DAS-59132.

Molecular analysis of NHEJ targeting of the Corn Event DAS-59132 genomic

25 **locus**

[0408] The 24 events that were recovered from the WHISKERS™ mediated

transformation, as described above, were analyzed using several different molecular

confirmation tools. As a result of the analysis events which contained a copy of the

AAD-1 transgene integrated within the Corn Event DAS-59132 genomic locus were

30 identified. Initially the 24 various events were confirmed to contain a copy of the

AAD-1 transgene, next the events were analyzed to determine whether the genomic

locus of Corn Event DAS-59132 which would suggest that a copy of the AAD-1

transgene had integrated via NHEJ within the genome of the maize cells. The events

which were identified to contain a copy of the AAD-1 transgene within the genomic locus of Corn Event DAS-59132 were further confirmed via In-Out PCR and Southern blot reactions. These assays confirmed that events containing a copy of the AAD-1 transgene integrated within the Corn Event DAS-59132 genomic locus via an

5 NHEJ mechanism.

[0409] DNA extraction: DNA was extracted from lyophilized maize callus tissue using a QIAGEN BIOSPRINT 96™ DNA isolation kit per manufacturer's recommendations. A pre-defined program was used for the automation extraction and DNA was eluted in 200 µl of 1:1 TE Buffer/distilled water. Two microliters (2 µL) of 10 each sample was quantified on THERMOSCIENTIFIC NANODROP 8000™ and samples were normalized to 100 ng/µL using QIAGEN BIOROBOT 3000™. Normalized DNA was stored at 4°C until further analysis.

[0410] Copy Number Estimation: Transgene copy number determination by hydrolysis probe assay, analogous to TAQMAM® assay, was performed by real-time 15 PCR using the LIGHTCYCLER®480 system (Roche Applied Science, Indianapolis, IN). Assays were designed for AAD-1 and the internal reference gene Invertase using LIGHTCYCLER® Probe Design Software 2.0. For amplification, LIGHTCYCLER®480 Probes Master mix (Roche Applied Science, Indianapolis, IN) was prepared at 1X final concentration in a 10 µL volume multiplex reaction 20 containing 0.4 µM of each primer and 0.2 µM of each probe (Table 29). A two-step amplification reaction was performed with an extension at 60°C for 40 seconds with fluorescence acquisition. Analysis of real time PCR copy number data was performed using LIGHTCYCLER® software release 1.5 using the relative quant module and is based on the $\Delta\Delta Ct$ method. For this, a sample of gDNA from a single copy calibrator 25 and a known two-copy check were included in each run.

Table 29: Primer/Probe Sequences for hydrolysis probe assay of AAD1 and internal reference (Inv).

Primer Name	Sequence	Detection
GAAD1F	SEQ ID NO:454 ; 5' TGTCGGTCCCTCTACCAA 3'	--
GAAD1R	SEQ ID NO:455 ; 5' CAACATCCATCACCTTGACTGA 3'	--
GAAD1R	SEQ ID NO:456 ; 5' CACAGAACCGTCGCTTCAGCAACA 3'	FAM
IVF-Taq	SEQ ID NO:457 ; 5' TGGCGGACGACGACTTGT 3'	--
IVR-Taq	SEQ ID NO:458 ; 5'	--

	AAAGTTGGAGGCTGCCGT 3'	
IV-Probe	SEQ ID NO:459 ; 5' CGAGCAGACCGCCGTACTTCTACC 3'	HEX

Corn Event DAS-59132 Genomic Locus Disruption Assay:

[0411] A genomic locus disruption assay for Corn Event DAS-59132 was

5 performed by real-time PCR using the LIGHTCYCLER®480 system (Roche Applied Science, Indianapolis, IN). Assays were designed to monitor the specificity for which Event 32 ZFN6 (25716/25717) bound and cleaved genomic sequences of the Corn Event DAS-59132 locus and the internal reference gene *IVF* using the

LIGHTCYCLER® Probe Design Software 2.0. For amplification,

10 LIGHTCYCLER®480 Probes Master mix (Roche Applied Science, Indianapolis, IN) was prepared at 1X final concentration in a 10 µL volume multiplex reaction containing 0.4 µM of each primer and 0.2 µM of each probe (Table 30). A two-step amplification reaction was performed with an extension at 55°C for 30 seconds with fluorescence acquisition. Analysis for the disruption assay was performed using

15 target to reference ratio (Figure 37). Four of the eight events were identified as containing an AAD-1 transgene integrated into the genomic locus of Corn Event DAS-59132. The following events, consisting of; Event 100655/105906[1]-001, Event 100655/105906[5]-013, Event 100655/105906[5]-015, and Event

100655/105906[3]-018, were advance for further molecular analysis to confirm the

20 integration of the AAD-1 transgene within the genomic locus of Corn Event DAS-59132.

Corn Event DAS-59132 Locus Specific In-Out PCR:

[0412] The insertion of the AAD-1 donor within the genomic locus of Corn

25 Event DAS-59132 via NHEJ can occur in one of two orientations. The integration of the AAD-1 transgene and the orientation of this integration were confirmed with an In-Out PCR assay. The In-Out PCR assay utilizes an “Out” primer that was designed to bind to the genomic locus of Corn Event DAS-59132 target sequence. In addition, an “In” primer was designed to bind to the AAD-1 donor sequence. The amplification reactions which were completed using these primers only amplify a donor gene which is inserted at the target site. The resulting PCR amplicon was produced from the two

primers, and consisted of a sequence that spanned the junction of the insertion. For each sample, two sets of In-Out PCR primers were multiplexed into one reaction and were used to detect the NHEJ-mediated donor insertion which could occur in one of two different orientations. Positive and negative controls were included in the assay.

5 Two positive control plasmids, pDAB100664 and pDAB100665, were constructed to simulate donor insertion at the genomic locus of Corn Event DAS-59132 in one of the two different orientations.

[0413] A DNA intercalating dye, SYTO-13, was used in the PCR mix in order to detect amplification in real time on a thermocycler with fluorescence detection 10 capability. In addition, a melting temperature (Tm) analysis program was attached to a regular PCR program so the amplified products could be analyzed for their Tm profiles. The similarity in the Tm profile between an unknown sample and the positive control sample strongly suggests that the unknown sample has the same amplified product as that of the positive control. The PCR reactions were conducted 15 using 10 ng of template genomic DNA, 0.2 μ M dNTPs, 0.2 μ M forward and reverse primers, 4 μ M SYTO-13 and 0.15 μ l of Ex Taq HS. Reactions were completed in two steps: the first step consisted of one cycle at 94°C (2 minutes) and 35 cycles at 98°C (12 seconds), 66°C (30 seconds) and 68°C (1.3 minutes); the second step was a Tm program covering 60-95°C followed by 65°C (30 seconds) and 72°C (10 minutes) 20 (Table 30). The amplicons were sent out for sequencing to confirm that the AAD-1 gene had integrated within the genomic locus of Corn Event DAS-59132.

[0414] The results of the real-time, In-Out PCR amplicons were visualized 25 using the ABI software. These results were further confirmed using a gel shift assay, wherein the amplicons were run on a 1.2% TAE gel. Expected amplicon sizes were ~ 1.8kb for the first orientation (as in pDAB100664) and ~2kb for the second orientation (as in pDAB100665). The gel shift assay results confirmed the real-time, In-Out PCR reaction data. Both sets of data suggested that a copy of the AAD-1 transgene had integrated via NHEJ within the genome of the maize cells at the genomic locus of Corn Event DAS-59132.

Table 30: Primers for In-Out PCR to detect NHEJ mediated targeting at Corn Event DAS-59132 in maize cells.

Primer Name	Sequence	Expected Amplicon size / control
E32-3R2 NJ-AAD1- Pri2	Forward Primer SEQ ID NO:460 5' GCC CTT ACA GTT CAT GGG CG 3' Reverser Primer SEQ ID NO:461 5' GAC CAA GTC CTT GTC TGG GAC A 3'	1.8 kB pDAB100664
E32-5F1 NJ-AAD1- Pri2	Forward Primer SEQ ID NO:462 5' ACA AAC ACG TCC TCC AAG GCT 3' Reverse Primer SEQ ID NO:463 5' GAC CAA GTC CTT GTC TGG GAC A 3'	2.0 kB pDAB100665

[0415] Southern Blot Analysis: The maize callus events identified above were further screened using a Southern blot assay. This assay was used to further confirm that the AAD-1 transgene had integrated via NHEJ within the genome of the maize cells at the genomic locus of Corn Event DAS-59132. The Southern blot analysis experiments generated data which demonstrated the integration and integrity of the AAD-1 transgene within the soybean genome.

[0416] DNA Extraction: Genomic DNA was extracted from the callus tissue harvested from each individual event. Initially, the tissue samples were collected in 2 mL tubes and lyophilized for 2 days. Tissue maceration was performed with a KLECO TISSUE PULVERIZER™ and tungsten beads (Kleco, Visalia, CA). Following tissue maceration the genomic DNA was isolated using the DNEASY PLANT MINI KIT™ (Qiagen, Germantown, MD) according to the manufacturer's suggested protocol.

[0417] Southern Blot: Genomic DNA (gDNA) was quantified using the QUANT-IT PICO GREEN DNA ASSAY KIT™ (Molecular Probes, Invitrogen, Carlsbad, CA). Quantified gDNA was adjusted to 4 µg for the Southern blot analysis. DNA samples were then digested using the NcoI restriction enzyme (New England BioLabs, Ipswich, MA) overnight at 37 °C followed with a clean-up using QUICK-PRECIP™ (Edge BioSystem, Gaithersburg, MD) according to the manufacturer's suggested protocol. DNA was resuspended in 1X dye and electrophoresed for 5 hours on a .8% SEAKEM LE AGAROSE GEL™ (Lonza, Rockland, ME) at 110 volts in a cold room. The gel was denatured, neutralized, and then transferred to a nylon

charged membrane (Millipore, Bedford, MA) overnight and DNA was crosslinked to the membrane using the UV STRATA LINKER 1800™ (Stratagene, La Jolla, CA), and blots were prehybridized with 20 mL of PERFECTHYB PLUS™ (Sigma, St.

5 Louis, MO). The 226 bp probe (SEQ ID NO:464) was labeled using PRIME-IT RMT RANDOM™ (Stratagene, La Jolla, CA) according to manufacturer's suggested protocol and purified using the PROBE QUANT G-50 MICRO COLUMNS™ (GE Healthcare, Buckinghamshire, UK) per manufacturer's suggested protocol.

Approximately, 20,000,000 cpm of the labeled probe was added to the blots and 10 incubated overnight. Blots were washed twice for 15 minutes per wash and placed on a phosphor image screen for 24 hours and analyzed by a STORM 860 SCANNER™ (Molecular Dynamics).

[0418] Expected and observed fragment sizes with the NcoI digest and probe, based on the known restriction enzyme sites AAD-1 and Corn Event DAS-59132 resulted from the Southern blots. Two DNA fragments were identified from these 15 digests and hybridizations. Southern blots which produced results with bands at sizes of around 2.9 and 5.5 kb indicated that the AAD-1 transgene had integrated into the genomic locus of Corn Event DAS-59132 via an NHEJ mechanism.

Example 9: Targeted integration and disruption of corn Engineered Landing

20 **Pad**

Characterization of ELP genomic target sequence

[0419] The genomic locus in which an Engineered Landing Pad (ELP) integrated was selected as an endogenous genomic loci for gene targeting. The construction of the ELP sequences which comprise Zinc Finger binding sites (eZFN1 and eZFN3) and about 1.0 kb of random artificial sequences which flank the Zinc 25 Finger binding sites, in addition to the Zinc Finger Nuclease proteins are described in International Patent Application WO2011091317, herein incorporated by reference in its entirety. To test NHEJ-mediated targeted integration within the ELP loci, two donor DNAs were constructed, both of which contain one of the two eZFN binding 30 sites in a 5.3 kb plasmid comprising an *aad-1* gene which confers resistance to the herbicide haloxyfop. Figure 38 presents a representative schematic of the integration.

Regeneration of transgenic plant events comprising an ELP

[0420] Four transgenic ELP events produced from the transformation of pDAB100640, pDAB100641 and pDAB106685 (two for each ELP) were generated as described in International Patent Application WO2011091317, herein incorporated by reference in its entirety. The events were obtained and confirmed to be single copy and contain an intact PTU comprising the ELP. These events were regenerated to produce donor material for targeting. Healthy growing tissue was transferred first to 28(1H) (MS medium (Murashige and Skoog (1962) *Physiol Plant* 15:473-497), 0.025 mg/L 2,-4D, 5 mg/L BAP, 1.0 mg/L Herbiace, 30 g/L sucrose, 2.5 g/L gelrite, pH 5.7)

10 and incubated in low light (14 μ E/m²·sec 16 hr photoperiod) for 7 days followed by high light (89 μ E/m²·sec 16 hr photoperiod) for another 7 days. Greening structures were transferred to 36(1H) which is the same as 28(1H) minus the BAP and 2,4-D and incubated in high light (40 μ E/m²·sec 16 hr photoperiod) until shoot structures developed sufficient roots for transplanting to greenhouse. Plants were grown to 15 maturity in greenhouse using mix of 95% METRO-MIX 360® and 5% clay/loam soil and pollinated dependent on health of plant. Vigorously growing plants were selfed or sibbed (plants from same event) and less vigorous plants were crossed with Hi-II or A188 to maintain embryogenic capacity of donor material. T₁ seed was planted in 4" pots and germinating seedlings were screened for zygosity via qPCR. Seedlings 20 determined to be homozygous for the PAT gene were transferred to 5 gallon pots, grown to reproductive stage, outcrossed to Hi-II and T₂ embryos used for targeting via NHEJ mediated integration.

NHEJ targeting of ELP protoplasts

25 [0421] Zea mays Hi-II protoplasts were obtained and transformed using the previously described protoplast transformation protocol. Donor plasmid DNA of pDAB100651 was transformed with the ZFN plasmid DNA of pDAB105941. Likewise, donor plasmid DNA of pDAB100652 was transformed with the ZFN plasmid DNA of pDAB105943. The donor DNAs were transformed with Zinc 30 Finger Nucleases into the ELP transgenic plants. Upon the introduction of the donor DNA and the eZFNs, both the donor DNA and the ELP loci with is integrated within the genomic target DNA were cleaved. The donor DNA was subsequently inserted into the genomic target. The insertion of the donor DNA within the ELP genomic loci

can occur in either orientation. Insertion of the donor DNA in the direct orientation will result in a junction sequence that corresponds to the expected annealing and ligation of the 4 bp single-stranded complementary ends generated from ZFN cleavage. Insertions and deletions (indels) at the junctions are common. Insertion of the donor DNA in the reverse orientation will result in both junction fractions that contain indels.

5 [0422] Protoplast Isolation: Maize Hi-II embryogenic suspension cultures were obtained and were maintained on a 3.5 day maintenance schedule. In a 50 mL sterile conical tube a 10 mL solution of sterile 6% (w/v) cellulase and a 10 mL 10 solution of sterile 0.6% (w/v) pectolyase enzyme solutions were pipette into the conical tube using a 10 mL pipette tip. Next, 4 pack cell volumes (PCV) of Hi-II suspension cells were added into the 50 mL tube containing the digest solution and wrapped with parafilm. The tubes were placed on a platform rocker overnight at room temperature for ~16-18 hrs. The next morning, the tubes were removed from the 15 shaker. In a sterile 50 mL conical tube the cells and enzyme solution were slowly filtered through a 100 μ m cell strainer. Next, the cells were rinsed using a 100 μ m cell strainer by pipetting 10 mLs of W5 media through the strainer. In a sterile 50 mL conical tube, the cells and enzyme solution were slowly filtered through a 70 μ m cell strainer. This straining step was followed by a second straining step, wherein the cells 20 and enzyme solution were slowly strained into a 50 mL conical tube through a 40 μ m cell strainer. Using a 10 mL pipette tip, the 40 μ m cell strainer was rinsed with 10 mL of W5 media to give a final volume of 40 mL and the tube was inverted. Very slowly, 8 mL of sucrose cushion was added to the bottom of the protoplast/enzyme 25 solution. Using a centrifuge with a swing arm bucket rotor, the tubes were spun for 15 minutes at 1500 rpm. The protoplast cells were removed using a 5 mL narrow bore pipette tip. These cells (7-8 mLs) which were observed as a protoplast bane were removed very slowly and put into a sterile 50 mL conical tube. Next, 25 mL of W5 media was used to wash the tubes. The W5 media was added and the tubes were inverted slowly and centrifuge for 10 minutes at 1500 rpm. The supernatant was 30 removed and 10mL of MMG solution was added with slow inversion of the tube to resuspend protoplast pellet. The density of protoplasts were determined using a haemocytometer, the 4 PCV yields ~30 million protoplasts.

[0423] Protoplast Transformation: The protoplast cells were diluted to 1.6 million protoplasts per ml using an MMG solution. The protoplasts were gently resuspended by slowly inverting the tube. Next, 300 μ L of protoplasts (~500k protoplasts) were added to a sterile 2 mL tube, the tubes were inverted to evenly 5 distribute the protoplast cells. Plasmid DNA of a concentration about 80 μ g was suspended in TE buffer was added to the protoplasts. Both the ZFN and donor 10 plasmid constructs were transformed. The eZFN expressing plasmids (pDAB105941 and pDAB105943) were added alone or in combination at a 1:1 or 10:1 ratio of Donor DNA (pDAB100651 and pDAB100652) to eZFN DNA for each of the eZFN1 and 15 eZFN3 treatments. The efficacy of the eZFNs had previously been tested and are described in more detail in International Patent Application WO2011091317, Figure 39 is provided as a representation of the eZFN relative to the donor DNA. The tubes were slowly rolled to suspend the DNA with the protoplasts and the tubes were 20 incubated for 5-10 minutes at room temperature. Next 300 μ L of PEG solution was added to the protoplast/DNA solution. Once all the PEG solution had been added, the PEG solution was mixed with the protoplast solution by gently inverting the tube. The cocktail was incubated at room temperature for 15-20 minutes with periodic inverting of the tube(s). After the incubation, 1 mL of W5 solution was slowly added to the tubes and the tubes were gently inverted. Finally, the solution was centrifuged at 25 1000 rpm for 15 minutes. The supernatant was carefully removed so as not to disturb the cell pellet. Finally, 1 mL of washing/incubating solution was added. The tubes were gently inverted to resuspend the cell pellet. The tubes were covered with aluminum foil to eliminate any exposure to light, and were laid on a rack on their side to incubate overnight. The cells were harvested 24 hours post-transformation for molecular analysis to identify ELP loci which contained a donor integrated via NHEJ-mediated DNA repair.

Table 31: DNA concentrations and of the eZFN and Donor plasmid DNA that were transformed into the maize cells and integrated within the ELP loci.

Treatment Groups	Donor DNA (pDAB)	Amount (μ g)	eZFNs (pDAB)	Amount (μ g)	Filler Plasmid (μ g)	Total DNA (μ g)
eZFN1 Donor alone	100651	40	---	---	40	80
eZFN1 alone	---	---	105941	40	40	80
eZFN1 Donor + eZFN1 (1:1)	100651	40	105941	40	0	80

eZFN1 Donor + eZFN1 (10:1)	100651	40	105941	4	36	80
---	---	---	---	---	---	---
eZFN3 Donor alone	100652	40	---	---	40	80
eZFN3 alone	---	---	105943	40	40	80
eZFN3 Donor + eZFN3 (1:1)	100652	40	105943	40	0	80
eZFN3 Donor + eZFN3 (10:1)	100652	40	105943	4	36	80

Molecular confirmation of maize ELP targeting by NHEJ in protoplasts

[0424] DNA extraction: DNA was extracted from maize tissue using a Qiagen BIOSPRINT 96™ robot via automation and DNA was eluted in 200 µl of 1:1 TE

5 Buffer/distilled water. DNA of each sample was quantified on THERMOSCIENTIFIC NANODROP 8000™ and samples were normalized to 100 ng/µL using QIAGEN BIOROBOT 3000™. Normalized DNA was stored at 4°C till further analysis.

[0425] ELP locus disruption assay: After harvesting the protoplasts at 24 hr,

10 DNA was extracted and analyzed using a disruption assay, junction analysis using PCR and sequencing of the DNAs produced by the junction PCR. The disruption assay is an indirect measure of relative eZFN cleavage activity. The TAQMAN™-based assay determines the loss of intact eZFN binding sites in the target DNA as would be expected due to misrepair of the DNA ends that can occur upon ligation of 15 the ends.

[0426] The data from the TAQMAN™-based assay suggests that eZFN1 has a higher activity than does eZFN3 and also that there is significant cleavage of the target DNA by the eZFNs as demonstrated by the statistically significant reduction in the signal in the eZFN samples compared to the donor alone samples.

20 [0427] Cleavage of genomic target DNA by eZFNs. DNA was isolated from each treatment group (6 replicates each) as indicated. Taqman assays were used to measure cleavage of the target DNA. ELP locus disruption assay was performed by real-time PCR using the LIGHTCYCLER®480 system (Roche Applied Science, Indianapolis, IN). Assays were designed to monitor eZFN1 and eZFN3 binding 25 sequences within ELP1 and the internal reference gene *IVF* using LIGHTCYCLER®

Probe Design Software 2.0. For amplification, LIGHTCYCLER®480 Probes Master mix (Roche Applied Science, Indianapolis, IN) was prepared at 1X final concentration in a 10 µL volume multiplex reaction containing 0.4 µM of each primer and 0.2 µM of each probe (Table 33). A two-step amplification reaction was 5 performed with an extension at 55°C for 30 seconds with fluorescence acquisition. Analysis for the disruption assay was performed using target to reference ratio. The location of the primers are shown in Figure 40.

Table 33: Primer and probe sequences for the disruption assay

Primer Name	Sequence	Detection
MAS622	SEQ ID NO:471 ; 5' TAGGAGTTCTCTTTATGCCACCC 3'	---
MAS621	SEQ ID NO:472 ; 5' CCTTGGGATTTCAGTTGGTAGGTT 3'	---
UPL67 Probe	Obtained from Roche Biosciences, Indianapolis, IN	FAM
IVF-Taq	SEQ ID NO:473 ; 5' TGGCGGACGACGACTTGT 3'	---
IVR-Taq	SEQ ID NO:474 ; 5' AAAGTTGGAGGCTGCCGT 3'	---
IV-Probe	SEQ ID NO:475 ; 5' CGAGCAGACCGCCGTGTACTTCTACC 3'	HEX

10 [0428] Locus Specific In-Out PCR: Donor insertion at the expected ELP site using NHEJ-mediated repair can result in two different orientations. Two positive control plasmids, pDAB100660 and pDAB100662 were constructed and transformed into plants to simulate donor insertion within the ELP site. Transgenic plants which were produced with the pDAB100660 and pDAB100662 control constructs were 15 assayed using an In-Out PCR design and the conditions for the In-Out PCR assay were determined on these plant materials.

[0429] The In-Out PCR assay has an “Out” primer in the ELP sequence while an “In” primer is placed in the donor sequence, so that only when the donor is inserted in the target site, the two primers would amplify a sequence that spans the junction of 20 the insertion (Table 34). For each sample, two In-Out PCR reactions were used to detect donor insertion with two different orientations. Positive and negative controls were included in the assay.

[0430] A DNA intercalating dye, SYTO-13, was used in the PCR mix in order to detect amplification in real-time on a thermocycler with fluorescence detection capability. In addition, a melting temperature (Tm) analysis program was attached to a regular PCR program so the amplified products can be analyzed for their Tm profiles.

5 Similarity of the Tm profile between an unknown sample and the positive control strongly suggests the unknown sample has the same amplified product as that of the positive control. Positive targeted samples identified in the real-time, In-Out PCR assay were further visualized using gel shift analysis. Expected amplicon sizes are ~1.5kb for Orientation 1 (as in pDAB100662), and ~1.4kb for Orientation 2 (as in 10 pDAB100660).

Table 34: Primers for In-Out PCR to detect NHEJ mediated targeting at ELP1 in maize

Primer Name	Sequence	Expected Amplicon size / control
ELP1-PriF1 NJ-AAD1-Pri2	SEQ ID NO:476 5' AGA CCT ACC ACC CAT TAG GGC 3' SEQ ID NO:477 5' GAC CAA GTC CTT GTC TGG GAC A 3'	1.5 kB pDAB100662
NJ-AAD1-Pri2 ELP2-PriR1	SEQ ID NO:478 5' GAC CAA GTC CTT GTC TGG GAC A 3' SEQ ID NO:479 5' GAT GGT GGT TAT GAC AGG CTC CT 3'	1.4 kB pDAB100660

[0431] PCR reactions were conducted using 10 ng of template genomic DNA, 15 0.2 µM dNTPs, 0.2 µM forward and reverse primers, 4 µM SYTO-13 and 0.15 µl of Ex Taq HS. Reactions were completed in two steps: the first step consisted of one cycle at 94°C (2 minutes) and 35 cycles at 98°C (12 seconds), 66°C (30 seconds) and 68°C (1.3 minutes); the second step was a Tm program covering 60-95°C followed by 65°C (30 seconds) and 72°C (10 minutes). Products were visualized using the ABI 20 software as well as by running on a 1.2% TAE gel.

[0432] The In-out PCR junction analysis for eZFN1 and eZFN3 of NHEJ-directed targeting of donor treatments that included donor DNA alone, eZFN alone, or Donor DNA with eZFN DNA (at a ratio of either 1:1 or 10:1) were run out on agarose gels. The results indicated that the PCR amplicon size of the donor and eZFN DNA 25 was that expected for an NHEJ targeted event.

[0433] Sequence of Target/donor Junctions: From the ELP targeted events which were confirmed via in-out PCR analysis, the PCR amplicon products were

confirmed via sequencing and the target-donor junctions were validated by standard Sanger sequencing. Briefly, junction PCR analysis was performed on all replicates of each treatment group. PCR primers were chosen to amplify one side of the insert junction sequences that were either in the direct or reverse orientation. PCR products 5 were observed in samples generated from the eZFN and Donor DNA samples, but not from the control samples, comprising the Donor DNA alone or eZFN alone samples. PCR products were evident in the majority of replicate samples from both ratios of eZFNs and Donor DNAs used.

[0434] Representative samples of the PCR products were cloned and 10 sequenced. For both the direct and reverse orientation, sequences of the PCR products from four different reactions are shown in Figure 41. Nine unique haplotypes were observed for the direct orientation of the insert, as expected from misrepair to the junction ends. Three of the 16 sequences aligned with the sequence expected from annealing and ligation of intact ends of the inserted Donor DNA and 15 the target sequence. All sequences of the PCR products in the reverse orientation had indels at the junctions as expected since the single-stranded ends of the Donor DNA and target DNA are not complementary.

Results of NHEJ-Mediated Donor Targeting in Maize Protoplasts

[0435] A maize protoplast-based transient assay system was developed that 20 showed high, reproducible expression of reporter genes. Protoplasts were derived from a Hi-II maize suspension culture that was developed at DAS. A transgenic line of maize (maize line 106685[1]-007) that harbored an insert of the ELP was used for NHEJ mediated integration of the donor DNA sequence.

25

NHEJ targeting of ELP embryos via microparticle bombardment

[0436] Three days prior to microparticle bombardment, 1.5-2.2 mm embryos 30 were isolated from surface sterilized ears and placed (scutellum-up) onto N6 basal medium and vitamins (Phytotechnology Laboratories, Shawnee Mission, KS) with 2.0 mg/L 2, 4-D, 2.8 g/L proline, 30 g/L sucrose, 100 mg/L casein enzymatic hydrolysate, 100 mg/L myo-inositol and 4.25 mg/L silver nitrate solidified with 2.5 g/L Gelzan (Phytotechnology Laboratories, Shawnee Mission, KS). Four hours prior to microparticle bombardment, ~35-40 embryos were placed (scutellum up) onto in the

center of a 100 x 15 mm Petri dish containing the same medium with the addition of 36.4 g/L sorbitol and 36.4 g/L mannitol.

[0437] Microparticle gold (0.6 micron, BioRad, Hercules, CA,) was prepared for DNA precipitation by weighing out 15 mg into a sterile, siliconized 1.7 mL microcentrifuge tube (Sigma-Aldrich, St. Louis, MO) and 500 μ L of ice cold 100% ethanol was slowly added. After a 15 second sonication in an FS-14 ultrasonic water bath (Fisher Scientific, Nazareth, PA), the gold was allowed to settle for 30 minutes at room temperature prior to centrifugation at 3,000 rpm for 60 seconds using a MiniSpin (Eppendorf, Hauppauge, NY). After removing the supernatant, 1 mL of ice cold, sterile water was added, mixed up and down with the pipette and allowed to settle for 3-5 minutes prior to centrifugation at 3,000 for 60 seconds. The wash step was repeated one more time prior to suspending the gold in 500 μ L of ice cold, sterile water. The washed gold was then aliquoted into separate 1.7 mL sterile, siliconized microcentrifuge tubes (50 μ L per tube) being careful to keep the gold well mixed by pipeting up and down between tubes. The washed gold (~1.5 mg per 50 μ L) was then stored at -20°C until needed.

[0438] For DNA precipitation, one tube containing ~1.5 mg of gold in 50 μ L of water was thawed for each 10 targets to be bombarded and sonicated in an ultrasonic water bath for 15 seconds then placed on ice. Plasmid DNA (0.5 μ g ZFN + 4.5 μ g Donor) was premixed in 0.6 mL microcentrifuge tubes (Fisher Scientific, Nazareth, PA) and added to the gold suspension gently pipeting up and down several times to mix thoroughly. Fifty μ L of ice cold 2.5 M calcium chloride was added and gently mixed by pipeting up and down several times. Twenty μ L of cold 0.1 M spermidine was then added and gently mixed by pipeting up and down several times. The tube was then capped and placed onto a Vortex Genie 2 (Scientific Instruments Inc., Bohemia, NY) and allowed to mix (set at 'shake 2') for 10 minutes after which the mixture was allowed to settle for 3-5 minutes. After centrifuging for 15 seconds at 5,000 rpm, the supernatant was carefully removed and 250 μ L of ice cold, 100% ethanol was added, the tube capped and mixed vigorously by hand to dislodge the pellet. After a second centrifuge for 15 seconds at 5,000 rpm, 120 μ L of ice cold, 100% ethanol was added, the tube capped and mixed vigorously by hand to dislodge the pellet.

[0439] For microparticle bombardment, sterilized macrocarriers (BioRad, Hercules, CA) were fit into stainless steel holders (BioRad, Hercules, CA) and autoclaved. Ten μ L of gold/DNA suspension was evenly spread in the center of the macrocarrier being sure to pipette up and down so as to keep well mixed then placed 5 onto a piece of sterile 125 mm Whatman #4 filter paper (GE Healthcare, Buckinghamshire, UK) on a bed of 8-mesh Drierite (W.A Hammond Drierite Co., Xenia, OH) in a 140 x 25 mm glass Petri dish. The gold/DNA was allowed to dry completely for about 10 minutes. Rupture discs (650 psi, BioRad, Hercules, CA) were sterilized by soaking for a few minutes in isopropyl alcohol then loaded into the 10 retaining cap of a microparticle bombardment devise (PDS-1000, BioRad, Hercules, CA). An autoclaved stopping screen (BioRad, Hercules, CA) and a loaded macrocarrier was placed into the launch assembly, the lid was screwed on and slide into the bombardment chamber just under the nozzle. The Petri dish containing the screen-covered, leaf target was uncovered and placed in the bombardment chamber 6 15 cm below the nozzle. A vacuum was pulled (-0.9 bar) and the devise was fired.

[0440] Next day (16 – 20 hours after bombardment), the bombarded embryos were transferred (scutellum-up) to N6 basal medium and vitamins (Phytotechnology Laboratories, Shawnee Mission, KS) with 2.0 mg/L 2, 4-D, 2.8 g/L proline, 30 g/L sucrose, 100 mg/L casein enzymatic hydrolysate, 100 mg/L myo-inositol and 4.25 20 mg/L silver nitrate solidified with 2.5 g/L Gelzan (Phytotechnology Laboratories, Shawnee Mission, KS). After 7 days - 8 days-post-bombardment, 11 days from culture initiation - embryos were transferred to selection media N6 basal medium and vitamins (Phytotechnology Laboratories, Shawnee Mission, KS) with 2.0 mg/L 2, 4-D, 2.8 g/L proline, 30 g/L sucrose, 100 mg/L casein enzymatic hydrolysate, 100 mg/L 25 myo-inositol, 4.25 mg/L silver nitrate and 0.0362 mg/L R-haloxyfop acid solidified with 2.5 g/L Gelzan (Phytotechnology Laboratories, Shawnee Mission, KS). After two weeks, embryos were transferred to fresh selection N6 basal medium and vitamins (Phytotechnology Laboratories, Shawnee Mission, KS) with 2.0 mg/L 2, 4-D, 2.8 g/L proline, 30 g/L sucrose, 100 mg/L casein enzymatic hydrolysate, 100 mg/L 30 myo-inositol, 4.25 mg/L silver nitrate and 0.181 mg/L R-haloxyfop acid solidified with 2.5 g/L Gelzan (Phytotechnology Laboratories, Shawnee Mission, KS) and after an additional two weeks they were transferred to fresh medium of the same composition.

[0441] Callus growing on 0.181 mg/L R-haloxifop was sampled for molecular analysis. Sampling involved placing either ~50 mg into 1.2 mL cluster tubes for PCR analysis or ~200 mg into 2.0 mL Safe Lock tubes (Eppendorf, Hauppauge, NY) for Southern blot analysis surrounded by dry ice for rapid freezing.

5 The tubes were then covered in 3M micropore tape (Fisher Scientific, Nazareth, PA) and lyophilized for 48 hours in a Virtual XL-70 (VirTis, Gardiner, NY). Once the tissue was lyophilized, the tubes were capped and stored at 8°C until analysis.

Molecular confirmation of maize ELP targeting by NHEJ in plants

10 [0442] DNA extraction: DNA was extracted from maize tissue using a Qiagen BIOSPRINT 96™ robot via automation and DNA was eluted in 200 µl of 1:1 TE Buffer/distilled water. Two µL of each sample was quantified on THERMOSCIENTIFIC NANODROP 8000™ and samples were normalized to 100 ng/µL using QIAGEN BIOROBOT 3000™. Normalized DNA was stored at 4°C till 15 further analysis.

15 [0443] Copy Number Estimation: Transgene copy number determination by hydrolysis probe assay, analogous to TAQMAN® assay, was performed by real-time PCR using the LIGHTCYCLER®480 system (Roche Applied Science, Indianapolis, IN). Assays were designed for the *aad-1* transgene and the internal reference gene 20 Invertase using LIGHTCYCLER® PROBE DESIGN SOFTWARE 2.0. For amplification, LIGHTCYCLER®480 Probes Master mix (Roche Applied Science, Indianapolis, IN) was prepared at 1X final concentration in a 10 µL volume multiplex reaction containing 0.4 µM of each primer and 0.2 µM of each probe (Table 32). A two-step amplification reaction was performed with an extension at 60°C for 40 25 seconds with fluorescence acquisition. Analysis of real time PCR copy number data was performed using LIGHTCYCLER® software release 1.5 using the relative quant module and is based on the $\Delta\Delta Ct$ method. For this, a sample of gDNA from a single copy calibrator and known 2 copy check were included in each run. The location of the primers are shown in Figure 41.

30 **Table 32: Primer/Probe Sequences for hydrolysis probe assay of AAD1 and internal reference (Inv)**

Primer Name	Sequence	Detection
GAAD1F	SEQ ID NO:465 ; 5' TGTCGGTTCCCTCTACCAA 3'	---

GAAD1R	SEQ ID NO:466 ; 5' CAACATCCATCACCTTGACTGA 3'	---
GAAD1R	SEQ ID NO:467 ; 5' CACAGAACCGTCGCTTCAGCAACA 3'	FAM
IVF-Taq	SEQ ID NO:468 ; 5' TGGCGGACGACGACTTGT 3'	---
IVR-Taq	SEQ ID NO:469 ; 5' AAAGTTGGAGGCTGCCGT 3'	---
IV-Probe	SEQ ID NO:470 ; 5' CGAGCAGACCGCCGTACTTCTACC 3'	HEX

[0444] ELP locus disruption assay: The ELP locus disruption assay was performed by real-time PCR using the LIGHTCYCLER®480 system (Roche Applied Science, Indianapolis, IN) as previously described above. Assays were designed to monitor eZFN1 and eZFN3 binding sequences within ELP1 and the internal reference gene *IVF* using LIGHTCYCLER® Probe Design Software 2.0. Analysis for the disruption assay was performed using target to reference ratio. The results are shown in Figure 42.

[0445] Locus Specific In-Out PCR: The In-Out PCR assay was completed as previously described above. The In-out PCR junction analysis for eZFN1 and eZFN3 of NHEJ-directed targeting of donor treatments that included donor DNA alone, eZFN alone, or Donor DNA with eZFN DNA (at a ratio of either 1:1 or 10:1) were run out on agarose gels. The results indicated that the PCR amplicon size of the donor and eZFN DNA was that expected for an NHEJ targeted event.

[0446] Sequence of Target/donor Junctions: From the ELP targeted events which were confirmed via in-out PCR analysis, the PCR amplicon products were confirmed via sequencing and the target-donor junctions were validated by standard Sanger sequencing. Briefly, junction PCR analysis was performed on all replicates of each treatment group. PCR primers were chosen to amplify one side of the insert junction sequences that were either in the direct or reverse orientation. PCR products were observed in samples generated from the eZFN and Donor DNA samples, but not from the control samples, comprising the Donor DNA alone or eZFN alone samples. PCR products were evident in the majority of replicate samples from both ratios of eZFNs and Donor DNAs used.

[0447] Representative samples of the PCR products were cloned and sequenced. For both the direct and reverse orientation, sequences of the PCR products from four different reactions are shown in Figure 43. Nine unique haplotypes were observed for the direct orientation of the insert, as expected from

misrepair to the junction ends. Three of the 16 sequences aligned with the sequence expected from annealing and ligation of intact ends of the inserted Donor DNA and the target sequence. All sequences of the PCR products in the reverse orientation had indels at the junctions as expected since the single-stranded ends of the Donor DNA and target DNA are not complementary.

Southern Blot Analysis: Maize callus that were initially identified as containing a donor sequence integrated within the ELP locus via the locus specific disruption assay and the in-out PCR assay was selected for further analysis by Targeted Integration (TI) Southern blots. For T₁ Southern blots, DNA was digested and probed with enzymes and probes at the target locus. For DNA extraction for Southern blots, tissue samples were collected in 2ml eppendorf tubes (Eppendorf) and lyophilized for 2 days. Tissue maceration was performed with a Kleco tissue pulverizer and tungsten beads (Kleco, Visalia, CA). Following tissue maceration the genomic DNA was isolated using the DNEASY PLANT MINI KIT™ (Qiagen, Germantown, MD) according to the manufacturer's suggested protocol.

[0448] Genomic DNA (gDNA) was quantified by QUANT-IT PICO GREEN DNA™ assay kit (Molecular Probes, Invitrogen, Carlsbad, CA). Quantified gDNA was adjusted to 4 µg for the Southern blot analysis. DNA samples were then digested using the *PmeI* restriction enzyme (New England BioLabs) overnight at 37 °C followed with a clean-up using QUICK-PRECIP™ (Edge BioSystem, Gaithersburg, MD) according to the manufacturer's suggested protocol. DNA was then resuspended in 1X dye and electrophoresed for 5 hours on a .8% SEAKEM LE agarose gel (Lonza, Rockland, ME) at 110 volts in a cold room. The gel was denatured, neutralized, and then transferred to a nylon charged membrane (Millipore, Bedford, MA) overnight and DNA was crosslinked to the membrane using the UV STRATA LINKER 1800™ (Stratagene, La Jolla, CA), and blots were pre-hybridized with 20 ml of PERFECTHYB PLUS™ (Sigma, St. Louis, MO). The probe was labeled using PRIME-IT RMT™ random (Stratagene, La Jolla, CA) according to manufacturer's suggested protocol and purified using PROBE QUANT G-50 MICRO COLUMNS™ (GE Healthcare, Buckinghamshire, UK) according to manufacturer's suggested protocol. Approximately 20,000,000 cpm of the labeled probe was added to the blots and incubated overnight. Blots were washed 2 x15 minutes per wash, placed on a phosphor image screen for 24 hours and analyzed by a STORM 860 SCANNER™

(Molecular Dynamics). Results showed expected bands for targeted integration (~6.9 kB).

Results of ELP loci targeting via NHEJ-mediated integration

5 [0449] The results of this study demonstrate precision insertion in maize of a donor DNA plasmid by NHEJ, subsequent to *in vivo*, ZFN-generated cleavage of the target DNA and the Donor DNA. Targeting of Donor DNAs occurred using two different donor DNAs (different by which eZFN binding sequences were contained in the ELP) and two different eZFNs within protoplasts and maize embryos. Integration 10 within the ELP loci via an NHEJ repair mechanism occurred in both orientations. The donor DNA insert was detected in both of these orientations in the samples tested.

15 [0450] Precision targeting of genes using the NHEJ repair mechanism in the tissues of differing plant species provides significant advantage over known repair mechanisms such as homologous recombination mediated repair. The NHEJ repair mechanism is the dominate repair mechanism which operates in most if not all plant tissues. Conversely, the activity of the homologous recombination mediated repair pathway operates throughout the G2 phase of the cell cycle in plants. Resultingly, many plant tissues capable of transformation do not actively undergo cell division and, hence, would not support gene targeting by a homologous recombination 20 mediated repair pathway. Another advantage of NHEJ repair mediated pathway for donor insertion within the genome of plants, is that unlike the homologous recombination mediated repair pathway, the NHEJ-mediated repair does not require extensive regions of homology, which reduces the size of the donor DNA sequences necessary for genomic insertion. Finally, donor DNA sequences of larger sizes can be 25 successfully inserted into genomic loci using NHEJ repair mediated pathway as compared to the homologous recombination mediated repair pathway. Donor-DNA mediated integration within a genomic locus via the homologous recombination mediated repair pathway requires that a DNA polymerases copy the DNA template contained in the Donor DNA. In contrast, the NHEJ repair mediated pathway only 30 requires interaction of the donor DNA with the target DNA at two points (their ends) and does not require template-dependent DNA synthesis. Accordingly the NHEJ repair mediated pathway can be utilized to integrate larger size donor DNA sequences within the targeted genomic locus.

Example 10: Exemplary Sequences**[0451]** SEQ ID NO:116

5 TCGCCCAAACCCCTGCCGCCGCCATGGCCGCAGCCACCTCCCCGCCGTGCATT
 CTCGGCGCCACCGCCGCCATGCCAAACCCGCCATCCTCTCCCGCGC
 CACCAGCCCCTCGCCGCCGCGCTCCCCGCCGTGCAGGTGTTGCCCG
 CGTCCCCGCCACCTCCGCCGCGCTCCCGCAACCGCGCTCCGGCCATGGGG
 CCCGTCCGAGCCCCGCAAGGGGCCGACATCCTCGTGAGGCCGCTGAGCGCTGC
 GGCATCGTCGACGTCTCGCCTACCCCGGCCGCTCCATGGAGATCCACCAGG
 10 CGCTGACCGCGCTCGCCGTATCACCAACCACCTCTCCGCCACGAGCAGGGGA
 GGCCTTCGCGCGCTCCGGCTACGCCCGCGCTCCGGCCGCGTCCGCTGC
 GCCACCTCCGGCCGGGGGCCACCAACCTCGTCCTCGCGCTGCCGACGCCCTCC
 TCGACTCCATCCCCATGGTCGCCATCACGGGCCAGGTCCCCGCCGATGATCGG
 15 CACGGACGCCTCCAGGAGACGCCATAGTGGAGGTACCGCGCTCCATACCAA
 GCACAACATACCTGGTCCTTGACGTGGAGGATATCCCCGCCGTATCCAGGAAGCC
 TTCTCCTTGACATCCTCTGGCCGCCGGGGCCGGTGCTAGTTGATATCCCCAAGGA
 CATCCAGCAGCAGATGGCTGTGCCGTGGACACTCCAATGAGTTGCCAGGG
 TACATCGCCCGCCTGCCAACGCCACCATCTACTGAATCGCTTGAGCAGGTCTGC
 GTCTGGTGGCGAGTCACGGGCCAATTCTGTATGTTGGTGGCTGCCGTGC
 20 GTCTGGCGAGGAGTTGCGCCGCTTGTGAGCTACTGGGATTCCAGTTACA
 ACTCTGATGGGCCTTGGCAACTTCCCCAGCGACGACCCACTGTCTGCCATGC
 TTGGGATGCATGGCACTGTGTATGCAAATTATGCACTGAGATAAGGCTGACCTGTT
 GCTCGCATTGGTGTGCCGTTGATGATCGTGTGACTGGGAAAATCGAGGCTTT
 GCAAGCAGGTCCAAGATTGAGCACATTGACATTGACCCAGCTGAGATTGGCAGA
 25 ACAAGCAGCCACATGTCTCCATTGTGCAAGATGTTAAGCTGCTTACAGGGGTT
 GAATGATCTATTAAATGGGAGCAAAGCACAACAGGGCTGGATTGGTCCATGG
 CACAAGGAGTTGGATCAGCAGAAGAGGGAGTTCCCTCTAGGATTCAAGACTTTG
 GCGAGGCCATCCGCCAATATGCTATCCAGGTACTGGATGAGCTGACAAAAG
 GGGAGGCATATTGCCACTGGTGTGGCAGCACCAAGATGTGGCGGCTCAGT
 30 ATTACACTTACAAGCGGCCACGGCAGTGGCTGCTTCGTCGGTTGGGGCAAT
 GGGATTGGTTACCAAGCTGCAGCTGGCGCTGCTGTGGCCAACCCAGGTGTTACA
 GTTGTGACATTGATGGTGTGAGTTCCCTCATGAACAATTCAAGGAGTTGGCGTT
 GATCCGCATTGAGAACCTCCCAGTGAAGGTGATGATATTGAACAACCAGCATCTG
 GGAATGGTGGTGCAGTGGAGGATAGGTTACAAGGCCATGGCGCACACA
 35 TACCTGGCAACCCAGAAAATGAGAGTGAAGATATATCCAGATTGACGATTG
 CTAAAGGATTCAACGTTCCAGCAGTGCAGTGACGAAGAACAGCGAAGTCACTG
 CAGCAATCAAGAACATGCTTGAGACCCAGGGCATACTTGTGGATATCATAGT
 CCCGCATCAGGAGCACGTGCTGCCATGATCCCAAGCGGTGGTGCCTTCAAGGAC
 ATGATCATGGAGGGTGTGGCAGGACCTCGTACTGAAATTTCGACCTACAAGACC
 40 TACAAGTGTGACATCGCAATCAGCATGATGCCCGCGTGTGTATCAACTACTAG
 GGGTTCAACTGTGAGCCATCGCTTCTAGTTGCTTGCATTGATATAAGCTT
 GTATTACTTAGTCCGAACCTGTAGTTGTAGTCTATGTTCTCTTTGTAGGGA
 TGTGCTGTGATAGTGTGACATGCAAGTTCTGCTACATATCAATAATAAGTAC
 TTCCATGNAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
 45 AAAAAAAA

[0452] SEQ ID NO:117

50 TCGCCCAAACCCCTGCCGCCGCCATGGCCGCAGCCACCTCCCCGCCGTGCATT
 CTCGGCGCCGCCGCCGCCATACCCAAACCCGCCAGCCTCTC
 CCGCGCCACCAAGCCCCGCCATCGCGCCGCGCTCCCCGCCGATCGTCAGGTGCT
 GCGCCCGCTCCCCGCCACCTCGCGCCTCCGCCACCGCGCTCCGGCC
 GTGGGGCCCTCCGAGCCCCGCAAGGGGCCGACATCCTCGTGAGGCCGCTGGA

GCGCTGCGGCATCGTCGACGTCTCGCCTACCCCTGGCGGCGTCCATGGAGATC
 CACCAGGCGCTGACGCGCTCGCCAGTCATACCAACCACCTCTCCGCCACGAGC
 AGGGGGAGGCGITCGCGCGTCCGGGTACGCCCGCGTCCGGCCGCGTCCGCG
 TCTGCGTCGCCACCTCCGGCCCCGGGGCACCAACCTCGTCTCCGCGTCGCCGA
 5 CGCTCTCCTCGACTCCATCCCCATGGTCGCCATCACGGGCCAGGTCCCCGCCGC
 ATGATCGGCACGGATGCGTCCAGGAGACGCCATCGTGGAGGTACGCGCTCC
 ATCACCAAGCACAACCTACCTGGTCTTGACGTGGAGGAATATCCCCGCGTCATCC
 AGGAAGCCTTCTCCTCGCATCCTCTGGCCGCCGGGCGGTGCTGGTTGATAT
 10 CCCCAAGGACATCCAGCAGCAGATGGCTGTGCCTGTCTGGGACACGCCGATGAG
 TTTGCCAGGGTACATGCCCGCCTGCCAAGCCACCATCTACTGAATCGCTTGAG
 CAGGTCCCTGCGTCTGGTGGCGAGTCACGGCGCCAATTCTGTATGTTGGTGGTG
 GCTGCGCTGCATCTGGTGGAGGAGTTGCGCCGCTTGTGAGCTCACTGGGATTCC
 AGTTACAACACTCTTATGGGCCTGGCAACTTCCCCAGTGACGACCCACTGTCTC
 15 TGCGCATGCTGGGATGCATGGCACTGTGTATGCAAATTATGCAGTAGATAAGGC
 TGACCTGTTGCTTGCATTGGTGTGCGGTTGATGATCGTGTGACCGGGAAAATC
 GAGGCTTTGCAAGCAGGTCCAAGGATTGAGCACATTGACATTGACCCAGCTGAGA
 TTGGCAGAACAAAGCAGGCCACATGTCTCCATTGTCAGATGTTAAGCTTGTCTTA
 CAGGGGTTGAATGCTCTAAATGGGAGCAAAGCACAACAGGGTCTGGATTTG
 20 GTCCATGGCACAAGGAGTTGGATCAGCAGAAGAGGGAGTTCCCTCTAGGATTCA
 AGACTTTGGTGAGGCCATCCCGCCTGCCAATATGCTATCCAGGTACTGGATGAGCT
 GACAAAAGGGGAGGCATATTGCCACCGGTGTTGGGCAGCATTGACATGAGATGTGGC
 GGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTCTTCATCCGGTTG
 GGTGCAATGGGATTGGTGCAGCTGCACTGGCGCTGCTGTGGCCAACCCAG
 25 GTGTTACAGTTGACATTGATGGGGATGGTAGTTCCCTCATGAACATTAGGA
 GTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAAGGTGATGATATTGAACAAAC
 CAGCATCTGGGAATGGTGGTGCAGTGGGAGGATAGGTTTACAAGGCCAACCGG
 GCGCACACATACCTTGGCAACCCAGAAAATGAGAGTGAGATATATCCAGATTTG
 TGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGACGAAGAAGAGCG
 30 AAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCAGGGCCATACTTGTGG
 ATATCATTGTCCCGCATCAGGAGCACGTGCTGCCATGATCCCAAGCGGTGGTGC
 TTTAAGGACATGATCATGGAGGGTGATGGCAGGACCTCGTACTGAAATTGAC
 CTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATACCTGCGTGTGTTAT
 35 CAACTACTGGGGTTCAACTGTGAACCATGCGTTCTAGTTGCTGTGTTCATTC
 ATATAAGCTTGTGTTACTTAGTCCGAACCGTAGTTGCTAGTCTCTGTTCTCTT
 TTGTTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTGCTCATAATCAAT
 AATAAGCACTCCATGNAANAAAAAAA
 AAAAAAAAAAAAAAAA

[0453] SEQ ID NO:118
 40 TCGCCCAAACCCCTGCCGCCATGGCGNCCACCTCCCCGCCGTCGCATT
 CTCGGCGCCNCCGCCGCCATNCCCAAACCCGCCANCCCTCTCCGCCGC
 CACCAGCCCGNCTCGCGCCGCGCTCCCCGCCGNTCGTCAGGTGNTGCGCG
 CGTCCCCCGCCACCTCCGCCGCCACCGCGCTCCGGCCCTGGG
 CCCGTCCGAGCCCCGCAAGGGCGCCGACATCCTCGTCAAGGCGCTCGAGCGCTGC
 45 GGCATCGTCACGTATTGCTTACCCCGCGCGCTCCATGGAGATCCACCAAGG
 CGCTGACGCGCTCGCCCGTCATACCAACCAACCTCTCCGCCACGAGCAGGGGGA
 GGCCTCGCGCGTCCGGCTACGCCCGCGTCCGGCCGCGTGGCGTCTCGTC
 GCCACCTCCGGCCGGGGCCACCAACCTCGTCTCCGCCGCTGACGCCCTCC
 TCGACTCCATCCCCATGGTCGCCATCACGGGCCAGGTCCCCGCCGATGATCGG
 50 CACGGACGCGTTCCAGGAGACGCCATAGTGGAGGTACCGCGCTCCATCACCAA
 GCACAACCTGGTCCTGACGTGGAGGAATATCCCCGCCGTACCCAGGAAGCC
 TTCTCCTCGCGTCCTCTGGCCGCCGGGGCGGTGCTGGTTGATATCCCCAAGG
 ATATCCAGCAGCAGATGGCGTGCCTACTGGGACACGCCGATGAGTTGCCAGG
 GTACATCGCCCGCTGCCAAGCCACCATCTACTGAATCGCTTGAGCAGGTCTG

5 CGTCTGGTTGGCGAGTCACGGCGCCAATTCTGTATGTTGGTGGCTGCGCTG
 CATCCGGCGAGGAGTTGCGCCGCTTGTGAGCTCACTGGGATTCCGGTTACAAC
 TACTCTGATGGGCCTGGCAACTCCCCAGCGACGACCCACTGTCTCTGCGCATG
 CTTGGGATGCATGGCACTGTGTATGCAAATTATGCAGTCGATAAGGCTGACCTGT
 TGCTTGATTTGGTGTGCGGTTGATGATCGCGTGAUTGGAAAATCGAGGCC
 TGCAAGCAGGCTCAAGATTGAGCACATTGACATTGACCCAGCTGAGATTGGCAG
 AACAAAGCAGCCACATGTCTCCATTGTGCAGATGTTAAGCTTGCTTACAGGGGT
 TGAATGCTCTATTAAATGGGAGCAAAGCACAAACAGGGTCTGGATTTGGTCCATG
 GCACAAGGAGTTGGATCAGCAGAAGAGGGAGTTCCCTCTAGGATTCAAGACTTT
 10 GGCAGGGCCATCCCCGCGCAATATGCTATCCAGGTACTGGATGAGCTGACAAAAA
 GGGGAGGCGATCATTGCTACTGGTGTGGGAGCACCAGATGTGGCGGCC
 TATTACACTTACAAGCGGCCACGGCAGTGGCTGTCTCGTCTGGTTGGGGCAA
 TGGGATTTGGGTTACCAAGCTGCAGCTGGCGCTGTTGGCCAACCCAGGTGTTAC
 AGTTGTTGACATTGATGGAGATGGTAGTTCCCTCATGAACATTCAAGGAGTTGGCA
 15 TTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATGATATTGAACAACCAGCATC
 TGGGAATGGTGGTGCATGGGAGGATAGGTTTACAAGGCCATCGGGCGCACA
 CATACTTGGCAACCCAGAAAATGAGAGTGAGATATATCCAGATTTGTGACGAT
 TGCTAAAGGATTCAACGTTCCGGCAGTCGTGACGAAGAAGAGCGAAGTCAC
 TGCAGCAATCAAGAAGATGCTGAGACCCCAGGGCCACTTGTGGATATCATC
 20 GTCCCGCATCAGGAGCACGTGCTGCCTATGATCCAAGCGGTGGTGGCTTCAAGG
 ACATGATCATGGAGGGTGTGGCAGGACCTCGTACTGAAATTGACCTACAAG
 ACCTACAAGTGTGACATGCGCAATCAGCATGGTGGCGTGTGTTATCAACTAC
 TAGGGGTTCAACTGTGAACCATGCGTTCTAGTTGCTGTTCATTCATAAAG
 CTTGTGTTACTTAGTCGAACCCCTGTAGCTTGTAGTCTATGCTCTTTGTTAGG
 25 GATGTGCTGTCATAAGATATCATGCAAGTTCTGCTCCTACATATCAATAAAGT
 ACTTCCATGNAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
 AAAAAAAAAA

30 [0454] SEQ ID NO:119
 CGTTGTGCCCTGGCAGTCTCAGGTTGAGCCCTCACCATTGAAGTAGCATGGTCA
 TTGGATTGACCCGATTGACGGCGGATCTATTGGATCTTCCCTTGTGTCGTTTA
 TACTGGTATAGATGTTAACACATATTGGAAAATATATTCAAAACATGTTCTAT
 35 AAAAAAGTTAAACTATACATGTATAATGGAAGTCATTATAAGAAATGTTTAC
 ATGTATAAAAGATGTACATCATATGTGAAAAGTAGACATGTGTTAGAAAAAAT
 AAACAAACAAATACATAAAAGAAAATCAAAGAAAAACACCCAAAAACCA
 AAGAAAATAAAGAAGAAGAAGAAAAAGAGAAAAACATTGAAAATCAAAGAAG
 AAAAAAAACATAAAGAAAAGAAAACCGAAAAATACTGGCAAAACACACAAAAA
 ATGAAAAGAAAAATAAAGAAAACCGGACTTACCAATCGAACGGAGCGATCGG
 40 ACACGAATGAGCGAAGGCATGCATCGAGCAACACCGCTAATTGACCGGCCGTA
 GTCGTTGCCGTAGACCATTATAAGAATCGGTATCGGAGAGACATAGGGGTT
 TTTGGTTCTAACCATATCTGTCACACTTACCATACATCACCTAGTC
 GATCAAATTAGGTGAGTATTGTTCTAGCCACATCTAAGGCAAGATTGTTTC
 TGAGCAGTGAACCCATATGTATAGACAGAAAAATTGTGAAAAGATTCTCTTAC
 45 ACGGTCAAAGCGTGGTTAACATTAACTCAAGTAAGATAATGCGATAA
 ATGTGACAAAAATAATGTGTTAGAAGTATGACAAAATAATCACAATCCAAA
 CAGTCTGATAGCTGGCGAGTGCAAAATAGATACGAAATCTCTGGTATATCACA
 CGGGTCCAAAATAATTGCTGTTGAGCATCAGCCTTCTGCACAAAAAGCTA
 GCCCAAACAAACGAGTGGCGTCCCATCTGAACCACACGCTACCCGCCGCGTGA
 50 CAGCGCCAAAGACAAAACCATACCCCTCCCCATTCCAACCCCTCTCCGCC
 ACAGAAATCTCTCCCTCGCCCAAACCCCTCGCCGCCATGGCCGCCACCT
 CCCCCGCCGTCGATTCTCCGGCGCCGCCGCCGCCATGCCAAGGCC
 CGCCCCGCCAGCCTCTCCCGCGCACCAGCCCCGCTCGCGCCGCGCTCCCCGCC
 CGCGTCGTCAGGTGCTGCGCCGCGCCCCCGCTGCTGCCACCTCCGCCGCC

5 CCGCCACCGCGCTCCGCCCTGGGCCCGTCCGAGCCCCGCAAGGGCGCCGACA
 TCCTCGTCGAGCGCTCGAGCGCTCGGCCATCGACGTATTGCCCTACCCCG
 CGGCGCGTCCATGGAGATCCACCAAGGGCGCTGACCGCCTCGCCCGTCA
 CACCTCTCCGCCACGAGCAGGGGAGGGCGTCCGCCGCTACGCCCGCG
 CGTCCGGCCCGCGTCCGCGTCTCGCTGCCACCTCCGCCGCGGACCAACCT
 CGTCTCCCGCGCTCGACTGCCCTCGACTCCATCCCCATGGTCGCCATCACGG
 GCCAGGTCCCCCGCCGATGATCGGCACGGACCGCTTCCAGGAGACGCCATAG
 TGGAGGTACCGCGCTCCATACCAAGCACAACCTACCTGGTCCTGACGTGGAGGA
 TATCCCCCGCGTCATCCAGGAAGCCTTCTCGCTCTGGCCGCCGGGC
 10 CGGTGCTGGTTGATATCCCCAAGGATATCCAGCAGCAGATGGCCGTGCCTATCTG
 GGACACGCCGATGAGTTGCCAGGGTACATCGTCCCCTGCCAAGCCACCATC
 TACTGAATCGCTTGAGCAGGTCTCGCTGGTTGGYAGTCACGGCGCCAAATT
 CTGTATGTTGGTGGCTCGCTGCATCCGGCGAGGAGTTGCCGCTTGTG
 AGCTCACTGGGATTCCGGTTACAACACTACTCTGATGGGCCCTGGCAACTTCCCAG
 15 CGACGACCCACTGTCTCGCATGCTTGGGATGCATGGCACTGTGATGCAAAT
 TATGCAGTCGATAAGGCTGACCTGTTGCTGCATTGGTGTGCGGTTGATGATCG
 CGTACTGGAAAATCGAGGCCATTGCAAGCAGGTCCAAGATTGTGCACATTGAC
 ATTGACCCAGCTGAGATTGGCAAGAACAGCAGCCACATGTCTCCATTGTGAG
 ATGTTAAGCTGCTTACAGGGGTTGAATGCTCTAAATGGGAGCAAAGCACA
 20 ACAGGGTCTGGATTGGCCATGGCACAAGGAGTTGGATCAGCAGAAGAGGGA
 GTTCCCTCTAGGATTCAAGACTTTGGCGAGGCCATCCGCCGCAATATGCTATCC
 AGGTACTGGATGAGCTGACAAAAGGGGAGGCCATATTGCTACTGGTGTGGC
 AGCACCAAGATGTGGCGCTCAGTATTACACTACAAGCGGCCACGGCAGTGGC
 TGTCTCGTCTGGTTGGGCAATGGGATTGGGTTACAGCTGCAGCTGGC
 25 TGCTGTGGCCAACCCAGGTGTTACAGTTGTTGACATTGATGGAGATGGTAGTTTC
 CTCATGAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGG
 TGATGATATTGAACAACCAGCATCTGGGAATGGTGGTGAATGGGAGGATAGGT
 TTTACAAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTG
 AGATATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTTG
 30 TGTGACGAAGAACAGAGCGAAGTCACTGCAGCAATCAAGAACAGATGCTTGAGACCC
 AGGCCATACTTGGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCATAG
 ATCCCAGCGGTGGTGCCTTCAAGGACATGATCATGGAGGGTGATGGCAGGACC
 TCGTACTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCAT
 GGTGCCCGCTGTTGATCAACTACTAGGGGTTCAACTGTGAACCATGCGTTTCT
 35 AGTTGCTGTTTCATTATATAAGCTTGTGTTACTTAGTCCGAACCCCTGTAGCT
 TTGTAGTCTATGCTCTTGTAGGGATGTGCTGTCAAGATATCATGCAAGTT
 TCTTGTCCCTACATATCAATAAAAGTACTTCCATGGAATAATTCTCAGTTCTGTT
 TGAATTGACATCTCTACAAACAGTGTGCTGGTCTTCTGTTACTTACATGT
 CTGCCGTGCCGGTTATGACATAATGACCGATGGAGGGTGGTCAGCAGGTTTAG
 40 ACGGGGAGTTGAAACTTTTGGGGGAAGAAATCTGAATACAGTTGGGAGG
 AAAGATAAAAGCATATACCTTGATTAATTATTGAGCCAAATATCCAGCCTAATT
 TATCAAGCAATAGGCAGTGTAGGGTGTGGCATTCTCTTCCCTGAGATCTGGT
 GTCGGGACCCGATTCTAAGTCACACCGATCTAGCATGTAACACCTCATATCACT
 TTGCCGCTCACGCACGGTATCCTCACGGGTGTCGCCCTACCATGGCCCAGGACC
 45 GTTGCCTTGGCTACGTATATGATGGTGTGCTAGYATCCATATGACAGA
 GAACCCGGCCGACATRGCTAGTCGTGAACCCAAAGCGGCACAGACCTATGGAG
 ACAGGCATACATGAATCACATCGAGCATGTCGGTCAACAGCGTATGAATCCGGG
 CTGTAGCACTGGCTAACAGGACTCCGGGAACCCGGCTGAGCAGGCTAGGC
 AGGACTCCGGAAGTCACCGCGTGACATTCCCCAAGGGACAGACATAGGAACG
 50 AAGTGGAACACATGCCGCCAGTCAAGTGTGAGCAGTAGTGCTGGCTAGC
 AGGACTCCGGTGAACCGGGCTGAGCGGACTACTATGGCTCGAGGTAGCACTAG
 ACTACATTCCCCATAAGAGAGGGCTKCCAAGGATAAGCAACTAGATTGTCGGRTC
 YCRSRYWTTGTCTCCGTGTGTTATTGTTGATGCAAGTATGTGTTGACAAAC
 ATGGCATCACAAACATAACGCAAACATAGATAGGCTCAGAGAGGCCACAT

AGCATTAAATACGAACAGGGTCACATGACCCATCATTAGCAGCATACAGCATGAA
GCATCATGTCTGAGTACAGACACTAC

5 [0455] SEQ ID NO:120
CTGAAAATTCAATATGCCCTCGGGCACCAATGCTCTGCTTCCAATTTCTATAAT
TCCCATTGTAAAAAACACACCACAAAAATCACACTGTAGTAATCTACATGTTG
TTGAGCCTATAAAATCTTCATAAAAATTGAGATTAATGCGGTTGTGCAAAAT
ATGGGGTTGGTCATGTTCTACATATTCTATTGCAATTGTTAACCTTGAGATGGTA
10 TATTTTGTCATAATGCATATCTCATTGTTATTATTTAACCTTGAGATGGTA
ACGAAGATCCAAACATGCATAGATGATTCTCCGGATGATTTTGAGCCTGCAC
TAGGAACTCCCAAGAGGCCAGAAGGTTGGGTTGTACAAGATAACATTGTTGAA
CACACTCATAACCTGCATGTGACATACATGACGTAACCTATAGTGTGATTGAC
AAATGTCCTTGTCCAATTGTTATATATCCCGTGGCAACGCACGGGCATTGCA
15 CTAGTATATGTAAAGATATCAATGTCAGGAGTCCCCATGGTCGTTGCGCTTGTCC
ACTACCGGCTCGCTAGAGGCAGCTCACCTAGAAGTCGCTACGAGCAATACATA
GTCGTTCTGGGCGCAGCTATGTTCTGCCTTTGCGACGCTCAGGCACGGCTTGCCT
ACAGCCTGAGGGTCGGCTAGGAACCACTAATTGTCATGCTGATGTCACAATG
ACATCATGCATATTTCATTTCGCTTCTCTTTAATTTATTGTTATTTC
20 AAAATATTATATATTGAAATTTCATGTTGATTGAAAAATGTTAAA
CCTGTATAGAGAAAAATATTGATATATAAAAGTATATAACATGAATGAAA
AATGTATAAATGTTAATTATGTTACAAAAATGTTGATAACAATTAGCAGTC
ACATATTCAAAATAATGTTGGAATTAAAAATATGTTAAGTTAA
AAAAAATGTTCATGTAATGTTGTTAAATGTTGATAACATTCAATAAAATATG
25 TCACATTGAATAATTCTCTCAAGCTTAACAAATGCGCTCATTATATTATCAAAA
ATTGTCGTCAGTGTACACAAATGTTATGTTAGTCACAAAAAAATGTTTTTCAG
AAAAAATATATTGATCATGTTAATTTAATTTAACTGTTAATATATATTAGAAA
ATATATTCAAAACATGTTCTGAAAAAGTTAAAATACATGTATAATGTAAG
TCATTATAATAATGTTTACATGTTAATTTAACTATGACAACATATGTCAGG
30 TAGACATGTTGAAAAAATAACAAATAACTAAATAAAAAAGAAAATCAAAGAA
AAACACCAAAACCAAGAAATAAAATAAAACCAAGTATAAGAAGARAAAG
GAGAAAAAACATTGAAAATCAAAGARAAAACATAAAGAAGAAAAACCGAA
AAAAACTAGCAAAACACACACACAAAAAGAAAATGAAAAGAAATAATAAA
GAAAGCCGGACTGAACCGATCAAACGCAGCGATCGAACATGGATGAGCTAAGGC
35 ATGCATCGAACACACGGCTAATTGGCCGGCCGTAGTCGTTGCCGTAGACCA
TTCCTACGAATCGGTACCGGAGAGACA TAGGGGCTGTATGGTCTTAACCATA
TTGCCACACTTGTACACCTCATCTTAGGCAAATTAAATCAAGTTATGTTAGGT
TTGGTTTAGCCACATCTAAGGCAAGATTATTTCTGAGCAGTGAACCCATAT
GTTATAGACATAAAAGTGTGGGAAGATTCCCTTACTGAAACTGTGGCTAACAA
40 TTTATTAAGAATTAACCTAAGTAAGATAGGTGCAACAAATGTAGCAAAAATAATG
TGGTATATATAGCAAAGATAGCCACAACCGCGAGTGGAAATACCAAGATACGAGA
TCTCTGGTCATATCACACGAGTCAAATTAAATTGCTTGTGAGGTTGAGCCTT
TGCATAAAAAGCTAGCCAAACAAACGAGTGGCGTCCCATCTGAACCAACAC
TCACCCGCCCGTGACAGGCCAAAGACAAAACCATCACCCCTCCCAATTCAA
45 CCCTCTCTGCCACAGAAATCTCTCCCTGCCAAACCCCTGCCGCCATG
GCCGAGCCACCTCCCCGCCGTGCAATTCTCGGGCGCCGCCGCCGCCCG
CCATACCCAAACCCGCCGCCAGCCTCTCCCGCCACCAGCCCCCTCGCGCC
CGCGCTCCCCGCCGATCGTCAGGTGCTGCGCCGTCCTCCGCCACCTCC
GTCGCGCCTCCCCGCCACCGCGCTCCGGCGTGGGGCCCTCGAGCCCCGCAAGG
50 GCGCCGACATCCTCGTCAGGGCGCTGGAGCGCTGCGGACATCGTCGACGTCTCGC
CTACCCCTGGCGCGCGTCCATGGAGATCCACCAAGGGCGCTGACGCGCTCGCCAGTC
ATCACCAACCACCTCTCCGCCACGAGCAGGGGAGGGCGTTCGCGCGTCCGGGT
ACGCCCGCGCGTCCGGCGCTGGCGTCTGCGCCACCTCCGGCCGGGG
CACCAACCTCGTCTCCCGCGCTGCCGACGCTCCTCGACTCCATCCCCATGGTCG

5 CCATCACGGGCCAGGTCCCCGCCATGATCGGCACGGATGCGTCCAGGAGA
 CGCCCATCGTGGAGGTACCGCGCTCCATCCAAGCACAACACTACCTGGTCCTGA
 CGTGGAGGATATCCCCCGTCATCCAGGAAGCCTCTTCCTCGCATCCTCTGGC
 CGCCCGGGCCGGTCTGGTGTATCCCCAAGGACATCCAGCAGCAGATGGCT
 10 GTGCCTGTCTGGGACACGCCGATGAGTTGCCAGGGTACATGCCCGCTGCCA
 AGCCACCATCTACTGAATCGCTGAGCAGGTCTCGCTGGTGGCGAGTCACG
 GCGCCAATTCTGTATGTTGGTGGCTGCCTGCATCTGGTGGAGGAGTTGCGC
 CGCTTGTGAGCTACTGGATTCCAGTTACAACACTACTTATGGCCTGGCAA
 CTTCCCCAGTGACGACCCACTGTCTCGCATGCTGGGATGCATGGCACTGTG
 15 TATGCAAATTATGCAGTAGATAAGGCTGACCTGTTGCTTCATTGGTGTGCGGT
 TTGATGATCGTGTGACCGGGAAAATCGAGGCTTGCAAGCAGGTCCAAGATTGT
 GCACATTGACATTGACCCAGCTGAGATTGGCAAGAACAAAGCAGGCCACATGTCTCC
 ATTGTCAGATGTTAAGCTTGTACAGGGTTGAATGCTCTATTAAATGGGA
 GCAAAGCACAACAGGGCTGGATTTGGCATGGCACAAGGAGTTGGATCAGC
 20 AGAAGAGGGAGTTCCCTAGGATTCAAGACTTTGGTGGCCATCCCAGCGCA
 ATATGCTATCCAGGTACTGGATGAGCTGACAAAAGGGGAGGCGATCATGCCAC
 CGGTGTTGGCAGCATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCA
 CGGCAGTGGCTGTCTCATCCGGTTGGTGAATGGATTGGTTGCCAGCTG
 CAGCTGGCGCTGCTGTGGCCAACCCAGGTGTTACAGTTGTTGACATTGATGGGGA
 25 TGGTAGTTCCATGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTC
 CCAGTGAAGGTGATGATATTGAACAAACCAGCATCTGGGAATGGTGGTGCAGTGG
 GAGGATAGGTTTACAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAA
 AATGAGAGTGGAGATATCCAGATTGACATTGCTAAAGGATTCAACGTT
 CGGCAGTTCGTGTGACGAAGAAGAGCGAAGTCAGTCAGCAATCAAGAAGATGC
 30 TTGAGACCCCAGGGCCATACTTGTGGATATCATTGTCCTGCATCAGGAGCACGT
 GCTGCCTATGATCCCAAGCGGTGGTCTTAAAGGACATGATCATGGAGGGTGAT
 GGCAGGACCTCGTACTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGC
 AATCAGCATGATACTCGTGTGTTACTCATATAAGCTTGTGTTACTTAGTTCCGAAC
 TGCGTTTCTAGTTGCTTGTGTTACTCATATAAGCTTGTGTTACTTAGTTCCGAAC
 35 CGTGTAGTTTGTAGTCTCTGTCTCTTTGTAGGGATGTGCTGTACATAAGATATC
 ATGCAAGTTCTGTCTACATATCAATAATAAGCACTCCATGGAATAATTCTCA
 GTTCTGTTGAATTCACATCTTCACGAACAGTGTGCTGGTCTTCTGTTAC
 TTTACATGCCCTGCCGTGTCAGGTTATGACATAACGACCGATGGAGGATTGGAGGG
 TGGTCGGCTGGTTAGACGGGAATTGAAACATTTCTGGAAGAAATCTGAAT
 40 ACAGTTGGAGGGAAATGGAAGCATATATTATCGAGCCGCTATCCAGGCTA
 ATTATCAAGCACTAGACAGTGTAGGGTGTGGCATTCTCTCTTCTGATATCC
 GGCTTGAGAGGAGAGATTGAGGCTTCGGCTGTGTTGGTGTGATTTCTACAGCA
 TTTGAGAGAGAGAGAGAGATGTTGCAACTGTGTTGTCTGGTGTGTTACA
 GAGAAAGAGAGATGACATTAGAGATATGCAGATCGITTACAGTTGTGCTGCGTT
 45 ATTCTGACTGATTGTTATTGTTGCTATCATGTGCAAATTGTTGATGGAAAA
 TCAACAAAATTGATATTGCAAAGCGAGTTGGATTGAATGATTGAGAAATG
 GTGACTTGTGAGTGGCTTGAGAAATTGGTGTTCATAGGTGTGCAAGTTGGTAAT
 GAAAGGCGGGCTTGAAATTCCGAAAGGCAGGCAATGATACTTCTGAAAGT
 GATGTTTCTTCCAGGTTCCGGTGGAAACAAGTCTACGTTGAGCCAATGTTGT
 50 CAGCTTATTCTGCTTTAGTTCAAGTTGTTGTTGTCACAGATTGCTGGCAGAGC
 CCCATGATCGGCTGAGCCTCCAGGAGATCCTGATTGCTCGACTGCGGATACGTT
 GAATCCTTAAAATACTATAAGCTCCCTAGTTAGTTAGAGAACTGAGAATC
 AATTGAGGGCAACATTAGTCGATTGGCTTCCGATTGACTGGGTCGCCCTCC
 GGGTCCTACAGTTGTGGCCCTATATGTAAGTGCCCAAGTGTGTTGGCTT
 CTGGTCTTCTGATGAAAGCGCGTGGTGGCTGGGGCTTAGAATATTTCATT
 GATTAACTAAAACAATCAGATCCCTTTCCCTGCTTCATGTGTGTTGACCAATC
 TTTTTAAAATTCTTGATTGTTATATTGATGGAGTAAATCTGGCTGTGCAA
 CGGTAGTCCATTGAAACCTGGAAATCGAAATCATTGACTGCAAGGTCTGTTGCC
 TGTTAGTTGTTCTTATATAAGATCTTGACAGTTATGAATTGCTTTGGAATT

GTATAAAGTTCACAGATAGACAGGCCCTGTTAAATACGTTCGTGCATTAA
 GTGTAAACATATCTGCCAGTGTATTTCTCGGCTCGCATTAGTACGCATAAATT
 TAGCACTCTCTGAATTCTCATATGCAGACCACTATGAAAAAAACGACATGC
 AAGTAAATAAAACGATTTCAGGTTCAATTAGTAGCAAACCGTTTATGTCCTTA
 5 AAAATCAATTAGCAGAGCCACTCCATTACCGGTCAGCAGAAAAGAAGCATGTG
 TGTGTTTGGGCTATCATAGAGCTAAATAATTGATTCCCATCTGTAATGTTCA
 TCGTTGTTACATCAGTGTGGCTGTCGTGGCTGGAGACTAGCCTGTCAGA
 CAATATGTTGACAAGAGTGTGTTGTGAGATCGGGATGCGGTGCTGCATCT
 GTACTTGTGTTGTGAATACCAGTTAGATGATCAGTTTGTCACCTCTGCCAT
 10 GAATGGCTGTTAAATTGTCACTTTGTAGGAACCTGTTGCCGTAAATATCAATTAAAT
 AATCAATTGTCATGGTATATCAATTAGATGGTCATTGTTCTAGTAGAGAT
 GTCTATACATGCCAATGCAATGTTAGAGTTCAAGGTCTCGACGGCGCGCA
 AAGCGCGTCCTATGCTCTAGTTAAGATGACAACAAACACGACCCAAGTGTAT
 GCTATGCTCATCCGGTGGCTTGTGATGTTCAATGGCGTGTCTCCATGGCA
 15 TCGACGGCGACAATGTTATCTCTTCAACTGTCGTATATGCTCATTGGCATT
 TGAAACTTGCAAGCAAGGTGATAACTGGTCTGGGATGTTGACGCCCTATG
 TATCTAGATTAGGGTGTGCTCCGCCAGTATTGGAACGATTATCAACATTG
 CGGCTGGTACTATTGTCGATGTTTAATGTATTGTCGATGACTCAATTCTACGTC
 20 TGAACATTTCATTGAGCCAAGAGGCAGAACACAGGTACATGTAACCGCCAGT
 GAAAAAGGTTCAAAGAAGAAAAAGATACGAACGACAGCGAGTTGTATKKCAGT
 TTTGAACTAAGAGTAACACGGAGTRCAGTAGTACGATCCTGTTGMYTCTGTA
 TTTGGWTAKTTTCCGGAGTTGAGTATTGWAACTTCTTGCTTTAA
 CATTAGTACAGATGCAAGTGTCTACATACAGCGCTTTGATTGTAACAATT
 25 ATGAAAGACGTAGTAATTGTTGCAAGTCAATAAGCTAGCCATCGTGTGGT
 TTCCCAAGAAAAAGATATTCACTATAGATTCACTACATCTCTAAACAAACTACA
 CTGTAGATTCACTACAGACCAACAGAACATTGTCAGTGGATAAAACTGACTCTAT
 CTTTGTGTAACCTCTGCAACGAACCTGGATAAAGATGGAGGCCAGTCCGTTCTGG
 30 TTACTAGGAGTATCCATTCTGAAGACCATGGAGCAACCACGGCGATGGCG
 ATCGGCAGCCTCCCAGCCGGCGACCATGGCGGATGCCACGAGCGCAGGAGCGAC
 GCCTCTCCCTGGCTCCTCGACGACATCGTAATCTGAGATCCTGTCCTGGCG
 TCGCCCCCAAGCCATCCTCCGCTGCCGCCGTACGCCGTGCTGGCGCC
 ACCACCTCCACCCCGACTTCCTCCTCGCCACACGCCGCCAGCCCCGCCCT
 35 CATCACCTCCGGCACAGTT

[0456] SEQ ID NO:121

40 AAATTITATAATATTGTTTCAAATTATGTTAAACTCATTGTTCAATT
 TTTGTAATATATTAAATCCATTGATAGATTGAAAATATAATAATTGTTCAA
 AAACATTCTATAATTCTAAACCTTTAACATTCAAGAATAAGATTAGGAAA
 TTTGATTCTAAAATATATTAAATCTGCAACTACATTATATAATTACA
 TGAGCCAATTATTGGTAGAAATCAACTGAAAAACAAAAGAAAAATTGGA
 ATAGCGGGAGTTCTCGCGAACCTGGGGGGGGGACAACCCCTATCA
 45 ATGAGCTAGGGATTCTTACATCTCGCTACAGCCGACTAGTTTCTYCCCA
 TTTGTTTATATCGGTTTTACTACTTTGCACTGTTCTGTTTGTATTATTCA
 TTTTCTCTATACTTCTGTTCTTCTGTTTCAACGATTACTTGTGTTTCACGTT
 TCTACAGTTCTGTTCTTCTGTTTCAACGATTACTTGTGTTTCACGTT
 TTTAAATTAAATTAAATCTCAGATAACATTACATTAAATTATATAC
 TTTATGTCAAGTTTCATACACATTGTGATTTATACATTAGGATTCTAA
 ATACATGATTAATATTATTCACTAGACATAGAGTACTGTTGAACACTTTCAA
 ATACATGTTGAAATAATTATTATGATATGAAATATGTTTTATTATGCAA
 CATTTTATACACTTATGTTTGTGAAATATTACAAAATTGCTGAAACGTG
 TGAACATTTTAAAATGTAACATAATTGTAATGGTATGAAACTTTGAAC

5 TCGCGAACATTATTTACATTGTATATTATTTGATTCACTTCTGTAAAGTTATC
 GCCTGAATTGCTGAAAAACGTGATTTTTAAATGCCACATATATTGTTTTGA
 ATGGTTCATGCATTTCTGAAAGTTGATCGAACATGTTTATATTGCACTTTAA
 AATGTAATAACCACTTGAAAATTAACTAATGTATTTCATAATATATGTATTAA
 10 ATATTATTAAAAAATAAAAAAAGGTAAAAGAAAAACAGATCAACCGCGATGAG
 ACCCATGGTTGCGCTTGTCCACTACCGGCTCACTGAAGACGTCTCACAGT
 AGGAGTCGCTACGAAGAATACATAGTCGCGCTGGCGCGGTTATGTTCCGCCTGT
 TCGACGCCAAGCATGGCTGCCTACAGCTAGAGGGTCGGCTAGGAACCACT
 AATTGTCATGCTGATGTACAATGACATCATACATGCTTTATTTAATTTC
 15 GCTTTCTCTTAAATTGTTGATTCAAAATATTCTGTTTTAAGAATGCTAG
 TATTGATTGAAAAATGTTAACCTGTATAGAAAAATATATAACATGAATGAAA
 AATGTATAGATGTTAACATGTGTACAAAAATGATTGACAATTAAGAATGTC
 ACATATTCAAAATAATGTATGTGGATTGAAAAATGTGTATATAATT
 20 AATGGTCATGTAATTAAAAAAATGTGTGATACATTCAACAAAAATATTTCAC
 ATTGAAATAATTCTTCTGAGCTTAAGAAATGTGTTATTGTTCAATT
 GTACAGTGTACAAAATGTTACATAGTTCAAAAAATGTTTCAGTAAAATTA
 CATTTCATTGTTGTTAACACACATTGGAAAAATATATTGAAACAT
 GTTTTGAAAAAAATTAAAACATGCTTGTACTCCCTCCGTCCGAAAAAGG
 TTTACATGTATAAAAGTTTTCGGAGGGAGGGATTATAATGTTAGTCATTATAA
 25 GAAATGTTTACATGTATGAAATGTATAGCATATGTGTAAAAGTAGACATGTGT
 TGAAAAAAAGTAAAACAACCCAAAAACCAATGAAAATAAAATAAAACCA
 AAGTACCAAGAAGAAGAAAAGGAGAATAAACATTGAAAACAAAGAAAATAA
 AAAACATAAAGAAGAAAGAAACCCAAAGAAAATGGCAAAATTAGACACAGA
 AAAGAAAAACGAAAAATATAATAAARAACCGGACTGAACCGATCGGAC
 30 ACGGATGAGCGAAGGCATGCATCGAGCAACACAGCTAATTGGCCGCCATAGT
 CGTCGCCCGCAGACCATTACAGAACATCGGTACCGGAGAGACATAGGGCTATT
 TGGTTTGTAGCCACATTGTCATACTTGTGACACCGCATCTATGCAAGTTGA
 CCAAATTAGGTGGATGTTAGTTCTAACACATGTAAGGAAAGATT
 AGCATTGAACCCGTAGACACAAAAAGTGTAGGAAGATTACTTAAACAAGCTAA
 35 AGTGTGGCTAACATTAAAGCATCTCAGGTAAGATAAGTGCACAAATATGGCA
 AAAATAATGTGGTATATGACAAAGATAGTCACAATCCAAACAGCCCATAGCC
 TGGCGAGTGCAAATAGATACGAGATCTCTGGTATACACAACCGTCAAATTAA
 TTGCTTGTTCAGCATCAGCCTTTGCATAAAGAAGCTAGCCAATCTGAACCAC
 ACACCTACCCGCCGTGACAGGCCAAAGACAAAAACATCACCCCTCCCCAATT
 40 CCAACCTCTCTGCCACAGAAATCTCCCCCTGCCAACCCCTGCCGCC
 CCATGGCCGCCACCTCCCCGCCGTGACATCTCGGGGCCACGCCGCC
 CATGCCAACCCGCCATCCTCTCCGCCACCAGCCCCTCGGCC
 GCGCTCCCCGCCGTGTCAGGTGTTGCCGTCCCCGCCACCTCCG
 CCGCCCTCCGCAACCGCGCTCCGCCCTGGGGCCCGTCCGAGCCCCGCAAGGG
 45 CGCCGACATCCTCGTCAGGCCTCGAGCGCTCGAGCGCTGCCATCGTCGACGTCTCGCC
 TACCCCGCGGCCCTCCATGGAGATCCACCAAGCGCTGACCGCGCTCGCCCGTCA
 TCACCAACCACCTCTCCGCCACGAGCAGGGGAGGCCTCGCGCGTCCGGCTA
 CGCCCGCGCTCCGCCGTGCGCTGCCACCTCCGCCGGGGGCC
 ACCAACCTCGTCTCCGCCGTGCGCCACGCCCCCTCTCGACTCCATCCCCATGGTCGC
 50 CATCACGGGCCAGGTCCCCGCCGTGACATCGGCACGGACCGCTCCAGGAGAC
 GCCCATAGTGGAGGTACCGCGCTCCATACCAAGCACAACACTACCTGGTCCTTGAC
 GTGGAGGATATCCCCCGCTCATCCAGGAAGCCTTCTTGCATCCTCTGGCC
 GCCCGGGGCCGGTGCAGGTGATATCCCCAAGGACATCCAGCAGCAGATGGCTGT
 GCCCGTCTGGGACACTCCAATGAGTTGCCAGGGTACATGCCCGCTGCCAAG
 CCACCATCTACTGAATCGCTTGAGCAGGTCTCGCTGGTGGCGAGTCACGGC
 GCCCAATTCTGTATGTTGGTGGCTGCGCTCGCTGGCGAGGAGTTGCGCC
 CTTTGTGAGCTTACTGGGATTCCAGTTACAACACTACTCTGATGGCCTGGCAACT
 TCCCCAGCGACGACCCACTGTCTCGCATGCTGGGATGCATGGCACTGTGTA
 TGCAAATTATGCAGTAGATAAGGCTGACCTGTTGCTCGATTGGTGTGCGGTTT

GATGATCGTGTGACTGGGAAAATCGAGGGCTTGCAAGCAGGTCCAAGATTGTGC
 ACATTGACATTGACCCAGCTGAGATTGGCAAGAACACAAGCAGCCACATGTCTCCAT
 TTGTGCAGATGTTAACGTTGCTTACAGGGGTTGAATGATCTATTAAATGGGAGC
 AAAGCACAACAGGGTCTGGATTGGTCCATGGCACAAGGAGTTGGATCAGCAG
 5 AAGAGGGAGTTCCCTAGGATTCAAGACTTTGGCGAGGCCATCCGCCAAT
 ATGCTATCCAGGTACTGGATGAGCTGACAAAAGGGGAGGCATATTGCCACTG
 GTGTTGGGCAGCACCAGATGTGGCGGCTCAGTATTACACTACAAGCGGCCACG
 GCAGTGGCTGCTCGTGGTCACTGGGCAATGGGATTGGTACAGCTGCA
 GCTGGCGCTGCTGTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTATG
 10 GTAGTTCCATGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCC
 AGTGAAGGTGATGATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGA
 GGATAGTTTACAAGGCCAATCGGGCGCACACATACCTTGGCAACCCAGAAAAA
 TGAGAGTGAGATATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCA
 15 GCAGTCGAGTGACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTT
 GAGACCCCAGGGCCATACTTGGATATCATAGTCCGCATCAGGAGCACGTGC
 TGCCTATGATCCCAAGCGGTGGTCTTCAGGACATGATCATGGAGGGTATGG
 CAGGACCTCGTACTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAA
 TCAGCATGATGCCGCGTGTGTTCAACTACTAGGGGTTCAACTGTGAGCCATG
 CGTTTCTAGTTGCTTGTTCATTGATAAGCTGTATTACTTAGTTCCGAACCC
 20 TGTAGTTTGTAGTCTATGTTCTTTGTAGGGATGTGCTGTCATAAGATGTCAT
 GCAAGTTCTGCTACATATCAATAATAAGTACTTCCATGGAATAATTCTCAGT
 TCTGTTGAATTTCGATCTTCACAAACAGTGTGCTGGTCTTCTGTTACTT
 TACATGTCGCTGTGTCAGGTTCTGACATAACGACCGATGGAGGGTGGTGGCAG
 GTTTAGAAGGGGAATTGAAACITTTGGAGAAGAAGTCTGAATACAGTTGGG
 25 AGGAAAAATAGAAGTATATACTCGATTAATTCAAGCCGCTATCCAGTCTA
 ATTATCAAGCACTAGACAGTGTAGGGTGTGGCATTCTCTCTTCTGAGATCC
 GGCTTGAGAGGGAGAGACCGAGGGCTTCGGCTGTGTTGGTGTGATTCTACAGC
 TTTGAGATAGAGAGAGAGATCCTGCAACTGTGGTTGTCTGCTGCTGTACAG
 CGAGAGAGACATTGAGAGATATGTAGATGTTACCAAGTTGCTGCTGTTATT
 30 GTACTGGTACTGATTGTTACTGTTGCTATCATGTGCAAATTGTTGTATGGAA
 AATCAACAAAATTGATATTGCAAAGCGAGTTGGATTGAATGATTGAGAAA
 TGGTGAATGCTTCCCTCAGACTGTTGAGTGGCCTTGAGAATTGGTGTTCATAG
 GTGGTGTATGCAAGTTGCTAATGAAAGGCAGGGCTTGAAATTCCGAAAGGCAG
 CCAATGATACTTCTGAAAGTGTGTTTCTGTCAGCTTCTGCTCTTCTGTT
 35 TCTAGACACACGTTGAGCCAATGTTGTCAGCTTCTGCTCTTCTGTT
 TAGGTGCAGTTGTTTGTACAGATTGCTGGCAGAGCCCCGTGATGGCTGAG
 CCTCCAAGAGATCCTGCTGCTGACTGCGGATACGCTGAATCCTTAAACGC
 TCCCTAGTTAAGTTAGAGAACTGAGAATCAATTGGGGCAACATTACTGGG
 TCGCCTCCCTGGGCCTCTACAGTTGTTGGGCCTATATGTAAGTGCCTGTT
 40 GTGGGGATTGCGGGCTGGCGGGCATTGCGTCTCTTCTGCGCTGT
 TTCCCCCTCTTCTGCTGCTTCTGGAGGGAGGTGGTGGCGGGCGGGTGTGTT
 GGGTCGATTGGAGCGCGCAACGCCGCTGCTGCATCTGCCATTGGTT
 GTT
 45 [0457] SEQ ID NO:140
 CGTCGCCCCGTAGACCATTCTACAGAACGAGAGACATAGGGGTTCTT
 TGTTCTAACCATATCTTGTACACTTACCATACATCACCTTAGTC
 TCAAATTAGGTGAGTATTGGTCTAGCCACATCTAAGGCAAGATTGTTCTG
 AGCAGTGAACCCATATGTCATAGACAGAAAAATTGTGAAAGATTCTTCTG
 50 GGTCAAAGCGTGGTTAACATTAACTCAAGTAAGATAATGCGATAAAT
 GTGACAAAAATAATGTGTTATAGAAGTATGACAAAAATAATCACAATCCAAACA
 GTCTGATAGCTGGCGAGTGCAAAATAGATACGAAATCTGTTGATATCACACG
 GGTCCAAAATAATTGCTTGTGAGCATCAGCCTTCTGCACAAAAAGCTAGC
 CAAACAAACGAGTGGCGCCATCTGAACCACACGCTCACCGCCCGTGACA

5 GCGCCAAAGACAAAACCATACCCCTCCCCAATTCCAACCCCTCTCCGCCTCAC
 AGAAATCTCTCCCTCGCCAAACCCTCGCCGCCATGGCCGCCACCTCC
 CCCGCCGTCGATTCTCCGGCGCCGCCATGCCAACGCCCC
 CCCGCCAGCCTCTCCCGGCCACCAGCCCCCTCGGCCGCGCTCCCCGCCCC
 CGTCGTAGGTGCTCGCCGCCGGGGGGCTGCTGCCACCTCCGCCGCGCCCC
 10 GCCACCGCGCTCCGGCCCTGGGGCCGCTCCAGGCCCCGAAGGGCGCCGACAT
 CCTCGTCGAGGCGCTCGAGCGCTGCCATCGTCGACGTATTGCCCTACCCGGC
 GGC CGTCCATGGAGATCCACCAAGGCCTGACCGCTGCCGCTACCAACC
 ACCTCCTCCGCCACGAGCGAGGGGAGGCCTGCCGCTACGCCGC
 15 GCGTCCGGCCGCGTCGGCTCGCTCGCCACCTCCGGCCGGGGGCCACCAACC
 TCGTCCCGCGCTCGCTGACGCCCTCTGACTCCATCCCCATGGTCGCCATCACG
 GCCAGGTCCCCCGCCGCATGATCGGACGGACGCGTCCAGGAGACGCCATA
 GTGGAGGTACCGCGCTCCATACCAAGACAACTACCTGGTCCTTGACGTGGAGG
 ATATCCCCCGCGTCATCCAGGAAGCCTCTCCTCGCTCCTCTGCCGCCGGG
 20 GCCGGTGTGGTGTGATATCCCCAAGGATATCCAGCAGCAGATGGCCGTGCCTATC
 TGGGACACGCCGATGAGTTGCCAGGGTACATCGTCCCGCCTGCCAAGGCCACCA
 TCTACTGAATCGCTTGAGCAGGTCTCGCTCTGGTTGGCGAGTCACGGGCCCAA
 TTCTGTATGTTGGTGGCTCGCTGCATCCGGCAGGAGTTGCCGCTTGTG
 GAGCTCACTGGGATTCCGGTTACAACACTACTCTGATGGGCCTGGCAACTCCCCA
 25 GCGACGACCCACTGTCTCGCATGCTGGGATGCATGGCACTGTGTATGCAA
 TTATGCAGTCGATAAGGCTGACCTGTTGCTGCATTGGTGTGCCGTTGATGATC
 GCGTGAUTGGAAAATCGAGGCCCTTGCAAGCAGGTCCAAGATTGTCACATTG
 ACATTGACCCAGCTGAGGATTGGCAAGAACAAAGCAGCCACATGTCTCCATTGTG
 AGATGTTAAGCTTGCTTACAGGGGTGAATGCTCTATTAAATGGGAGCAAAGCA
 30 CAACAGGGTCTGGATTTGGTCCATGGCACAAGGAGTTGGATCAGCAGAACAGG
 GAGTTCCCTCTAGGATTCAAGACTTTGGCGAGGCCATCCGCCAATATGCTA
 TCCAGGTACTGGATGAGCTGACAAAAGGGGAGGCCATATTGCTACTGGTGTG
 GGCAGCACCAGATGTGGCGGCTCAGTATTACACTACAAGCGGCCACGGCAGT
 GGCTGTCTCGCTGGTGGGGCAATGGGATTGGTACCAAGCTGCAGCTGGC
 35 GCTGCTGTGGCCAACCCAGGTGTACAGTTGACATTGATGGAGATGGTAGTT
 TCCTCATGAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAA
 GGTGATGATATTGAACAAACCAGCATCTGGAAATGGTGTGCAATGGGAGGATAG
 GTTTACAAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAG
 TGAGATATATCCAGATTGACGATTGCTAAAGGATTCACGTTCCGGCAGTT
 40 CGTGTGACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACAGATGCTTGAGACC
 CCAGGGCCATACTTGTGGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCTA
 TGATCCAAGCGGTGGTGCTTCAAGGACATGATCATGGAGGGTGTGGCAGGA
 CCTCGTACTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGC
 ATGGTCCCCCGGTGTGTTCAACTACTAGGGGTTCAACTGTGAACCATGCGTT
 45 TCTAGTTGCTGTTCAATTATAGCTGTGTTACTTAGTCCGAACCCGTAA
 GCTTTGAGTCTATGCTCTTTGTAGGGATGTGCTGTGATAGATATCATGCAA
 GTTTCTGCTTACATATCAATAATAAGTACTTCCATGGAATAATTCTCAGTTCTG
 TTTGAAATTGACATCTCTACAAACAGTGTGCTGGTCTTCTGTTACTTACA
 TGTCTGCCGTGCCGGTTATGACATAATGACCGATGGAGGGTGGTCAGCAGGTT
 TAGACGGGGAGTTGAAACTTTTTGGGGGAAGAAATCTGAATACAGTTGGGA
 GGAAAGATAAAAGCATATACCTGATTAATTGAGCCAATATCCAGCCTAA
 TTTATCAAGCAATAGGCAGTGTAGGGTGTG

50 [0458] SEQ ID NO:141
 CGTCGCCGTAGACCATTCTACGAATCGGTACGGAGAGACATAGGGCTGTA
 TGGTCTTAACCATAACCTGCCACACTTGTGTCACACCTCATCTTAGGCAAATTAA
 TCAAGTTATGTAGGTGTTGGTTAGCCACATCTAAGGCAAGATTATTTCTG
 AGCAGTGAACCCATATGTTAGACATAAAAAGTGTGGAAGATTCCCTTAGT
 CAAACTGTGGCTAACAAATTATTAAGAATTAACTTAAGTAAGATAGGTGCAACAA

ATGTAGCAAAAATAATGTGGTATAATAGCAAAGATAGGCCACAACCGCGAGTGG
 AAATACCAAGATACGAGATCTCTGGTCATATCACACGAGTCCAATTAAATTGCTTT
 GTTTGAGGTTTCAGCCTTTGCATAAAAAAAGCTAGCCCCAACAAACGAGTGGCGT
 CCCATCTGAACCACACACTCACCCGCCGTGACAGGCCAACAGACAAAACCAT
 5 CACCCCTCCCCAATTCCAACCCCTCTCTGCCTCACAGAAATCTCTCCCTGCCCA
 AACCCCTGCCGCCATGGCCGCAGCCACCTCCCCGCCGTGCATTCTCGGGC
 GCCGCCGCCGCCGCCATACCCAAACCCGCCAGCCTCTCCCGGCC
 ACCAGCCCCTCGCGCGCTCCCCGCCATCGTCAGGTGCTGCCGC
 GTCCCCCGCCGCCACCTCCGTGCCCTCCGCCACCGCGCTCCGGCGTGGGGC
 10 CCCTCCGAGCCCCGCAAGGGCGCCGACATCCTCGTCGAGGCCTGGAGCGCTGC
 GGCATCGTCGACGTCTCGCCTACCCCTGGCGCGTCCATGGAGATCCACCAAGG
 CGCTGACGCGCTCGCCAGTCATCACCAACCAACCCACTTCCGCCACGAGCAGGGGA
 GGCCTTCGCGCGTCCGGGTACGCCCGCGTCCGGCCCGTCCGCTCGC
 GCCACACTCCGGCCGGGGGCCACCAACCTCGTCTCCGCCCTGCCGACGCTCTCC
 15 TCGACTCCATCCCCATGGTCGCCATCACGGGCCAGGTCCCCGCCATGATCGG
 CACGGATGCGTTCCAGGAGACGCCATCGTGGAGGTACCGCGCTCCATACCAA
 GCACAACACTACCTGGTCCTTGACGTGGAGGAATACCCCCCGGTCA
 TTCTTCCTCGCATCCTCTGGCCGCCGGGCGGTGCTGGTGA
 20 ATCCCCAAGGACATCCAGCAGCAGATGGCTGTGCCTGTGGACACGCCGATGAGTTGCCAGG
 GTACATGCCCGCCTGCCAAGCCACCATCTACTGAATCGCTTGAGCAGGTCTG
 CGTCTGGTTGGCGAGTCACGGGCCAATTCTGTATGTTGGTGGCTCGC
 CATCTGGTGAGGAGTTGCAGCTTGTGAGCTCACTGGGATTCCAGTTACAAC
 TACTCTATGGGCCCTGGCAACTCCCCAGTGACGACCCACTGTCTCGC
 25 CATGGGATGCATGGCACTGTGTATGCAAATTATGCA
 GAGTAGATAAGGCTGACCTGTTGCTTCATTGGCTTACAGGGGT
 GCTTGCATTGGTGTGCCAGTGCACATTGACATTGACCCAGCTGAG
 30 GATTGGCAAGGTTGGATCAGCAGAAGAGAGGAGTTCCCTAGGATTCAAGACTTT
 GGTGAGGCCATCCGCCCAATATGCTATCCAGGTACTGGATGAGCTGAC
 35 GAAAAA
 GGGGAGGCATTCAGGCCACGGTGTGGGCAGCAGCATCAGATGTGGCG
 TATTACACTTACAAGCGGCCACGGCAGTGGCTGTCTCATCCGG
 40 TGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCTGGCCA
 AGTTGTGACATTGATGGGATGGTAGTTCCCTCATGAACATT
 CAGGAGTTGGCG
 45 TTGATCCGTATTGAGAACCTCCAGTGAAAGGTGATGATATT
 GAACAACCAGCATC
 TGGGAATGGTGGTGCAGTGGGAGGATAGGTTTACAAGGCC
 ACACCAGGCGCACA
 CATACTTGGCAACCCAGAAAATGAGAGTGAG
 50 ATATATCCAGATTGTGACGAT
 TGCTAAAGGATTCAACGTTCCGCAGTTGAGACCCAGGGCC
 ACTTGAGGATATCATT
 GTCCCGCATCAGGAGCACGTGCTGCCTATGAT
 40 CCAAGCGGTGGCTTTAAGG
 ACATGATCATGGAGGGTGTGGCAGGACCTCGTACTGAA
 ATTTGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTCGCGTGTGTTGATCAACTAC
 TGGGGGTTCAACTGTGA
 45 ACCATGCGTTCTAGTTGCTTGTAGTTGCTCTGTTCT
 CTTGTGTTACTTAGTCCGAACCGTGTAGTTGCTTGTAG
 50 GATGTGCTGTCA
 TAAGATATCATGCAAGTTCTCAGTTCTGTTGA
 ATTTCACATCTTCACGAACA
 GTGTGCTGGTTCTGTTACTTACATGCC
 CGCGTGTGAGGTTATGACATAA
 CGACCGATGGAGGATTGGAGGGTGGCTGGCTGGTT
 ATTTTCTGGAAGAAATCTGA
 55 ATACAGTTGGAGGGAAATGGAAGCATATATT
 ATCGAGCCGCTATCCAGGCTA
 ATTATCAAGCACTAGACAGTGTAGGGTGTGG
 CATTCTCTCTCCTGATATCCGGCTGAGAGGAGAGATT
 GAGGCTCGCGCTGTG
 TTGTTGCTGATTCTACAGCA
 TTTGAGAGAGAGAGAGAGATGTTGCA
 ACTGTG
 TTTGCTTGGTTGCTGTACAGAGAAAGAGATGAC
 ATTAGAGATATGCAGATC
 GTTACCAAGTTGTGCTGC
 TTATTGTTATTGTTGCTATCATGT

5 GCAAATTGTTGATGGAAAATCAACAAAATTGATATTGCAAAGCGAGTTG
 GATTGAATGATTGAGAAATGGTACTTGTGAGTGGCCTGAGAATTGGTGT
 CATAGGTGTGCAGTTGTAATGAAAGGCGCGCTGAAATTCCGAAAGGCAG
 GCAATGATACTTCTGAAAGTGTGTTCTCCAGGTTCCGGTGGAACAAAGT
 CTACGTTGAGCCAATGTTGTCAGCTTATTCTGCTTTAGTTCAAGTTGTTGTT
 CACAGATTGCTGGCAGAGCCCCATGATCGGCTGAGCCTCCAGGAGATCCTGAT
 TGCTCGACTGC

[0459] SEQ ID NO:142

10 CGTCGCCGTAGACCATTACGAATCGGTACCGGAGAGACATAGGGCTATT
 TGGTTGTAGCCACATTGTCATACTTGTGACACCGCATCTATGCAAGTTGA
 TCAAATTAGGTGGATGTTAGTCTAACCATGTAAGGAAAGATTTTTTTA
 TGAGCATTGAACCGTAGACACAAAAAGTGTAGGAAGATTACTTAAACAAGCT
 AAAGTGTGGCTAACAAATTAAAGCATCTCAGGTAAGATAAGTGCACAAATATGG
 15 CAAAAATAATGTGGTATATATGACAAAGATAGTCACAATCCAAACAGCCCATAG
 CCTGGCGAGTGCAAATAGATACGAGATCTCTGGTGTATCACAACCGTCCAAATT
 AATTGCTTGTTCAGCATCAGCCTTTGCATAAAGAAGCTAGCCCAATCTGAAC
 CACACACTCACC CGCCGTGACAGCGCAAAGACAAAACCATACCCCTCCCC
 20 AATTCCAACCCCTCTCTGCCTCACAGAAATCTCCCCCTGCCAAACCCCTGCC
 GCCGCCATGGCCGCCACCTCCCCGCCGCGCATTCTCGGGGCCACCGCCG
 CCGCCATGCCAAACCCGCCATCCTCTCCCGGCCACAGCCGCTCTCGCG
 CCGCGCGCTCCCCGCCGCGTCAGGTGTTGCCCGCTCCCCGCCACC
 TCCGCCGCGCCTCCGCAACCGCGCTCCGCCCTGGGGCCCGTCCGAGCCCCGCA
 AGGGCGCCGACATCCTCGTCGAGGCGCTCGAGCGCTGCCGATCGTCGACGTCTT
 25 CGCCTACCCCGCGCCGCTCCATGGAGATCCACCAAGGGCGCTGACCGCCTCGCC
 GTCATCACCAACCACCTCTCCGCCACGAGCAGGGGGAGGCCTCGCGCGTCCG
 GCTACGCCGCGCGTCCGCCGCGTCGGCGTCTGCGTCGCCACCTCCGCCGG
 GGCCACCAACCTCGTCTCGCGCTCGCCACGCCCTCTCGACTCCATCCCCATG
 GTCGCCATCACGGGCCAGGTCCCCCGCCGATGATGGCACGGACCGCGTCCAGG
 30 AGACGCCATAGTGGAGGTACCGCGCTCCATACCAAGCACAACACTACCTGGTCT
 TGACGTGGAGGATATCCCCCGCGTCATCCAGGAAGCCTCTTCCTGCATCCTCTG
 GCCGCCGGGGCCGGTCTAGTTGATATCCCCAAGGACATCCAGCAGCAGATGG
 CTGTGCCGTCTGGACACTCCAATGAGTTGCCAGGGTACATGCCGCCCTGCC
 CAAGCCACCATCTACTGAATCGCTTGAGCAGGTCTGCCGCTGGTGGCGAGTCA
 35 CGGCCCAATTCTGTATGTTGGTGGTGGCTGCCGCTGCCGCTGGCGAGGAGTTGC
 GCCGCTTGTGAGCTACTGGGATTCCAGTTACAACACTACTCTGATGGGCCCTGGC
 AACTCCCCAGCGACGACCCACTGTCTCGCATGCTGGATGCAATGGCACTG
 TGTATGCAAATTATGCACTAGATAAGGCTGACCTGCTCGCATTGGTGTGCG
 GTTGATGATCGTGTGACTGGAAAATCGAGGCTTGTCAAGCAGGTCAAAGATT
 40 GTGCACATTGACATTGACCCAGCTGAGATTGGCAAGAACAAAGCAGGCCACATGTCT
 CCATTGTGCAGATGTTAAGCTGCTTACAGGGTTGAATGATCTATTAAATGG
 GAGCAAAGCACAACAGGGCTGGATTGGTCCATGGCACAAGGAGTTGGATCA
 GCAGAAGAGGGAGTTCCCTAGGATTCAAGACTTTGGCGAGGCCATCCCGCCG
 CAATATGCTATCCAGGTACTGGATGAGCTGACAAAAGGGGAGGGCATATTGCC
 45 ACTGGTGTGGCAGCACAGATGTGGCGGCTCAGTATTACACTACAAGCGGC
 CACGGCAGTGGCTGTCTCGTGGTTGGGGCAATGGGATTGGGTACCGC
 TGAGCTGGCGCTGCTGTGGCCAACCCAGGTGTTACAGTTGTCGACATTGATGGT
 GATGGTAGTTCCCTCATGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAAC
 TCCCAGTGAAGGTGATGATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGT
 50 GGGAGGATAGGTTACAAGGCCAATCGGGCGCACACATACCTGGCAACCCAG
 AAAATGAGAGTGAGATATCCAGATTGTCAGGATTGCTAAAGGATTCAACGT
 TCCAGCAGTTGAGTGACGAAGAAGAGCGAAGTCAGCAATCAAGAACAGT
 GCTTGAGACCCAGGGCCATACTTGTGGATATCATAGTCCCGCATCAGGAGCAC
 GTGCTGCCTATGATCCAAGCGGTGGTCTTCAAGGACATGATCATGGAGGGTG

5 ATGGCAGGACCTCGTACTGAAATTGACCTACAAGACCTACAAGTGTGACATGC
 GCAATCAGCATGATGCCGCGTGTGTTATCAACTACTAGGGGTTCAACTGTGAGC
 CATCGTTTCTAGTTGCTTGTTCATTCAATAAGCTTGTATTACTAGTTCCGA
 ACCCTGTAGTTGTAGTCTATGTTCTTTGTAGGGATGTGCTGTCAAGATG
 10 TCATGCAAGTTCTTGTACATATCAATAAGTACTTCCATGGAATAATTCT
 CAGTTCTGTTGAATTTCATCTTCTACAAACAGTGTGCTGGTCTTCTGTT
 ACTTTACATGTCTGCTGTGACATAACGACCGATGGAGGGTGGTCG
 GCAGGTTTAGAAGGGATTGAAACTTTGGAAAGAAGTCTGAATACAGT
 TGGGAGGAAAAATAGAAGTATATACTTCGATTAATTATCAAGCCCCTATCCAG
 15 TCTAATTATCAAGCACTAGACAGTGTAGGGTGTGGCATTCTCTTCTGAG
 ATCCGGCTTGAGAGGAGACCGAGGGCTCGGCTGTGTTGGTGTGATTCTAC
 AGCTTTGAGATAGAGAGAGAGATCCTGCAACTGTGGTTGTCTGCTGCTTGT
 ACAGCGAGAGAGACATTGAGAGATATGTAGATGTTACCACTGTGCTGCTTGT
 ATTGTTACTGGTACTGATTGTTACTGTTGCTATCATGTGCAAATTGTTGTGAT
 20 GGAAAATCAACAAAATTGATATTGCAAAGCGAGTGGATTGAATGATTGA
 GAAATGGTGACTGCTTCCCTCAGACTTGTGAGTGGCCTTGAGAATTGGTGTTC
 ATAGGTGGTGTATGCAGITGCTAATGAAAGGCGACGGCTTGAAATTCCGAAAG
 GCAGCCAATGATACTTCTGAAAGTGTGTTTCAGCTTATTCTGCTTCTAGTTTC
 CAAGTCTAGACACACGTTGAGCCAATGTTGTCACTTATTCTGCTTCTAGTTTC
 AGTTAGGTGCAGTTGTTTGTACAGATTGCTGGCAGAGCCCCGTGATCGGC
 TGAGCCTCCAAGAGATCCT

25 [0460] SEQ ID NO:175
 tnantggta ggtgctggtg gtccgaaggt ccacgccc aactacg

[0461] SEQ ID NO:176
 CNANTACGTAGTTGGCGGCGTGGACCTTCGGACCACCAGCACCTAAC

30 [0462] SEQ ID NO:177
 TNANTGGTTAGGTGCTGGTGGTCCGAAGGTCCACGCCAACACTACG

[0463] SEQ ID NO:178
 ANGNGTCGTAGTTGGCGGCGTGGACCTTCGGACCACCAGCACCTAAC

35 [0464] SEQ ID NO:179
 TGGATATCATAGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCAAGCGGTGG
 TGCTTCAAGGACATGATCATGGGTTAGGTGCTGGTGGTCCGAAGGTCCACGCCG
 CCAACTACGTGGATATCATAGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTGTGTTCAAGGACATGATCATGG

40 [0465] SEQ ID NO:180
 CCATGATCATGTCCTGAAAGCACCACCGCTTGGGATCATAGGCAGCACGTGCTC
 CTGATGCGGGACTATGATATCCACGTAGTTGGCGGCGTGGACCTTCGGACCACCA
 GCACCTAACCCATGATCATGTCCTGAAAGCACCACCGCTTGGGATCATAGGCAG
 45 CACGTGCTCCTGATGCGGGACTATGATATCCA

[0466] SEQ ID NO:181
 TGGATATCATAGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCAAGCGGTGG
 TGCTTCAAGGACATGATCATGGGTTAGGTGCTGGTGGTCCGAAGGTCCACGCCG
 CCAACTACGGATGGCAGGACCTCGTACTGAAATTGACCTACAAGACCTACAAG
 50 TGTGACATGCGCAATCAGCATGATGCCCGGT

[0467] SEQ ID NO:182

ACGCGGGCATCATGCTGATTGCGCATGTCACACTTGTAGGTCTTGTAGGTCGAAA
TTTCAGTACGAGGTCTGCCATCCGTAGTTGGCGCGTGGACCTTCGGACCA
GCACCTAACCCATGATCATGTCCTTGAAAGCACCACCGCTGGGATCATAGGCAG
5 CACGTGCTCCTGATGCGGGACTATGATATCCA

[0468] SEQ ID NO:183

GGAGTTGGCGTTGATCCGNC

10 **[0469]** SEQ ID NO:184

AACTACAGGGTTCGGAACTAAGTAANT

[0470] SEQ ID NO:185

15 TGGCAGGATATATTGTGGTGTAAACAAATTGACGCTTAGACAACCTAATAACACA
TTGCAGCAGTTTAATGTAATGACTGAATTAACGCCATTGAATTGAGCTCGGTAC
CACTGGATTTGGTTTAGGAATTAGAAATTATTGATAGAAGTATTTACAAAT
ACAAATACATACTAAGGGTTCTTATGCTAACACATGAGCGAAACCCATAA
20 GAACCTTAATTCCCTTATCTGGGAACTAACACATTATTCTGGAGAAAAATAG
AGAGAGATAGATTGTAGAGAGAGACTGGTGTAGTTGCGGACTCTATTAGATCT
GGGTAACTGGCCTAATGGCCTGGAGGAGCTGGCAACTCAAATCCCTTGCCA
AAAACCAACATCATGCCATTCCACCATGCTTGTATCCAGCTGCGCGCAATGTACCC
25 CGGGCTGTGTATCCAAAGCCTCATGCAACCTAACAGATGGATGTTGGAAAGGC
CTATAACAGCAACCACAGACTAAAACCTTGCCTCCATAGACTTAAGCAAATG
TGTGTACAATGTGGATCCTAGGCCAACCTTGTATGCCTATGTGACACGTAAACA
GTACTCTCAACTGTCCAATCGTAAGCGTCTCTAGCCTCCAGGGCCAGCGTAAG
CAATACCAGCCACAACACCCCTAACCTCAGCAACCAACCAAGGGTATCTATCTG
30 CAACCTCTCGAGATCATCAATCCACTCTGTGGTGTGGCTCTGCTAAAGT
TCACTGTAGACGTCTCAATGTAATGGTAACGATATCACAACCGCGGCCATATC
AGCTGCTGTAGCTGGCTAATCTCAACTGGTCTCCTCTCCGGAGACATGGCTTCTA
CCTACAAAAAAAGCTCCGCACGGCTGCATTGTCACAAATCATGAAAAGAAAA
ACTACCGATGAACAATGCTGAGGGATTCAAATTCTACCCACAAAAAGAAGAAAG
AAAGATCTAGCACATCTAACGCTGACGAAGCAGCAGAAATATATAAAAATATAA
35 ACCATAGTGCCCTTCCCTCTGATCTTGTAGCATGGCGAAATTTAA
ACCCCCCATCATCTCCCCAACAAACGGGGATCGCAGATCTACATCCGAGAGGCC
CATTCCCCGCGAGATCCGGGCCGGATCCACGCCGGCGAGAGCCCCAGCCCGAG
ATCCCGCCCCCTCCCGCGACCGATCTGGCGCGCACGAAGCCGCTCTGCCAC
CCAAACTACCAAGGCCAAAGATCGAGACCGAGACGGAAAAAAACGGAGAA
40 AGAAAGAGGAGAGGGCGGGGTGGTTACCGGCCGGCGCGAGGGAG
GGGGGAGGAGCTCGTCTCCGGCAGCGAGGGGGAGGGAGGTGGAGGTGGTGGT
GGTGGTGGTGGTAGGGTTGGGGGATGGGAGGAGGGGGGGTATGTATATAG
TGGCGATGGGGGGCGTTCTTGGAAAGCGGAGGGAGGGCCGGCTCGCTGG
CTCGCGATCCTCTCGCTTCCGGCCCCACGACCCGGACCCACCTGCTTTTT
45 TCTTTTCTTTTTCTTCTTTCTTTCTTTGGCTGCGAGACGTGCGGTGCGTGC
GGACAACTCACGGTGTAGTGGGGGGGTGGAGGACTATTGTCAGTTGGCTGG
ACTGGGGTGGGTTGGGTTGGGTTGGGCTGGCTGGCTATGGATCGTGGAT
AGCACCTTGGCTTACCGAACTTACGGGTTGTTGTAAATGTTGAGTCTAA
GTTTATCTTTATTTACTAGAAAAAAACCCATGCGCTGCAACGGGGAAAGC
50 TATTTAATCTTATTATTGTCATTGTGAGAATTGCCCTGAATATATATTTCTCA
AAAATTATGTCAAATTAGCATATGGTTTTAAAGATATTCTTACAAATCC
CTCTGTATTACAAAAGCAAACGAACTAAACCCGACTCAAATACAGATATGCA
TTTCCAAAAGCGAATAAAACTAAAAACCAATTCTACACAAATGACGTATCAA
GTACCGACAAAAACATCCTCAATTCTATAATAGTAGAAAAGAGTAAATTCACT
TTGGGCCACCTTTATTACCGATTTACTTACCCACCTTAACGTGATGTTT

CACTTTGACCAGGTAATCTAACCTTGTAAATTTGGACTATCCCGACTCTCTTC
 TCAAGCATATGAATGACCTCGAGATGCTAGTCTAGAGTCGACCTGCAGGGTCA
 GCGTACCCGGTCGTGCCCTCTCTAGAGATAATGAGCATTGATGTCTAAGTTA
 TAAAAAAATTACACATATTTTTGTACACTGTTGAAGTGCAGTTATCTAT
 5 CTTTATACATATTTAAACTTACTCTACGAATAATATAATCTATAGTACTACAA
 TAATATCAGTGTAGAGAATCATATAAATGAACAGTTAGACATGGTCTAAAGG
 ACAATTGAGTATTTGACAACAGGACTCTACAGTTATCTTTAGTGTGCATGT
 GTTCTCCTTTTTGCAAATAGCTCACCTATAATAACTCATCCATTATTAA
 GTACATCCATTAGGGTTAGGGTTAATGGTTTATAGACTAATTTTTAGTAC
 10 ATCTATTTTATTCTATTAGCCTCTAAATTAAGAAAACAAAACCTATTAGT
 TTTTTATTAAATAATTAGATATAAAATAGAATAAAAATAAAGTACTAAAAATT
 AAACAAATACCCCTTAAGAAATTAAAAAAACTAAGGAAACATTTCCTGTTCG
 AGTAGATAATGCCAGCCTGTTAACGCCGTCGACGAGTCTAACGGACACCAACC
 15 AGCGAACCCAGCAGCGTCGCGTCGGGCCAAGCGAAGCAGACGGCACGGCATCTCT
 GTCGCTGCCTCTGGACCCCTCTCGAGAGTCCGCTCCACCGTTGGACTTGCTCCGC
 TGCGGCATCCAGAAATTGCGTGGCGAGCGGCAGACGTGAGCCGGCACGGCAG
 GCGGCCTCCTCCTCCTCACGGCACGGCAGCTACGGGGATTCTTCCACCG
 CTCCTCGCTTCCCTCCTCGCCGCCGTAAATAATAGACACCCCCCTCACACCC
 20 TCTTCCCCAACCTCGTGTGTTCGGAGCGCACACACACACAACCAGATCTCCCC
 CAAATCCACCGTCGGCACCTCCGCTCAAGGTACGCCGCTCGTCCCTCCCCCCCC
 CCCCCCTCTACCTCTAGATCGCGTTCCGGTCCATGGTTAGGGCCCGTAGT
 TCTACTCTGTTCATGTTGTAGATCCGTGTTAGATCCGTGCTAG
 CGTTCGTACACGGATGCGACCTGTACGTCAGACACGTTCTGATTGCTAACTGCC
 AGTGTCTCTGGGAATCCTGGGATGGCTCTAGCCGTTCCGCAGACGGGATC
 25 GATTCATGATTTTTGTTCGTTGCATAGGGTTGGCTTGCCTTCTTAT
 TTCAATATATGCCGTGACTTGTGTTGGGTCTAGATCGGAGTAGAATTCT
 TGTTGTGATGATGTGGTCTGGTTGGCGGTCTAGATCGGAGTAGAATTCT
 GTTCAAACACTACGGGATTATTAAATTGGATCTGTATGTGTGCCATACA
 TATTCAAGTTACGAATTGAAGATGATGGATGGAAATATCGATCTAGGATAGGTA
 30 TACATGTTGATGCCGGTTTACTGATGCATATACAGAGATGCTTTGTTCGCTTG
 GTTGTGATGATGTGGTGTGGTGGCGGTCTGTCATTGCTCTAGATCGGAGTAG
 AATACTGTTCAAACACTACCTGGTATTATTAAATTGGAACTGTATGTGTGT
 CATACTTCACTAGTTACGAGTTAACAGATGGATGGAAATATCGATCTAGGATAG
 GTATACATGTTGATGTGGGTTTACTGATGCATATACATGATGGCATATGCAGCA
 35 TCTATTCACTGCTCAACCTTGAGTACCTATCTATTATAATAAAAGTATGTT
 TATAATTATTTGATCTGATATACCTGGATGATGGCATATGCAGCAGCTATATGT
 GGATTGCTACCCCTGCTTACAGCTATTATTGCTGGTACTGTTCTTGTAC
 TCGATGCTCACCCCTGTTGGTACTCTGAGGAGGATCACAAGTTGTAC
 40 AAAAAAGCAGGCTATGCCGCCACCTCCCCGCCGTCGCTTCTCGGGCGCC
 ACCGCCGCCCATGCCAACCCGCCATCCTCTCCGCCACCAGCCCG
 TCTCGCGCCGCGCCTCCCCGCCGCGTCGTCAGGTGTTGCGCCGCTCCCCGC
 CGCCACCTCCGCCGCGCTCCCGCAACCGCGCTCCGCCCTGGGCCGCTCCGAG
 CCCCAGCAAGGGCGCCGACATCCTCGTCAAGGCGCTCGAGCGCTCGGCATCG
 ACGTCTCGCCCTACCCCGCGGCCTCATGGAGATCCACCAAGGGCGCTGACGCG
 45 CTCGCCGTCATACCAACCAACCTCTCCGCCACGAGCAGGGGGAGGCCTCGCG
 GCGTCGGCTACGCCCGCGCTCCGGCGCGTCGGCGTCTGCGTCGCCACCTCCG
 GCGGGGGGCCACCAACCTCGTCTCCGCCGCTCGGCCACGCCCTCTGACTCCAT
 CCCCAGGTCGCCATCACGGGCCAGGTCCCCGCCGATGATGGCACGGACGCG
 TTCCAGGAGACGCCATAGTGGAGGTACCGCCTCCATACCAAGCACAACCTACC
 50 TGGTCCTGACGTGGAGGATATCCCCCGCTCATCCAGGAAGCCTTCTCCTTGC
 ATCCTCTGGCCGCCGGGGCCGGTGCTAGTTGATATCCCCAAGGACATCCAGCAG
 CAGATGGCTGTGCCCCTGGGACACTCCAATGAGTTGCCAGGGTACATCGCCC
 GCCTGCCCAAGCCACCATCTACTGAATCGCTTGAGCAGGTCTCGTCTGGTGG
 CGAGTCACGGCGCCCAATTCTGTATGTTGGTGGCTCGCTCGTCTGGCGAG

GAGTTGCGCCGCTTGTGAGCTTACTGGGATTCCAGTTACAACACTACTCTGATGG
 GCCTTGGCAACTTCCCCAGCGACGACCCACTGTCTCTGCGCATGCTGGGATGCA
 TGGCACTGTGTATGCAAATTATGCACTAGATAAGGCTGACCTGTTGCTCGCATT
 5 GGTGTGCGGTTGATGATCGTGTGACTGGAAAATCGAGGCTTGTCAAGCAGGT
 CCAAGATTGTGCACATTGACATTGACCCAGCTGAGATTGGCAAGAACAAAGCAGC
 CACATGTCTCCATTGTGCAGATGTTAAGCTTACAGGGGTTGAATGATCTA
 TTAAATGGGAGCAAAGCACAAACAGGGTCTGGATTTGGTCCATGGCACAAGGAG
 TTGGATCAGCAGAAGAGGGAGTTCTCTAGGATTCAAGACTTTGGCGAGGCCA
 10 TCCC GCCGCAATATGCTATCCAGGTACTGGATGAGCTGACAAAAGGGGAGGCAG
 TCATTGCCACTGGTGTGGCAGCACCAGATGTGGCGGCTCAGTATTACACTTA
 CAAGCGGCCACGGCAGTGGCTGTCTCGTCTGGTTGGGGCAATGGGATTGGG
 TTACCAAGCTGCAGCTGGCGCTGCTGTGGCCAACCCAGGTGTTACAGTTGTGACA
 TTGATGGTGATGGTAGTTCTCATGAACATTCAAGGAGTTGGCGTTGATCCGCATT
 15 GAGAACCTCCCAGTGAAGGTGATGATATTGAACAACCAGCATCTGGAAATGGTG
 GTGCAGTGGGAGGATAGGTTACAAGGCCAATCAGGGCGCACACATACCTGGC
 AACCCAGAAAATGAGAGTGAGATATATCCAGATTGTGACGATTGCTAAAGGA
 TTCAACGTTCCAGCAGTCAGTGACGAAGAAGAGCGAAGTCAGCAGCAATC
 AAGAAGATGCTGAGACCCCAGGGCCAATTGTGGATATCATAGTCCCACATC
 20 AGGAGCACGTGCTGCCTATGATCCAAATGGTGGTGTCAAGGACATGATCAT
 GGAGGGTGATGGCAGGACCTCGTACTGATAACCCAGTTCTGTACAAAGTGGTG
 ATCCTACTAGTAGAAGGAGTGCCTCGAAGCAGATCGTTCAAACATTGGCAATAA
 AGTTCTTAAGATTGAATCCTGTTGCCGGTCTGCGATGATTATCATATAATTCT
 GTTGAATTACGTTAACGATGTAATAATTAAACATGTAATGCATGACGTTATTATG
 AGATGGGTTTTATGATTAGAGTCCCGCAATTATACATTAAACCGCGATAGAAA
 25 ACAAAATATAGCGCGCAAACTAGGATAAAATTATCGCGCGGGTGTCACTATGTT
 ACTAGATCGAAAGCTAGCTTAGGAGCTTGGATCAGATTGCGTTCCGCCTCAGT
 TTAAACTATCAGTGTGACAGGGATATTGGCGGGTAAAC

30 [0471] SEQ ID NO:186
 ATTTCCATTCACTTGGCCC

[0472] SEQ ID NO:187
 TGCTATCTGGCTCAGCTGC

35 [0473] SEQ ID NO:188
 ATGGTGGAAAGGGCGGTTGTGA

[0474] SEQ ID NO:189
 CTCCCGCGCACCGATCTG

40 [0475] SEQ ID NO:190
 CCCGCCCTCTCCTCTTC

[0476] SEQ ID NO:191
 45 AAGCCGCCTCTGCCACCCA

[0477] SEQ ID NO:192
 AYCAGATGTGGCGGGCTCAGTAT

50 [0478] SEQ ID NO:193
 GGGATATGTAGGACAAGAAACTGCGATGA

[0479] SEQ ID NO:194

ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTCTGGTTGGGGCAATGGGATTGGGTTACAGCTGCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTACAGTTGACATTGATGGAGATGGTAGTTCTCAT
 5 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGGTATG
 ATATTGAACAACCAGCATCTGGGAATGGTGGTGCAATGGGAGGATAGGTTTAC
 AAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 10 CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACATGCTTGAGACCCCAGGGC
 CATACTTGGATATCATCGTCCGCATCAGGAGCACGTGCTGCCTATGATCCc
 AatGGcGGcGCTTCAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTACTG
 AAATGGTCCGAAGGTCCACGCCAACACTACGAGTATGATCCCAGCGGTGGT
 15 CTTCAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTACTGAAATTGCA
 CCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGGCCCGTGTGTA
 TCAACTACTAGGGGTTCAACTGTGACCATGCGTTCTAGTTGCTTGTTCATT
 CATATAAGCTTGTGTTACTTAGTCCGAACCCCTGTAGCTTGTAGTCTATGCTCTC
 TTTGTAGGGATGTGCTGTACATAAGATATCATGCAAGTTCTGTCCATACATAC

[0480] SEQ ID NO:195

ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTCTGGTTGGGGCAATGGGATTGGGTTACAGCTGCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTACAGTTGACATTGATGGAGATGGTAGTTCTCAT
 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGGTATG
 ATATTGAACAACCAGCATCTGGGAATGGTGGTGCAATGGGAGGATAGGTTTAC
 25 AAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACATGCTTGAGACCCCAGGGC
 CATACTTGGATATCATCGTCCGCATCAGGAGCACGTGCTGCCCTCCAAAGC
 GGTGGTGTTCAGGACATGATCATGGAGGGTATGGCAGGACCTCGTACTG
 30 AATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGGCCGC
 GTGTTGTATCAACTACTAGGGGTTCAACTGTGACCATGCGTTCTAGTTGCTT
 GTTCTCATATAAGCTTGTGTTACTTAGTCCGAACCCCTGTAGCTTGTAGTCT
 ATGCTCTCTTGTAGGGATGTGCTGTACATAAGATATCATGCAAGTTCTGTCT
 ACATATC

[0481] SEQ ID NO:196

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTACAGTTGACATTGATGGGATGGTAGTTCTCA
 40 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGGTATG
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAAGTGGGAGGATAGGTTTA
 CAAGGCCAACCAGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACATGCTTGAGACCCCAGGG
 45 CCATACTTGGATATCATTGTCCGCATCAGGAGCACGTGAatGGcGGcGCTTTC
 AAGGACATGATCATGGAGGGTATGGCAGGACCTCGTACTGAAATGGTCCGAAG
 GTCCACGCCGCAACTACGAGTATGATCCCAAGCGGTGGTGTCTTAAAGGACATG
 ATCATGGAGGGTATGGCAGGACCTCGTACTGAAATTCGACCTACAAGACCTAC
 AAGTGTGACATGCGCAATCAGCATGATACCTGCGTGTGTACTCAACTACTGGGG
 50 TTCAACTGTGAACCATGCGTTCTAGTTGCTTGTGTTCTCATATAAGCTTGTG
 TACTTAGTCCGAACCGTGTAGTTGTAGTCTCTGTCTTGTAGGGATGTG
 CTGTACATAAGATATCATGCAAGTTCTGTCCATACATAC

[0482] SEQ ID NO:197

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCACTCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGGATGGTAGTTCCCTCA
 5 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTCAGTGGGAGGGATAGGTTTA
 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTG
 10 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGGATATCATTGCCCCATCAGGAGCAGTGCTGCCTATGATCCC
 AAGCGGTGGTCTTTAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTTCAACTACTGGGGTTCAACTGTGAACCATGCGTTCTAGTT
 15 GCTTGTGTTCAATTATAGCTGTGTTACTTAGTCCGAACCGTAGTTGTA
 GTCTCTGTTCTCTTTGAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTG
 TCCTACATATC

[0483] SEQ ID NO:198

ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 20 TTCTGGTTGGGGCAATGGGATTGGGTTACCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTATGGTAGTTCCCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTCAGTGGGAGGGATAGGTTTA
 CAAGGCCATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 25 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCGAGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGGATATCATAGTCCCGCATCAGGAGCAGTGCTGCCTAatGGcGGc
 GCTTCAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTACTGAAATGGTC
 CGAAGGTCCACGCCAACACTACGAGTCCCAAGCGGTGGTCTTCAAGGACAT
 30 GATCATGGAGGGTATGGCAGGACCTCGTACTGAAATTGACCTACAAGACCTA
 CAAGTGTGACATGCGCAATCAGCATGATGCCCGTGTGTTATCAACTACTAGGG
 GTTCAACTGTGAGCCATGCGTTCTAGTTGCTTCAATTATAGCTGT
 ATTACTTAGTCCGAACCCGTAGTTGAGTCTATGTTCTTTGAGGGATG
 TGCTGTCATAAGATGTCATGCAAGTTCTGCTACATATC
 35

[0484] SEQ ID NO:199

ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCTGGTTGGGGCAATGGGATTGGGTTACCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTATGGTAGTTCCCTCA
 40 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTCAGTGGGAGGGATAGGTTTA
 CAAGGCCATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCGAGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 45 CCATACTTGGATATCATAGTCCCGCATCAGGAGCAGTGCTGCCTATGATCC
 CAAGCGGTGGTCTTCAAGGACATGATCATGGAGGGTATGGCAGGACCTCGT
 ACTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
 CCCCGTGTGTTATCAACTACTAGGGGTCAACTGTGAGGCCATGCGTTCTAGTT
 TGCTTGTGTTCAATTATAGCTGTATTACTTAGTCCGAACCCGTAGTTGTA
 50 GTCTATGTTCTCTTTGAGGGATGTGCTGTCATAAGATGTCATGCAAGTTCTG
 TCCTACATATC

[0485] SEQ ID NO:200

ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGGATTGGTTACCAGCTGCAGCTGGCGCTGCTG
 5 TGGCCAACCCAGGTGTACAGTTGACATTGATGGAGATGGTAGTTCCCTCAT
 GAACATTCACTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATG
 ATATTGAACAACCAGCATCTGGGAATGGTGGTGCAATGGGAGGATAGGTTTAC
 AAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 10 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGGC
 CATACTTGGATATCATCGTCCCGATCAGGAGCACAAGCGGTGGTGCTTCA
 AGGACATGATCATGGAGGGTATGGCAGGACCTCGTACTGAAATTGACCTAC
 15 AAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGCCCGTGTGTATCAAC
 TACTAGGGGTTCAACTGTGAACCATGCGTTCTAGTTGCTTGTATCATAT
 AAGCTTGTGTTACTTAGTCCGAACCCCTGTAGCTTGTAGTCTATGCTCTTTGT
 AGGGATGTGCTGTATAAGATATCATGCAAGTTCTGTCCATACATAC

[0486] SEQ ID NO:201

ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGGATTGGTTACCAGCTGCAGCTGGCGCTGCTG
 20 TGGCCAACCCAGGTGTACAGTTGACATTGATGGAGATGGTAGTTCCCTCAT
 GAACATTCACTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATG
 ATATTGAACAACCAGCATCTGGGAATGGTGGTGCAATGGGAGGATAGGTTTAC
 AAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 25 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGGC
 CATACTTGGATATCATCGTCCCGATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTGTTCAGGACATGATCATGGAGGGTATGGCAGGACCTCGTA
 30 CTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGC
 CCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGAACCATGCGTTCTAGTT
 GCTTGTTCATTATATAAGCTTGTGTTACTTAGTCCGAACCCCTGTAGCTTGTG
 GTCTATGCTCTTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTTG
 TCCTACATATC

[0487] SEQ ID NO:202

35 ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCATCCGGTTGGGTGCAATGGGATTGGGTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTACAGTTGACATTGATGGGATGGTAGTTCCCTCA
 TGAACATTCACTGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAAGTGGAGGATAGGTTTA
 40 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGGATATCATCGTCCCGATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTGTTCAGGACATGATCATGGAGGGTATGGCAGGACCTCGTA
 45 CTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTATCAACTACTGGGGTTCAACTGTGAACCATGCGTTCTAGTT
 GCTTGTTCATTATATAAGCTTGTGTTACTTAGTCCGAACCGTGTAGTTTGTG
 GTCTCTGTTCTCTTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTTG
 TCCTACATATC

50

[0488] SEQ ID NO:203

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCATCCGGTTGGGTGCAATGGGATTGGGTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTACAGTTGACATTGATGGGATGGTAGTTCCCTCA

TGAACATTCAGGAGTGGCGTTGATCCGTATTGAGAACCTCCCAGTGAAGGTGAT
GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
CAAGGCCAACCGGGCGCACACATACCTTGGCAACCCAGAAAATGAGAGTGAGAT
ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTG
5 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
CCATACTTGTGGATATCATTGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCC
AAGCGGTGGTGCTTTAAGGACATGATCATGGAGGGTGATGGCAGGACCTCGTA
CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
10 CTGCGTGTGTATCAACTACTGGGGGTTCACTGTGAACCATGCGTTCTAGTTT
GCTTGTTCATTCATATAAGCTTGTGTTACTTAGTCCGAACCGTGTAGTTTGTG
GTCTCTGTTCTCTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTG
TCCTACATATC

[0489] SEQ ID NO:204
15 ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
TTCGCTGGTTGGGGCAATGGGATTGGGTTACAGCTGCAGCTGGCTGCT
GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTATGGTAGTTCTCA
TGAACATTAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCAGTGAAGGTGAT
GATATTGAACAACCAGCATCTGGGAATGGTGGTGCACTGGGAGGATAGGTTTA
20 CAAGGCCAATCGGGCGCACACATACTTGCACACAGAAAATGAGAGTGAGAT
ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCGAGTG
ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACATGCTTGAGACCCCCAGGG
CCATACTGTTGGATATCATAGTCCGCATCAGGAGCACGTGCTGCCTATGATCC
25 cAatGGcGGcGCTTCAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTACT
GAAATGGTCCGAAGGTCCACGCCGCCAACTACGAGTCCAAAGCGGTGGTGCTTTC
AAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTACTGAAATTGACCTA
CAAGACCTACAAGTGTGACATGCCAATCAGCATGATGCCCGGTGTTGATCAA
CTACTAGGGGTTCAACTGTGAGCCATGCGTTTCTAGTTGCTTGTITCATTCAA
TAAGCTTGATTACTTAGTCCGAACCTGTAGTTGAGTCTATGTTCTTTG
30 TAGGGATGTGCTGTCATAAGATGTCATGCAAGTTCTGTCTACATATC

[0490] SEQ ID NO:205
35 ACCAGATGTGGCGGCTCAGTATTACACTACAAGCGGCCACGGCAGTGGCTGTC
TTCGTCTGGTTGGGGCAATGGGATTGGGTACCAAGCTGCAGCTGGCGCTGCT
GTGGCCAACCCAGGTGTTACAGTTGTCACATTGATGGTGATGGTAGTTCTCA
TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
GATATTGAACAACCAGCATCTGGGAATGGTGGTGCACTGGGAGGGATAGGTTTA
CAAGGCCAATCGGGCGCACACATACCTTGGCAACCCAGAAAATGAGAGTGAGAT
ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTTGAGTG
40 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACATGCTTGAGACCCCCAGGG
CCATACTTGTGGATATCATAGTCCGCATCAGGAGCAGTGCTGCCTATGATCC
CAAGCGGTGGTGCCTTCAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGT
ACTGAAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
CCCAGGTGTTGATCAACTACTAGGGGTTCAACTGTGAGCCATGCGTTCTAGTT
45 TGCTTGTTCATTGATATAAGCTTGATTACTAGTTCCGAACCCCTGTAGTTTGTG
GTCTATGTTCTTTGTAGGGATGTGCTGTCATAAGATGTCATGCAAGTTCTTG
TCCTACATATC

[0491] SEQ ID NO:206
50 ACCAGATGTGGCGGCTCAGTATTACACTACAAGCGGCCACGGCAGTGGCTGCT
TTCGTCTGGTTGGGGCAATGGGATTGGGTTACCAAGCTGCAGCTGGCGCTGCTG
TGGCCAACCCAGGTGTTACAGTTGACATTGATGGAGATGGTAGTTCCCTCAT
GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATG
ATATTGAACAACCAGCATCTGGGAATGGTGGTCAATGGGAGGGATAGGTTTAC

AAGGCCAACGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCAGGGC
 CATACTTGTGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCc
 5 AatGGcGGcGCTTCAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTACTG
 AAATGGTCCGAAGGTCCACGCCAACTACGAGTATGATCCCAAGCGGTGGTG
 CTTCAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTACTGAAATTGCA
 CCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGCCCGTGTGTA
 TCAACTACTAGGGGTTCAACTGTGACCATGCGTTCTAGTTGCTTGTTCATT
 10 CATATAAGCTTGTGTTACTTAGTCCGAACCCCTGTAGCTTGTAGTCTATGCTCTC
 TTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTGCTACATATC

[0492] SEQ ID NO:207

15 ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGGATTGGGTTACAGCTGCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTTACAGTTGACATTGATGGAGATGGTAGTTCTCAT
 GAACATTCAAGGAGTTGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATG
 ATATTGAACAACCAGCATCTGGGAATGGTGGTCAATGGGAGGATAGGTTTAC
 20 AAGGCCAACGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCAGGGC
 CATACTTGTGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTGTTCAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTA
 25 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGC
 CCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGACCATGCGTTCTAGTT
 GCTTGTTCATTATATAAGCTTGTGTTACTTAGTCCGAACCCCTGTAGCTTGTG
 GTCTATGCTCTTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTG
 TCCTACATATC

[0493] SEQ ID NO:208

30 ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGGATGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGTGAT
 35 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGATAGGTTTA
 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTG
 ACGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGTGGATATCATTGTCGGCATCAGGAGCACGTGCTGCCTATGATCCC
 40 AAGCGGTGGTGTCTTTAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTATCAACTACTGGGGTTCAACTGTGACCATGCGTTCTAGTT
 GCTTGTTCATTATATAAGCTTGTGTTACTTAGTCCGAACCGTGTAGTTTGTG
 GTCTCTGTTCTTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTG
 45 TCCTACATATC

[0494] SEQ ID NO:209

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 50 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGGATGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGATAGGTTTA
 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTG

5 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGTGGATATCATTGTCGGCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTCTTTAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTATCAACTACTGGGGGTCAACTGTGAACCATGCGTTCTAGTT
 GCTTGTTCATTATATAAGCTTGTACTTAGTCCGAACCGTGTAGTTTGTA
 GTCTCTGTCTCTTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTG
 TCCTACATATC

10 [0495] SEQ ID NO:210
 ACCAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGGATTGGGTACCAAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTGTGACATTGATGGTGTGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 15 CAAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCAGTGAGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGTGGATATCATAGTCCGCATCAGGAGCACGTGCTGCCTATGATCC
 20 CAAGCGGTGGTCTTCAGGACATGATGGAGGGTGTGGCAGGACCTCGT
 ACTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
 CCCCGTGTGTATCAACTACTAGGGGTCAACTGTGAGCCATGCGTTCTAGTT
 TGCTTGTTCATTATATAAGCTTGTATTACTAGTCCGAACCTGTAGTTGTA
 GTCTATGTTCTCTTTGTAGGGATGTGCTGTATAAGATGTCATGCAAGTTCTG
 25 TCCTACATATC

[0496] SEQ ID NO:211
 ACCAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGGATTGGGTACCAAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTGTGACATTGATGGTGTGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 CAAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCAGTGAGTG
 35 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGTGGATATCATAGTCCGCATCAGGAGCACGTGCTGCCTATGATCC
 CAAGCGGTGGTCTTCAGGACATGATGGAGGGTGTGGCAGGACCTCGT
 ACTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
 CCCCGTGTGTATCAACTACTAGGGGTCAACTGTGAGCCATGCGTTCTAGTT
 40 TGCTTGTTCATTATATAAGCTTGTATTACTAGTCCGAACCTGTAGTTGTA
 GTCTATGTTCTCTTTGTAGGGATGTGCTGTATAAGATGTCATGCAAGTTCTG
 TCCTACATATC

[0497] SEQ ID NO:212
 45 ACCAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGGATTGGGTACCAAGCTGCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTTACAGTGTGACATTGATGGAGATGGTAGTTCTCAT
 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATG
 ATATTGAACAACCAGCATCTGGGAATGGTGGTGCATGGAGGGATAGGTTTAC
 50 AAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGGC
 CATACTTGTGGATATCATCGTCCGCATCAGGAGCACGTGCTGCCTATGATCC
 AatGGcGGcGCTTCAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTACTG

AAATGGTCCGAAGGTCCACGCCAACACTACGAGTATGATCCCAAGCGGTGGTG
 CTTCAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTACTGAAATTGCA
 CCTACAAGACCTACAAGTGTGACATGCCAACATCAGCATGGTGCCCGTGTGTA
 5 TCAACTACTAGGGGTTCAACTGTGACCATGCCGTTCTAGTTGCTGTTCTATT
 CATATAAGCTTGTGTTACTTAGTCCGAACCCCTGTAGCTTGTAGTCTATGCTCTC
 TTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTGTCCATACATAC

[0498] SEQ ID NO:213

10 ACCAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTCTGGTTGGGGCAATGGGATTGGGTTACAGCTGCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTTACAGTTGTGACATTGATGGAGATGGTAGTTCTCAT
 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGGATGAT
 ATATTGAACAACCAGCATCTGGGAATGGTGGTGCAATGGGAGGAGAGTTTAC
 15 AAGGCCAACCGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTA
 CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACATGCTGAGACCCCCAGGGC
 CATACTGTTGGATATCATGTCCTCGCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTGTGCTTCAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTA
 20 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCCAACATCAGCATGGTGC
 CCGCGTGTGTTGATCAACTACTAGGGGTTCAACTGTGACCATGCCGTTCTAGTT
 GCTTGTGTTCATCATATAAGCTTGTGTTACTTAGTCCGAACCCCTGTAGCTTGTGTA
 GTCTATGCTCTTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTG
 TCCTACATAC

[0499] SEQ ID NO:214

25 ATCAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGTGACATTGATGGGATGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGTGAT
 30 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAAGTGGGAGGAGAGTTTA
 CAAGGCCAACCGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTA
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACATGCTGAGACCCCCAGGG
 CCATACTGTTGGATATCATTGTCCTCGCATCAGGAGCACGTGCTGAGTTGGCGG
 35 CGCTTCAAGGACATGATCATGGAGGGTGTGKCAAGGACCTCGTACTGAAATGGT
 CCGAAGGTCCACGCCCTCGTATGAAATGGTCCGAAGGTCCACGCCAACACTACG
 AGNNNNNNNNNNNNNNNNNNNNNNNNNNNTATGATTCCCAATGGCGGCTT
 CCCAATGGCGCGCTTCAAGGACATGATCATGGAGGGTGTGGCAGGACCTCG
 TACTGAAATGGTCCGAAGGTCCACGCCAACACTACGATGATCCCAAGCGGTGGT
 40 GCTTTAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTACTGAAATTGCG
 ACCTACAAGACCTACAAGTGTGACATGCCAACATCAGCATGATACCTGCGTGTG
 ATCAACTACTGGGGGTTCAACTGTGACCATGCCGTTCTAGTTGCTTGTGTTCT
 TCATATAAGCTTGTGTTACTTAGTCCGAACCGTGTAGTTGTAGTCTCTGTCT
 CTTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTGTCCATACATAC

[0500] SEQ ID NO:215

45 ATCAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGTGACATTGATGGGATGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAAGTGGGAGGAGAGTTTA
 CAAGGCCAACCGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTA
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACATGCTGAGACCCCCAGGG

CCATACTTGGATATCATTGCCCCCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTCTTAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTA
 CTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 5 CTGCGTGTGACTCAACTACTGGGGTTCAACTGTGACATGCGTCTAGTT
 GCTTGTTCATTCATATAAGCTTGTACTTAGTCCGAACCGTGTAGTTTGT
 GTCTCTGTTCTCTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTG
 TCCTACATATC

[0501] SEQ ID NO:216

10 ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGGATTGGGTACAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTATGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTCAGTGGGAGGGATAGGTTTA
 15 CAAGGCCAACCGGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCAGTG
 ACGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCCAGGG
 CCATACTTGGATATCATAGTCCGCATCAGGAGCAGTGCTGCCTATGATCC
 eAatGGcGGcGCTTCAAGGACATGATGGAGGGTATGGCAGGACCTCGTACT
 20 GAAATGGTCCGAAGGTCCACGCCAACACTACGAGTATGATCCAAAGCGGTGGT
 GCTTCAAGGACATGATGGAGGGTATGGCAGGACCTCGTACTGAAATTG
 ACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATGCCCGTGT
 ATCAACTACTAGGGGTCAACTGTGAGCCATGCGTTCTAGTTGCTGT
 25 TCATATAAGCTTGTATTACTTAGTCCGAACCTGTAGTTGTAGTCTATGTT
 CTTTGTAGGGATGTGCTGTATAAGATGTCATGCAAGTTCTGCTACATATC

[0502] SEQ ID NO:217

30 ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGGATTGGGTACAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTATGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTCAGTGGGAGGGATAGGTTTA
 CAAGGCCAACCGGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCAGTG
 35 ACGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCCAGGG
 CCATACTTGGATATCATAGTCCGCATCAGGAGCAGTGCTGCCTATGATT
 CCAATGGCGCGCTTCAGGACATGATGGAGGGTATGGCAGGACCTCGT
 ACTGAAAATGGTCCGAAGGTCCACGCCACCTCGTACTGAAATGGTCCRAAG
 GTCCACGCCAACACTACGAGTATGATCCAAAGCGGTGGTCTCAAGGACATG
 40 ATCATGGAGGGTATGGCAGGACCTCGTACTGARATTGACCTACAAGACCTAC
 AAGTGTGACATGCGCAATCAGCATGATGCCCGTGTGTATCAACTACTAGGGG
 TTCAACTGTGAGCCATGCGTTCTAGTTGCTTGTTCATTGATAAGCTTGTAT
 TACTTAGTCCGAACCTGTAGTTGTAGTCTATGTTCTTTGTAGGGATGTG
 CTGTCATAAGATGTCATGCAAGTTCTGCTACATATC

45

[0503] SEQ ID NO:218

50 ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGGATTGGGTACAGCTGCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTTACAGTTGACATTGATGGAGATGGTAGTTCTCAT
 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTAAGGGTGT
 ATATTGAACAACCAGCATCTGGGAATGGTGGTCATGGGAGGGATAGGTTTAC
 AAGGCCAACCGGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTA

CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTGAGACCCCAGGGC
 CATACTTGGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCc
 AatGGcGGcGCTTCAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTACTG
 5 AAATGGTCCGAAGGTCCACGCCCAACTACGAGTATGATCCCAAGCGGTGGTG
 CTTCAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTACTGAAATTGCA
 CCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGCCCGCTGTGTA
 TCAACTACTAGGGGTTCAACTGTGAAACCATGCGTTCTAGTTGCTGTTCATT
 CATATAAGCTTGTGTTACTTAGTCCGAACCCCTGTAGCTTGTAGTCTATGCTCTC
 10 TTTGTAGGGATGTGCTGTACATAAGATATCATGCAAGTTCTGTCCATACATC

[0504] SEQ ID NO:219

ACAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTGGGGCAATGGGATTGGGTTACAGCTGCAGCTGGCGCTGCTG
 15 TGGCCAACCCAGGTGTACAGTTGACATTGAGAACCTCCCTGTGAAGGTGATG
 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATG
 ATATTGAACAACCAGCATCTGGGAATGGTGGTGCAATGGGAGGATAGGTTTAC
 AAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTA
 20 CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTGAGACCCCAGGGC
 CATACTTGGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTCTTCAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTA
 CTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTG
 25 CCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGAAACCATGCGTTCTAGTT
 GCTTGTTCATTCAATAAGCTTGTGTTACTTAGTCCGAACCCCTGTAGCTTGTGTA
 GTCTATGCTCTTTGTAGGGATGTGCTGTACATAAGATATCATGCAAGTTCTG
 TCCTACATATC

[0505] SEQ ID NO:220

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 30 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTACAGTTGACATTGATGGGATGGTAGTTCCCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCACTGGGAGGATAGGTTTA
 35 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTA
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTGAGACCCCAGGG
 CCATACTTGGATATCATTGCCCCGATGCTTTAAGGACATGATCATGGAGG
 GTGATGGCAGGACCTCGTACTGAAATTGACCTACAAGACCTACAAGTGTGACA
 40 TCGCAATCAGCATGATACCTGCGTGTGTTACTCAACTACTGGGGTTCAACTGTG
 AACCATGCGTTCTAGTTGCTGTTCTTCAATAAGCTTGTGTTACTTAGTTC
 CGAACCGTGTAGTTGTAGTCTCTGTTAGGGATGTGCTGTACATAAG
 ATATCATGCAAGTTCTGTCCATACATATC

[0506] SEQ ID NO:221

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 45 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTACAGTTGACATTGATGGGATGGTAGTTCCCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCACTGGGAGGATAGGTTTA
 50 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTA
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTGAGACCCCAGGG
 CCATACTTGGATATCATTGCCCCGATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTCTTAAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTA

CTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTATCAACTACTGGGGGTTCAACTGTGAACCATGCGTTCTAGTT
 GCTTGTTCATTATAGCTGTACTAGTCCGAACCGTGTAGTTGTAG
 5 TCCTACATATC

[0507] SEQ ID NO:222

ACCAAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 10 TTCGCTGGTTGGGGCAATGGGATTGGGTTACCAAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 CAAGGCCAACGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGA
 15 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCAGTGAGTG
 ACGAAGAAGAGCGAAGTCAGCAATCAAGAACATGCTGAGACCCCCAGGG
 CCATACTTGGATATCATAGTCCGCATCAGGAGCACGTGCTGCCTATGATCC
 CAAGCGGTGGTGCCTCAAGGACATGATCATGGAGGGTGTAGGGCAGGACCTCGT
 ACTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
 20 CCCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGAGGCCATGCGTTCTAGTT
 TGCTTGTTCATTATAGCTGTATTACTAGTCCGAACCTGTAGTTGTAG
 GTCTATGTTCTCTTTGTAGGGATGTGCTGTATAAGATGTCATGCAAGTTCTTG
 TCCTACATATC

[0508] SEQ ID NO:223

ACCAAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 25 TTCGCTGGTTGGGGCAATGGGATTGGGTTACCAAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 CAAGGCCAACGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGA
 30 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCAGTGAGTG
 ACGAAGAAGAGCGAAGTCAGCAATCAAGAACATGCTGAGACCCCCAGGG
 CCATACTTGGATATCATAGTCCGCATCAGGAGCACGTGCTGCCTATGATCC
 CAAGCGGTGGTGCCTCAAGGACATGATCATGGAGGGTGTAGGGCAGGACCTCGT
 35 ACTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
 CCCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGAGGCCATGCGTTCTAGTT
 TGCTTGTTCATTATAGCTGTATTACTAGTCCGAACCTGTAGTTGTAG
 GTCTATGTTCTCTTTGTAGGGATGTGCTGTATAAGATGTCATGCAAGTTCTTG
 TCCTACATATC

40

[0509] SEQ ID NO:224

ACCAAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 45 TTCGCTGGTTGGGGCAATGGGATTGGGTTACCAAGCTGCAGCTGGCGCTGCTG
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGAGATGGTAGTTCTCAT
 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCAGTGAAGGTGATG
 ATATTGAACAACCAGCATCTGGGAATGGTGGTGCATGGGAGGGATAGGTTAC
 AAGGCCAACGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGA
 50 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCAGCAATCAAGAACATGCTGAGACCCCCAGGGC
 CATACTTGGATATCATCGTCCGCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTGCCTCAAGGACATGATCATGGAGGGTGTAGGGCAGGACCTCGTA
 CTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGC

CCCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGAACCATGCGTTCTAGTT
GCTTGTTCATTATATAAGCTTGTACTTAGTCCGAACCCTGTAGCTTGT
GTCTATGCTCTTTGTAGGGATGTGCTGATAAGATATCATGCAAGTTCTG
TCCTACATATC

5

[0510] SEQ ID NO:225

ACCAAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
TTCGTCGGTTGGGGCAATGGGATTGGGTTACAGCTGCAGCTGGCGCTGCTG
TGGCCAACCCAGGTGTTACAGTTGACATTGATGGAGATGGTAGTTCTCAT
10 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGGTATG
ATATTGAACAACCAGCATCTGGGAATGGTGGTGCAATGGGAGGATAGGTTTAC
AAGGCCAACGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
15 CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACAGATGCTTGAGACCCCCAGGGC
CATACTTGGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCC
AAGCGGTGGTCTTCAAGGACATGATCATGGAGGGTGATGGCAGGACCTCGTA
CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGC
CCCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGAACCATGCGTTCTAGTT
20 GCTTGTTCATTATATAAGCTTGTACTTAGTCCGAACCCTGTAGCTTGT
GTCTATGCTCTTTGTAGGGATGTGCTGATAAGATATCATGCAAGTTCTG
TCCTACATATC

[0511] SEQ ID NO:226

ATCAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
25 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGGATGGTAGTTCTCAT
TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGTATG
GATATTGAACAACCAGCATCTGGGAATGGTGGTGAGTGGGAGGATAGGTTTAA
CAAGGCCAACGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
30 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTG
ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACAGATGCTTGAGACCCCCAGGG
CCATACTTGGATATCATGGTCCCGCATCAGGAGCACGTGCTGCCTATGATCC
cAatGGcGGcGCTTCAAGGACATGATCATGGAGGGTGATGGCAGGACCTCGTACT
GAAATGGTCCGAAGGTCCAGGCCAAGTACAGAGTATGATCCCAAGCGGTGGT
35 GCTTTAAGGACATGATCATGGAGGGTGATGGCAGGACCTCGTACTGAAATTTCG
ACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATACCTGCGTGT
ATCAACTACTGGGGTTCAACTGTGAACCATGCGTTCTAGTTGCTTGT
TCATATAAGCTTGTACTTAGTCCGAACCGTGTAGTTGTAGTCTCTGTTCT
CTTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTGCTACATATC
40

[0512] SEQ ID NO:227

ATCAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGGATGGTAGTTCTCAT
45 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGTATG
GATATTGAACAACCAGCATCTGGGAATGGTGGTGAGTGGGAGGATAGGTTTAA
CAAGGCCAACGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTG
ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACAGATGCTTGAGACCCCCAGGG
50 CCATACTTGGATATCATGGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCC
AAGCGGTGGTCTTAAAGGACATGATCATGGAGGGTGATGGCAGGACCTCGTA
CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
CTGCGTGTGTATCAACTACTGGGGTTCAACTGTGAACCATGCGTTCTAGTT
GCTTGTTCATTATATAAGCTTGTACTTAGTCCGAACCGTGTAGTTGT
195

GTCTCTGTTCTTTGAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTTG
TCCTACATATC

[0513] SEQ ID NO:228

5 ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
TTCGTCGGTTGGGGCAATGGGATTGGGTACAGCTGCAGCTGGCGCTGCT
GTGGCCAACCCAGGTACAGTTGACATTGATGGTGATGGTAGTTCTCA
TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
GATATTGAACAACCAGCATCTGGGAATGGTGGTGAGTGGGAGGATAGGTTTA
10 CAAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCAGTG
ACGAAGAACAGCGAAGTCAGCAATCAAGAACAGATGCTGAGACCCCCAGGG
CCATACTTGGATATCATCGTCCGCATCAGGAGCAGTGCTGCCTATGATCC
15 CAAGCGGTGGTGCTTCAAGGACATGATCATGGAGGGTGATGGCAGGACCTCGT
ACTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTG
CCCGCGTGTGTATCAACTACTAGGGGTCAACTGTGAGGCCATGCGTTCTAGTT
TGCTTGTTCATTCAATAAGCTTGTATTACTTAGTCCGAACCCGTAGTTGT
GTCTATGTTCTTTGAGGGATGTGCTGTCATAAGATGTCATGCAAGTTCTTG
TCCTACATATC

20

[0514] SEQ ID NO:229

ACCGAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
TTCGTCGGTTGGGGCAATGGGATTGGGTACAGCTGCAGCTGGCGCTGCT
GTGGCCAACCCAGGTACAGTTGACATTGATGGTGATGGTAGTTCTCA
25 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
GATATTGAACAACCAGCATCTGGGAATGGTGGTGAGTGGGAGGATAGGTTTA
CAAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCAGTG
ACGAAGAACAGCGAAGTCAGCAATCAAGAACAGATGCTGAGACCCCCAGGG
30 CCATACTTGGATATCATCGTCCGCATCAGGAGCAGTGCTGCCTATGATCC
CAAGCGGTGGTGCTTCAAGGACATGATCATGGAGGGTGATGGCAGGACCTCGT
ACTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTG
CCCGCGTGTGTATCAACTACTAGGGGTCAACTGTGAGGCCATGCGTTCTAGTT
TGCTTGTTCATTCAATAAGCTTGTATTACTTAGTCCGAACCCGTAGTTGT
35 GTCTATGTTCTTTGAGGGATGTGCTGTCATAAGATGTCATGCAAGTTCTTG
TCCTACATATC

30

[0515] SEQ ID NO:230

ACCGAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
TTCGTCGGTTGGGGCAATGGGATTGGGTACAGCTGCAGCTGGCGCTGCTG
TGGCCAACCCAGGTACAGTTGACATTGATGGAGATGGTAGTTCTCAT
GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCGTGAAGGTGATG
ATATTGAACAACCAGCATCTGGGAATGGTGGTCAATGGGAGGATAGGTTAC
AAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
45 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTG
CGAAGAACAGCGAAGTCAGCAATCAAGAACAGATGCTGAGACCCCCAGGGC
CATACTTGGATATCATCGTCCGCATCAGGAGCAGTGCTGCCTATGATtCCc
AatGGcGGcGCTTCAAGGACATGATCATGGAGGGTGATGGCAGGACCTCGTACTG
AAATGGTCCGAAGGTCCACGCCAACACTACGAGATGATCCCAAGCGGTGGTG
50 CTTTCAAGGACATGATCATGGAGGGTGATGGCAGGACCTCGTACTGAAATTTCGA
CCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGGCCCGTGGTGT
TCAACTACTAGGGGTCAACTGTGAACCATGCGTTCTAGTTGCTTGTTCATT
CATATAAGCTTGTACTTAGTCCGAACCCGTAGCTTGTAGCTATGCTCTC
TTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTGCTACATATC

[0516] SEQ ID NO:231

ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTCTGGTTGGGGCAATGGGATTGGGTTACCGCTGCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTTACAGTTGTTGACATTGATGGAGATGGTAGTTCTCAT
 5 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATG
 ATATTGAACAACCAGCATCTGGGAATGGTGGTGCAATGGGAGGATAGGTTTAC
 AAGGCCAACGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGACGATTGCTAACAGGATTCAACGTTCCGGCAGTCGTGTA
 10 CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACATGCTTGAGACCCCAGGGC
 CATACTGTTGGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTCTTCAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGC
 15 CCGCGTGGTGTATCAACTACTAGGGGTTCAACTGTGAACCATGCGTTCTAGTT
 GCTTGTTCATTCAATATAAGCTTGTGTTACTAGTCCGAACCGTGTAGCTTGTGTA
 GTCTATGCTCTTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTTG
 TCCTACATATC

[0517] SEQ ID NO:232

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 20 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGTTGACATTGATGGGATGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAAGTGGAGGATAGGTTTA
 25 CAAGGCCAACGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGACGATTGCTAACAGGATTCAACGTTCCGGCAGTCGTGTA
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACATGCTTGAGACCCCAGGG
 CCATACTGTTGGATATCATTGTCGGCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTCTTAAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTA
 30 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTTGATCAACTACTGGGGTTCAACTGTGAACCATGCGTTCTAGTT
 GCTTGTTCATTCAATATAAGCTTGTGTTACTAGTCCGAACCGTGTAGTTTGTGTA
 GTCTCTGTTCTTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTTG
 TCCTACATATC

[0518] SEQ ID NO:233

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGTTGACATTGATGGGATGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGTGAT
 40 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAAGTGGAGGATAGGTTTA
 CAAGGCCAACGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGACGATTGCTAACAGGATTCAACGTTCCGGCAGTCGTGTA
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACATGCTTGAGACCCCAGGG
 CCATACTGTTGGATATCATTGTCGGCATCAGGAGCACGTGCTGCCTATGATCCC
 45 AAGCGGTGGTCTTAAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTTGATCAACTACTGGGGTTCAACTGTGAACCATGCGTTCTAGTT
 GCTTGTTCATTCAATATAAGCTTGTGTTACTAGTCCGAACCGTGTAGTTTGTGTA
 GTCTCTGTTCTTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTTG
 50 TCCTACATATC

[0519] SEQ ID NO:234

ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTCTGGTTGGGGCAATGGGATTGGGTTACCGCTGCAGCTGGCGCTGCT

GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTGTAGGGTAGTTCCCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGATAGGTTTA
 5 CAAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCAGTGAGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCCAGGG
 CCATACTTGTGGATATCATAGTCCCGCATCAGGAGCACGTGCTGCCTATGATCC
 10 CAAGCGGTGGTGCCTCAAGGACATGATCATGGAGGGTGTAGGGCAGGACCTCGT
 ACTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
 CCCCGTGTGTATCAACTACTAGGGTTCAACTGTGAGGCCATGCGTTTAGTT
 TGCTTGTTCATTCATATAAGCTGTATTACTTAGTCCGAACCCCTGTAGTTGT
 GTCTATGTTCTTTGTAGGGATGTGCTGTATAAGATGTCATGCAAGTTCTG
 TCCTACATATC

15 [0520] SEQ ID NO:235
 ACCAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGGATTGGTACCGAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTGTAGGGTAGTTCCCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 20 GATATTGAACAAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGATAGGTTTA
 CAAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCAGTGAGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCCAGGG
 CCATACTTGTGGATATCATAGTCCCGCATCAGGAGCACGTGCTGCCTATGATCC
 25 CAAGCGGTGGTGCCTCAAGGACATGATCATGGAGGGTGTAGGGCAGGACCTCGT
 ACTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
 CCCCGTGTGTATCAACTACTAGGGTTCAACTGTGAGGCCATGCGTTTAGTT
 TGCTTGTTCATTCATATAAGCTGTATTACTTAGTCCGAACCCCTGTAGTTGT
 GTCTATGTTCTTTGTAGGGATGTGCTGTATAAGATGTCATGCAAGTTCTG
 30 TCCTACATATC

[0521] SEQ ID NO:236
 ACCAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGGATTGGTACCGAGCTGCAGCTGGCGCTGCT
 35 TGGCCAACCCAGGTGTTACAGTTGACATTGATGGAGATGGTAGTTCCCTCAT
 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATG
 ATATTGAACAAACCAGCATCTGGGAATGGTGGTGCATGGAGGGATAGGTTTAC
 AAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 40 CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCCAGGGC
 CATACTTGTGGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTGTTCATTCAAGGACATGATCATGGAGGGTGTAGGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGC
 CCGCGTGTGTATCAACTACTAGGGTTCAACTGTGACCATGCGTTTAGTT
 45 GCTTGTTCATTCATATAAGCTGTGTACTTAGTCCGAACCCCTGTAGTTGT
 GTCTATGCTCTTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTG
 TCCTACATATC

[0522] SEQ ID NO:237
 50 ACCAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGGATTGGTACCGAGCTGCAGCTGGCGCTGCT
 TGGCCAACCCAGGTGTTACAGTTGACATTGATGGAGATGGTAGTTCCCTCAT
 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATG
 ATATTGAACAAACCAGCATCTGGGAATGGTGGTGCATGGAGGGATAGGTTTAC

5 AAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGGC
 CATACTTGGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCC
 10 AAGCGGTGGTCTTCAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGC
 CCGCGTGTGTATCAACTACTAGGGTTCAACTGTGAACCATGCGTTCTAGTT
 GCTTGTTCATTATAGCTTGTTACTTAGTCCGAACCCTGTAGCTTGTAG
 GTCTATGCTCTTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTTG
 TCCTACATATC

[0523] SEQ ID NO:238
 ATCAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 15 TTCACTCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGGATGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 20 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGGATATCATTGCCCCCATCAGGAGCACGTGCTGCCTATGATCCC
 25 AAGCGGTGGTCTTAAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTATCAACTACTGGGGTTCAACTGTGAACCATGCGTTCTAGTT
 GCTTGTTCATTATAGCTTGTTACTTAGTCCGAACCCTGTAGTTGTAG
 GTCTCTGTTCTCTTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTTG
 TCCTACATATC

[0524] SEQ ID NO:239
 ATCAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 30 TTCACTCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGGATGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 35 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGGATATCATTGCCCCCATCAGGAGCACGTGCTGCCTATGATCCC
 40 AAGCGGTGGTCTTAAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTATCAACTACTGGGGTTCAACTGTGAACCATGCGTTCTAGTT
 GCTTGTTCATTATAGCTTGTTACTTAGTCCGAACCCTGTAGTTGTAG
 GTCTCTGTTCTCTTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTTG
 TCCTACATATC

45 [0525] SEQ ID NO:240
 ACCAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTGCTCTGGTTGGGGCAATGGGATTGGGTTACAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTGTGGTAGTTCTCA
 50 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCAGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG

CCATACTTGTGGATATCATAGCCCCCATCAGGAGCACGTGCTGCCTATGATCC
 CAAGCGGTGGTCTTCAGGACATGATGGAGGGTATGGCAGGACCTCGT
 ACTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
 5 CCCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGAGGCCATGCGTTTCTAGTT
 TGCTTGTTCATTCATATAAGCTTGTATTACTTAGTTCCGAACCTGTAGTTTGTA
 GTCTATGTTCTCTTTGTAGGGATGTGCTGTATAAGATGTCATGCAAGTTCTTG
 TCCTACATATC

[0526] SEQ ID NO:241

10 ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGTCGGTTGGGGCAATGGGATTGGGTTACCGAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTACAGTTGACATTGATGGTATGGTAGTTCCCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGATAGGTTTA
 15 CAAGGCCAATCGGGCGCACACATACCTTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTTCGAGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGTGGATATCATAGCCCCCATCAGGAGCACGTGCTGCCTATGATCC
 20 CAAGCGGTGGTGTCTTCAGGACATGATGGAGGGTATGGCAGGACCTCGT
 ACTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
 CCCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGAGGCCATGCGTTTCTAGTT
 TGCTTGTTCATTCATATAAGCTTGTATTACTTAGTTCCGAACCTGTAGTTTGTA
 GTCTATGTTCTCTTTGTAGGGATGTGCTGTATAAGATGTCATGCAAGTTCTTG
 25 TCCTACATATC

[0527] SEQ ID NO:242

ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGTCGGTTGGGGCAATGGGATTGGGTTACCGAGCTGCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTACAGTTGACATTGATGGAGATGGTAGTTCCCTCAT
 30 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATG
 ATATTGAACAACCAGCATCTGGGAATGGTGGTGCATGGAGGATAGGTTTAC
 AAGGCCAATCGGGCGCACACATACCTTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGGC
 35 CATACTTGTGGATATCATGTCGGCATCAGGAGCACGTGCTGCCTATGATCC
 AatGGcGGcGCTTCAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTACTG
 AAATGGTCCGAAGGTCCACGCCCAACTACGAGTATGATCCCAAGCGGTGGT
 CTTCAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTACTGAAATTCA
 40 CCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTCCCCGTGTGTA
 TCAACTACTAGGGGTTCAACTGTGACCATGCGTTTCTAGTTGCTTGTTCATT
 CATATAAGCTTGTGTACTTAGTTCCGAACCTGTAGCTTGTAGTCTATGCTCTC
 TTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTGTCCATATC

[0528] SEQ ID NO:243

45 ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGTCGGTTGGGGCAATGGGATTGGGTTACCGAGCTGCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTACAGTTGACATTGATGGAGATGGTAGTTCCCTCAT
 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATG
 ATATTGAACAACCAGCATCTGGGAATGGTGGTGCATGGAGGATAGGTTTAC
 50 AAGGCCAATCGGGCGCACACATACCTTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGGC
 CATACTTGTGGATATCATGTCGGCATCAGGAGCACGTGCTGCCTATGATCC
 AatGGcGGcGCTTCAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTACTG

AAATGGTCCGAAGGTCCACGCCAACACTACGAGTATGATCCCAAGCGGTGGTG
 CTTCAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTACTGAAATTGCA
 CCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGCCCGTGTGTA
 5 TCAACTACTAGGGGTTCAACTGTGAACCATGCGTTTCTAGTTGCTTGTTCATT
 CATATAAGCTTGTGTTACTTAGTCCGAACCCCTGTAGCTTGTAGTCTATGCTCTC
 TTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTGTCCATACATAC

[0529] SEQ ID NO:244

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 10 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGGATGGTAGTTCCCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGTGAT
 GATATTGAACAAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTA
 15 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTTGTGTC
 ACGAAGAAGAGCGAAGTCAGCAATCAAGAACATGCTTGAGACCCCAGGG
 CCATACTTGTGGATATCATTGTCCCGCATCAGGAGCACGTGCTGCCtGcGCTTTC
 AAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTACTGAAATGGTCCGAAG
 20 GTCCACGCCGCCAACACTACGAGTATGATCCCAAGCGGTGGTGTGTTAAGGACATG
 ATCATGGAGGGTGTGGCAGGACCTCGTACTGAAATTGACCTACAAGACCTAC
 AAGTGTGACATGCGCAATCAGCATGATAACCTGCGTGTGTTATCAACTACTGGGGG
 TTCAACTGTGAACCATGCGTTTCTAGTTGCTTGTTCATTGATATAAGCTTGTG
 TACTTAGTCCGAACCGTGTAGTTGTAGTCTCTGTTGAGGGATGTG
 25 CTGTCATAAGATATCATGCAAGTTCTGTCCATACATAC

[0530] SEQ ID NO:245

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGGATGGTAGTTCCCTCA
 30 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGTGAT
 GATATTGAACAAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTA
 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTTGTGTC
 ACGAAGAAGAGCGAAGTCAGCAATCAAGAACATGCTTGAGACCCCAGGG
 35 CCATACTTGTGGATATCATTGTCCCGCATCAGGAGCACGTGCTGCCATGATCCC
 AAGCGGTGGTGTGTTAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTA
 CTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTTGATCAACTACTGGGGTTCAACTGTGAACCATGCGTTTCTAGTT
 40 GCTTGTGTTGATGACATGCGTGTGTTACTTAGTCCGAACCGTGTAGTTGTA
 GTCTCTGTTCTTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTG
 TCCTACATATC

[0531] SEQ ID NO:246

ACCAAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 45 TTCGTCCTGGTTGGGGCAATGGGATTGGGTTACAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTAGGGTAGTTCCCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCAGTGAAGGTGAT
 GATATTGAACAAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTA
 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 50 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGCAGTTCGAGTG
 ACGAAGAAGAGCGAAGTCAGCAATCAAGAACATGCTTGAGACCCCAGGG
 CCATACTTGTGGATATCATAGTCCCGCATCAGGAGCACGTGCTGCCATGATCC
 CAAGCGGTGGTGTGTTCAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGT
 ACTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG

CCCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGAGCCATGCGTTCTAGTT
TGCTTGTTCATTCAATATAAGCTTGTATTACTTAGTCCGAACCCTGTAGTTGT
GTCTATGTTCTCTTTGTAGGGATGTGCTGTCATAAGATGTCATGCAAGTTCTG
TCCTACATATC

5

[0532] SEQ ID NO:247

ACCAAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
TTCGCTCTGGTTGGGGCAATGGGATTGGGTACCAAGCTGCAGCTGGCGCTGCTG
TGGCCAACCCAGGTGTTACAGTTGACATTGATGGTAGGGTAGTTCCCTCA
TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCAGTGAAGGTGAT
GATATTGAACAACCAGCATCTGGGAATGGTGGTCAGTGGGAGGGATAGGTTTA
CAAGGCCAACCGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
ATATCCAGATTGTGACGATTGCTAACAGGATTCAACGTTCCAGCAGTCAGTG
ACGAAGAAGAGCGAAGTCAGCAATCAAGAACAGATGCTTGAGAACCCCCAGGG
CCATACTTGTGGATATCATAGTCCCACAGGAGCACGTGCTGCCTATGATCC
CAAGCGGTGGTGCCTTCAAGGACATGATCATGGAGGGTAGGGCAGGACCTCGT
ACTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
CCCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGAGCCATGCGTTCTAGTT
TGCTTGTTCATTCAATATAAGCTTGTATTACTTAGTCCGAACCCTGTAGTTGT
GTCTATGTTCTCTTTGTAGGGATGTGCTGTCATAAGATGTCATGCAAGTTCTG
TCCTACATATC

[0533] SEQ ID NO:248

ACCAAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
TTCGCTCTGGTTGGGGCAATGGGATTGGGTACCAAGCTGCAGCTGGCGCTGCTG
TGGCCAACCCAGGTGTTACAGTTGACATTGATGGAGATGGTAGTTCCCTCAT
GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATG
ATATTGAACAACCAGCATCTGGGAATGGTGGTCATGGAGGGATAGGTTTAC
AAGGCCAACCGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
TATCCAGATTGTGACGATTGCTAACAGGATTCAACGTTCCGGCAGTCGTGTGA
CGAAGAAGAGCGAAGTCAGCAATCAAGAACAGATGCTTGAGAACCCCCAGGGC
CATACTTGTGGATATCATCGTCCCACAGGAGCACGTGCTGCCTATGATCC
AatGGcGGcGCTTCAAGGACATGATCATGGAGGGTAGGGCAGGACCTCGTACTG
AAATGGTCCGAAGGTCCACGCCAACTACGAGTATGATCCCAAGCGGTGGTG
CTTCAAGGACATGATCATGGAGGGTAGGGCAGGACCTCGTACTGAAATTTCGA
CCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTCCCCCGTGTGTA
TCAACTACTAGGGGTTCAACTGTGAACCATGCGTTCTAGTTGCTGTTCTCATT
CATATAAGCTTGTGTTACTTAGTCCGAACCCTGTAGCTTGTAGTCTATGCTCTC
TTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTGCTACATATC

40

[0534] SEQ ID NO:249

ACCAAGATGTGGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
TTCGCTCTGGTTGGGGCAATGGGATTGGGTACCAAGCTGCAGCTGGCGCTGCTG
TGGCCAACCCAGGTGTTACAGTTGACATTGATGGAGATGGTAGTTCCCTCAT
GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATG
ATATTGAACAACCAGCATCTGGGAATGGTGGTCATGGAGGGATAGGTTTAC
AAGGCCAACCGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
TATCCAGATTGTGACGATTGCTAACAGGATTCAACGTTCCGGCAGTCGTGTGA
CGAAGAAGAGCGAAGTCAGCAATCAAGAACAGATGCTTGAGAACCCCCAGGGC
CATACTTGTGGATATCATCGTCCCACAGGAGCACGTGCTGCCTATGATCC
AAGCGGTGGTGCCTTCAAGGACATGATCATGGAGGGTAGGGCAGGACCTCGTA
CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGC
CCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGAACCATGCGTTCTAGTT
GCTTGTTCATTCAATATAAGCTTGTGTTACTTAGTCCGAACCCTGTAGCTTGT

GTCTATGCTCTTTGAGGGATGTGCTGCATAAGATATCATGCAAGTTCTG
TCCTACATATC

[0535] SEQ ID NO:250

5 ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGGATGGTAGTTCCCTCA
TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCCAGTGAAGGTGAT
GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGAGGGATAGGTTTA
10 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTC
ACGAAGAAGAGCGAAGTCAGTGCAGCAATCAAGAACAGATGCTGAGACCCCCAGGG
CCATACTTGTGGATATCATTGCCCCGATCAGGAGCAGTGCTGCCTATGATCCC
AAGCGGTGGTGTGCTTTAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTA
15 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
CTGCGTGTGTATCAACTACTGGGGGTTCAACTGTGAACCATGCGTTCTAGTT
GCTTGTTCATTATATAAGCTTGTGTTACTTAGTCCGAACCGTGTAGTTTGTA
GTCTCTGTTCTTTGAGGGATGTGCTGCATAAGATATCATGCAAGTTCTG
TCCTACATATC

20

[0536] SEQ ID NO:251

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
25 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGGATGGTAGTTCCCTCA
TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCCAGTGAAGGTGAT
GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGAGGGATAGGTTTA
CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTC
30 ACGAAGAAGAGCGAAGTCAGTGCAGCAATCAAGAACAGATGCTGAGACCCCCAGGG
CCATACTTGTGGATATCATTGCCCCGATCAGGAGCAGTGCTGCCTATGATCCC
AAGCGGTGGTGTGCTTTAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTA
CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
CTGCGTGTGTATCAACTACTGGGGGTTCAACTGTGAACCATGCGTTCTAGTT
35 GCTTGTTCATTATATAAGCTTGTGTTACTTAGTCCGAACCGTGTAGTTTGTA
GTCTCTGTTCTCTTTGAGGGATGTGCTGCATAAGATATCATGCAAGTTCTG
TCCTACATATC

40

[0537] SEQ ID NO:252

40 ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
TTCGTCTGGTTGGGGCAATGGGATTGGGTTACCAAGCTGCAGCTGGCGCTGCT
GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTGTAGTTCCCTCA
TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGAGGGATAGGTTTA
45 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCGAGTG
ACGAAGAAGAGCGAAGTCAGTGCAGCAATCAAGAACAGATGCTGAGACCCCCAGGG
CCATACTTGTGGATATCATAGTCCCCGATCAGGAGCAGTGCTGCCTATGATCC
CAAGCGGTGGTGTGCTTCAAGGACATGATGGAGGGTGTGGCAGGACCTCGT
50 ACTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
CCCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGAGGCCATGCGTTCTAGTT
TGCTTGTTCATTATATAAGCTTGTATTACTAGTCCGAACCGTGTAGTTTGTA
GTCTATGTTCTCTTTGAGGGATGTGCTGCATAAGATGTGATGCAAGTTCTG
TCCTACATATC

[0538] SEQ ID NO:253

ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 5 TTCGCTGGTTGGGGCAATGGGATTGGTTACAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAAACCAGCATCTGGGAATGGTGGTGCAGTGGAGGGATAGTTTCA
 10 CAAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGA
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCAGTG
 ACGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCCAGGG
 CCATACTTGTGGATATCATAGTCCCGCATCAGGAGCACGTGCTGCCTATGATCC
 15 CAAGCGGTGGTGTCTTCAGGACATGATCATGGAGGGTGTAGGGCAGGACCTCGT
 ACTGAAATTTCGACCTACAAGACCTAACAGTGTGACATGCGCAATCAGCATGATG
 CCCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGAGGCCATGCGTTTCTAGTT
 TGCTTGTCTCTTTGTAGGGATGTGCTGTATAAGATGTCATGCAAGTTCTTG
 GTCTATGCTCTCTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTTG
 TCCTACATATC

[0539] SEQ ID NO:254

ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 20 TTCGCTGGTTGGGGCAATGGGATTGGTTACAGCTGCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTTACAGTTGACATTGATGGAGATGGTAGTTCTCAT
 GAACATTCAAGGAGTTGGCATTGATCCGATTGAGAACCTCCCTGTGAAGGTGATG
 ATATTGAACAAACCAGCATCTGGGAATGGTGGTGCATGGAGGGATAGTTTAC
 25 AAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCCAGGGC
 CATACTTGTGGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCC
 30 AAGCGGTGGTGTCTTCAGGACATGATCATGGAGGGTGTAGGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGC
 CCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGAACCATGCGTTTCTAGTT
 GCTTGTCTCTTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTTG
 GTCTATGCTCTCTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTTG
 TCCTACATATC

35 [0540] SEQ ID NO:255

ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 40 TTCGCTGGTTGGGGCAATGGGATTGGTTACAGCTGCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTTACAGTTGACATTGATGGAGATGGTAGTTCTCAT
 GAACATTCAAGGAGTTGGCATTGATCCGATTGAGAACCTCCCTGTGAAGGTGATG
 ATATTGAACAAACCAGCATCTGGGAATGGTGGTGCATGGAGGGATAGTTTAC
 AAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCCAGGGC
 45 CATACTTGTGGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTGTCTTCAGGACATGATCATGGAGGGTGTAGGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGC
 CCGCGTGTGTATCAACTACTAGGGGTTCAACTGTGAACCATGCGTTTCTAGTT
 GCTTGTCTCTTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTTG
 GTCTATGCTCTCTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTTG
 TCCTACATATC

[0541] SEQ ID NO:256

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTACAGTTGACATTGATGGGATGGTAGTTCTCA
 5 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTG
 10 ACGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGTGGATATCATTGCCCCCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTGTCTTAAGGACATGATGGAGGGTATGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTATCAACTACTGGGGTTCAACTGTGAACCATGCGTTCTAGTT
 15 GCTTGTTCATTCAATATAAGCTTGTGTTACTTAGTCCGAACCGTGTAGTTGT
 GTCTCTGTTCTTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTG
 TCCTACATATC

[0542] SEQ ID NO:257

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTACAGTTGACATTGATGGGATGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 25 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTG
 ACGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGTGGATATCATTGCCCCCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTGTCTTAAGGACATGATGGAGGGTATGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 30 CTGCGTGTGTATCAACTACTGGGGTTCAACTGTGAACCATGCGTTCTAGTT
 GCTTGTTCATTCAATATAAGCTTGTGTTACTTAGTCCGAACCGTGTAGTTGT
 GTCTCTGTTCTTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTG
 TCCTACATATC

[0543] SEQ ID NO:258

ACCAAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGGATTGGGTACCAAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTACAGTTGACATTGATGGTGTAGGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 40 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 CAAGGCCATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCGAGTG
 ACGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGTGGATATCATAGTCCCCCATCAGGAGCACGTGCTGCCTATGATTC
 45 cAatGGcGGcGCTTCAGGACATGATGGAGGGTATGGCAGGACCTCGTACTGAAATTTCG
 GAAATGGTCCGAAGGTCCACGCCCAACTACGAGTATGATCCCAAGCGGTGGT
 GCTTCAAGGACATGATGGAGGGTATGGCAGGACCTCGTACTGAAATTTCG
 ACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATGCCCGCGTGT
 ATCAACTACTAGGGGTCAACTGTGAGCCATGCGTTCTAGTTGCTTGT
 50 TCATATAAGCTTGTATTACTTAGTCCGAACCCCTGTAGTTGTAGTCTATGTTCT
 CTTTGTAGGGATGTGCTGTCATAAGATGTGATGCAAGTTCTGCTACATATC

[0544] SEQ ID NO:259

ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTGCTCTGGTTGGGGCAATGGGATTGGGTTACAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTATGGTAGTTCTCA
 5 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCAGTGAAGGTGAT
 GATATTGAACAAACCAGCATCTGGGAATGGTGGCAGTGGGAGGATAGGTTTA
 CAAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCAGTGAGTG
 10 ACGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCCAGGG
 CCATACTTGGATATCATAGTCCCACAGGAGCACGTGCTGCCTATGATCC
 CAAGCGGTGGTCAAGGACATGATCATGGAGGGTATGGCAGGACCTCGT
 ACTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
 CCCGCGTGTGTATCAACTACTAGGGTTCAACTGTGAGCCATGCGTTCTAGTT
 TGCTTGTTCATTCAATAAGCTGTATTACTAGTCCGAACCCCTGTAGTTGT
 15 GTCTATGCTCTTTGTAGGGATGTGCTGCATAAGATGTCATGCAAGTTCTTG
 TCCTACATATC

[0545] SEQ ID NO:260

ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTGCTCTGGTTGGGGCAATGGGATTGGGTTACAGCTGCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTTACAGTTGACATTGATGGAGATGGTAGTTCTCAT
 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATG
 ATATTGAACAAACCAGCATCTGGGAATGGTGGTCAATGGGAGGATAGGTTTAC
 25 AAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCCAGGGC
 CATACTTGGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTGTTCAGGACATGATCATGGAGGGTATGGCAGGACCTCGTA
 30 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGC
 CCGCGTGTGTATCAACTACTAGGGTTCAACTGTGAACCATGCGTTCTAGTT
 GCTTGTTCATTCAATAAGCTGTGTACTTAGTCCGAACCCCTGTAGCTTGT
 GTCTATGCTCTTTGTAGGGATGTGCTGCATAAGATATCATGCAAGTTCTTG
 TCCTACATATC

35 [0546] SEQ ID NO:261

ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTGCTCTGGTTGGGGCAATGGGATTGGGTTACAGCTGCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTTACAGTTGACATTGATGGAGATGGTAGTTCTCAT
 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGTGATG
 40 ATATTGAACAAACCAGCATCTGGGAATGGTGGTCAATGGGAGGATAGGTTTAC
 AAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCCAGGGC
 CATACTTGGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCC
 45 AAGCGGTGGTGTTCAGGACATGATCATGGAGGGTATGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGC
 CCGCGTGTGTATCAACTACTAGGGTTCAACTGTGAACCATGCGTTCTAGTT
 GCTTGTTCATTCAATAAGCTGTGTACTTAGTCCGAACCCCTGTAGCTTGT
 GTCTATGCTCTTTGTAGGGATGTGCTGCATAAGATATCATGCAAGTTCTTG
 50 TCCTACATATC

[0547] SEQ ID NO:262

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTGCTCTGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT

GTGGCCAACCCAGGTACAGTTGACATTGATGGGGATGGTAGTTCCCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAAACCAGCATCTGGGAATGGTGGTCAGTGGGAGGGATAGGTTA
 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 5 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTG
 ACGAAGAAGAGCGAAGTCAGCAATCAAGAACATGCTGAGACCCCCAGGG
 CCATACTTGTGGATATCATTGTCCCCCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTCTTTAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTA
 10 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTATCAACTACTGGGGTTCAACTGTGAACCATGCGTTCTAGTT
 GCTTGTTCATTATAGCTGTGTTACTTAGTCCGAACCGTAGTTGTA
 GTCTCTGTTCTTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTG
 TCCTACATATC

15 [0548] SEQ ID NO:263
 ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCATCCGGTTGGGTGCAATGGATTGGGTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTACAGTTGACATTGATGGGGATGGTAGTTCCCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCCAGTGAAGGTGAT
 20 GATATTGAACAAACCAGCATCTGGGAATGGTGGTCAGTGGGAGGGATAGGTTA
 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTG
 ACGAAGAAGAGCGAAGTCAGCAATCAAGAACATGCTGAGACCCCCAGGG
 CCATACTTGTGGATATCATTGTCCCCCATCAGGAGCACGTGCTGCCTATGATCCC
 25 AAGCGGTGGTCTTTAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTATCAACTACTGGGGTTCAACTGTGAACCATGCGTTCTAGTT
 GCTTGTTCATTATAGCTGTGTTACTTAGTCCGAACCGTAGTTGTA
 GTCTCTGTTCTTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTG
 30 TCCTACATATC

[0549] SEQ ID NO:264
 ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGATTGGTTACAGCTGCAGCTGGCGCTGCT
 35 GTGGCCAACCCAGGTACAGTTGACATTGATGGTGTGGTAGTTCCCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAAACCAGCATCTGGGAATGGTGGTCAGTGGGAGGGATAGGTTA
 CAAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCGAGTG
 40 ACGAAGAAGAGCGAAGTCAGCAATCAAGAACATGCTGAGACCCCCAGGG
 CCATACTTGTGGATATCATAGTCCCCCATCAGGAGCACGTGCTCCAGCGGTGG
 TGCTTCAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTACTGAAATTTC
 GACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATGCCCGCGTGTG
 TATCAACTACTAGGGTTCAACTGTGAGCCATGCGTTCTAGTTGCTTGT
 45 TTCATATAAGCTGTATTACTTAGTCCGAACCCCTGTAGTTGTAGTCTATGTTCT
 CTTTGTAGGGATGTGCTGTCATAAGATGTGTCAGTTCTGCTACATATC

[0550] SEQ ID NO:265
 ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 50 TTCGCTGGTTGGGGCAATGGATTGGTTACAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTACAGTTGACATTGATGGTGTGGTAGTTCCCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAAACCAGCATCTGGGAATGGTGGTCAGTGGGAGGGATAGGTTA
 CAAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT

ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCGAGTG
 ACGAAGAAGAGCGAAGTCAGTCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGTGGATATCATAGTCCCGCATCAGGAGCAGTGTGCCTATGATtCC
 5 cAatGGcGGcGCTTCAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTACT
 GAAATGGTCCGAAGGTCAAGCGGTGGTCAAGGACATGATCATGGAGGGT
 GATGGCAGGACCTCGTACTGAAATTGACCTACAAGACCTACAAGTGTGACATG
 CGCAATCAGCATGATGCCGCGTGTGTTCAACTACTAGGGGTTCAACTGTGAG
 CCATGCGTTCTAGTTGCTTCAATTATAAGCTGTATTACTTAGTTCCG
 10 AACCTGTAGTTGTAGTCTATGTTCTTTGTAGGGATGTGCTGTCATAAGAT
 GTCATGCAAGTTCTGCTACATATC

[0551] SEQ ID NO:266
 ACCAGATGTGGCGGCTCAGTATTACACTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTGGGGCAATGGGATTGGGTTACAGCTGCAGCTGGCGCTGCTG
 15 TGGCCAACCCAGGTGTTACAGTTGTCAGATTGATGGAGATGGTAGTTCCCTCAT
 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGGTGTG
 ATATTGAACAACCAGCATCTGGGAATGGTGGTCAATGGGAGGATAGGTTTAC
 AAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 20 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCAGGGC
 CATACTTGTGGATATCATCGTCCCGCATCAGGAGCAGTGTGCCTATGATtCCc
 AatGGcGGcGCTTCAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTACTG
 AAATTGcAgGGTGGTGTGTTCAAGGACATGATCATGGAGGGTATGGCAGGACC
 25 TCGTACTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCAT
 GGTGCCCGCGTGTGTTCAACTACTAGGGGTTCAACTGTGAACCATGCGTTTCT
 AGTTGCTGTGTTCAATTATAAGCTGTGTTACTTAGTCCGAACCCCTGTAGCT
 TTGAGTCTATGCTCTTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTT
 TCTTGTCCATACATATC

30 [0552] SEQ ID NO:267
 ACCAGATGTGGCGGCTCAGTATTACACTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTGGGGCAATGGGATTGGGTTACAGCTGCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTTACAGTTGTCAGATTGATGGAGATGGTAGTTCCCTCAT
 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGGTGTG
 35 ATATTGAACAACCAGCATCTGGGAATGGTGGTCAATGGGAGGATAGGTTTAC
 AAGGCCAATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCAGCAATCAAGAAGATGCTTGAGACCCCAGGGC
 CATACTTGTGGATATCATCGTCCCGCATCAGGAGCAGTGTGCCTATGATCCC
 40 AAGCGGTGGTGTGTTCAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTA
 CTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGC
 CCGCGTGTGTTCAACTACTAGGGGTTCAACTGTGAACCATGCGTTTCTAGTT
 GCTTGTGTTCAATTATAAGCTGTGTTACTTAGTCCGAACCCCTGTAGCTTGTG
 GTCTATGCTCTTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTG
 45 TCCTACATATC

[0553] SEQ ID NO:268
 ATCAGATGTGGCGGCTCAGTATTACACTACAAGCGGCCACGGCAGTGGCTGTC
 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 50 GTGGCCAACCCAGGTGTTACAGTTGTCAGATTGATGGGGATGGTAGTTCCCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGGTGTG
 GATATTGAACAACCAGCATCTGGGAATGGTGGTCAAGTGGGAGGATAGGTTTA
 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTG

5 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGTGGATATCATTGTCGGCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTCTTTAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTATCAACTACTGGGGTTCAACTGTGAACCATGCGTTTAGTT
 GCTTGTTCATTATAGCTGTACTTAGTCCGAACCGTGTAGTTTGTA
 GTCTCTGTTCTTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTG
 TCCTACATATC

10 [0554] SEQ ID NO:269
 ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCATCCGGTTGGGTGCAATGGGATTGGGTTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTGTGACATTGATGGGATGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCCAGTGAAGGTGAT
 15 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 20 CCATACTTGTGGATATCATTGTCGGCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTCTTTAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTATCAACTACTGGGGTTCAACTGTGAACCATGCGTTTAGTT
 GCTTGTTCATTATAGCTGTACTTAGTCCGAACCGTGTAGTTTGTA
 GTCTCTGTTCTTTGTAGGGATGTGCTGTATAAGATATCATGCAAGTTCTG
 25 TCCTACATATC

[0555] SEQ ID NO:270
 ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGGATTGGGTTACCAGCTGCAGCTGGCGCTGCT
 30 GTGGCCAACCCAGGTGTTACAGTGTGACATTGATGGTGTAGGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 CAAGGCCATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCGAGTG
 35 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 CCATACTTGTGGATATCATAGTCCGCATCAGGAGCACGTGCTGCCTATGATCC
 eAatGGeGGcGCTTCAAGGACATGATCATGGAGGGTATGGCAGGACCTCGTACT
 GAAATggtccgaaggtCAAGCGGTGGTCTTCAGGACATGATCATGGAGGGTGTG
 40 GCAGGACCTCGTACTGAAATTGACCTACAAGACCTACAAGTGTGACATGCGCA
 ATCAGCATGATGCCCGCGTTGTATCAACTACTAGGGGTTCAACTGTGAGCCAT
 GCGTTTCTAGTTGCTGTTCAATTATAGCTGTATTACTTAGTCCGAACC
 CTGTAGTTTGTAAGTCTATGTTCTTTGTAGGGATGTGCTGTATAAGATGTCA
 TGCAAGTTCTGTCCATATC

45 [0556] SEQ ID NO:271
 ACCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTGGTTGGGGCAATGGGATTGGGTTACCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTGTGACATTGATGGTGTAGGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 CAAGGCCATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCGAGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAAGATGCTTGAGACCCCAGGG
 50 CCATACTTGTGGATATCATAGTCCGCATCAGGAGCACGTGCTGCCTATGATCC

CAAGCGGTGGTCTTCAGGACATGATCATGGAGGGTGTGGCAGGACCTCGT
 ACTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
 CCCGCGTGTGTATCAACTACTAGGGTTCAACTGTGAGCCATGCGTTCTAGTT
 5 TGCTTGTTCATTCAATAAGCTTGTATTACTTAGTCCGAACCCCTGTAGTTTGTA
 GTCTATGTTCTCTTTGTAGGGATGTGCTGTCATAAGATGTCATGCAAGTTCTTG
 TCCTACATATC

[0557] SEQ ID NO:272

ACCAAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 10 TTCGCTCTGGTTGGGGCAATGGGATTGGGTACAGCTGTCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTACAGTTGTCAGATTGATGGAGATGGTAGTTCTCAT
 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGGATGATG
 ATATTGAACAACCAGCATCTGGGAATGGTGGTCAATGGGAGGATAGGTTTAC
 15 AAGGCCAATCGGGCGCACACATACCTTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACAGATGCTTGAGACCCCAGGGC
 CATACTTGGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCc
 AatGGcGGcGCTTCAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTACTG
 20 AAATTgcAggTACAAGAGATCCCAAGCGGTGGTGTCAAGGACATGATCATGGAG
 GGTGATGGCAGGACCTCGTACTGAAATTGACCTACAAGACCTACAAGTGTGAC
 ATGCGCAATCAGCATGGTCCCCCGTGTGTATCAACTACTAGGGTTCAACTGT
 GAACCATGCGTTCTAGTTGCTTCAATTCAATAAGCTTGTGTACTTAGTT
 CCGAACCCCTGTAGCTTGTAGTCTATGCTCTTTGTAGGGATGTGCTGTCATAA
 25 GATATCATGCAAGTTCTTGTCTACATATC

[0558] SEQ ID NO:273

ACCAAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTCTGGTTGGGGCAATGGGATTGGGTACAGCTGTCAGCTGGCGCTGCTG
 TGGCCAACCCAGGTGTACAGTTGTCAGATTGATGGAGATGGTAGTTCTCAT
 30 GAACATTCAAGGAGTTGGCATTGATCCGTATTGAGAACCTCCCTGTGAAGGGATG
 ATATTGAACAACCAGCATCTGGGAATGGTGGTCAATGGGAGGATAGGTTTAC
 AAGGCCAATCGGGCGCACACATACCTTGGCAACCCAGAAAATGAGAGTGAGATA
 TATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTGA
 CGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACAGATGCTTGAGACCCCAGGGC
 35 CATACTTGGATATCATCGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTGTCTTCAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGGTGC
 CCGCGTGTGTATCAACTACTAGGGTTCAACTGTGAACCATGCGTTCTAGTT
 40 GCTTGTTCATTCAATAAGCTTGTGTACTTAGTCCGAACCCCTGTAGCTTGT
 GTCTATGCTCTCTTGTAGGGATGTGCTGTCATAAGATATCATGCAAGTTCTTG
 TCCTACATATC

[0559] SEQ ID NO:274

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 45 TTCATCCGGTTGGGTGCAATGGGATTGGGTGCCAGCTGCAAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTACAGTTGTCAGATTGATGGGATGGTAGTTCTCAT
 TGAACATTCAAGGAGTTGGCGTTGATCCGTATTGAGAACCTCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTCAGTGGAGGATAGGTTTA
 CAAGGCCAACCGGGCGCACACATACCTTGGCAACCCAGAAAATGAGAGTGAGAT
 50 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACAGATGCTTGAGACCCCAGGG
 CCATACTTGGATATCATGTCCCGCATCAGGAGCACGTGCTGCCTATGATCCC
 AAGCGGTGGTGTCTTAAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC

CTGCGTGTGTATCAACTACTGGGGTCAACTGTGAACCATGCCTTCTAGTT
 GCTTGTTCATTATATAAGCTTGTACTTAGTCCGAACCGTGTAGTTGT
 GTCTCTGTCTCTTTGTAGGGATGTGCTGATAAGATATCATGCAAGTTCTG
 TCCTACATATC

5

[0560] SEQ ID NO:275

ATCAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCATCCGGTTGGGTCAATGGATTGGGTGCCAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGGATGGTAGTTCTCA
 10 TGAACATTCAAGGAGTTGGCGTTGATCCGTTATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 CAAGGCCAACCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCGGCAGTCGTGTC
 15 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACAGATGCTTGAGAACCCAGGG
 CCATACTTGTGGATATCATTGTCCGCATCAGGAGCAGTGCTGCCTATGATCCC
 AAGCGGTGGTCTTTAAGGACATGATCATGGAGGGTGTGGCAGGACCTCGTA
 CTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATAC
 CTGCGTGTGTATCAACTACTGGGGTCAACTGTGAACCATGCCTTCTAGTT
 20 GCTTGTTCATTATATAAGCTTGTACTTAGTCCGAACCGTGTAGTTGT
 GTCTCTGTCTCTTTGTAGGGATGTGCTGATAAGATATCATGCAAGTTCTG
 TCCTACATATC

[0561] SEQ ID NO:276

ACCAAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 25 TTCGCTCTGGTTGGGGCAATGGATTGGGTACCAAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTGATGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCAATTGAGAACCTCCCAGTGAAGGTGAT
 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 CAAGGCCATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 30 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCGAGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACAGATGCTTGAGAACCCAGGG
 CCATACTTGTGGATATCATAGTCCGCATCAGGAGCAGTGCTGCCTATGATCC
 CAAGCGGTGGTCTTCAGGACATGATCATGGAGGGTGTGGCAGGACCTCGT
 ACTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
 35 CCCCGTGTGTATCAACTACTAGGGGTCAACTGTGAGGCCATGCGTTCTAGTT
 TGCTTGTTCATTATATAAGCTTGTATTACTTAGTCCGAACCCGTAGTTGT
 GTCTATGTTCTCTTTGTAGGGATGTGCTGATAAGATGTCATGCAAGTTCTG
 TCCTACATATC

40 **[0562]** SEQ ID NO:277

ACCAAGATGTGGCGGCTCAGTATTACACTTACAAGCGGCCACGGCAGTGGCTGTC
 TTCGCTCTGGTTGGGGCAATGGATTGGGTACCAAGCTGCAGCTGGCGCTGCT
 GTGGCCAACCCAGGTGTTACAGTTGACATTGATGGTGATGGTAGTTCTCA
 TGAACATTCAAGGAGTTGGCGTTGATCCGCAATTGAGAACCTCCCAGTGAAGGTGAT
 45 GATATTGAACAACCAGCATCTGGGAATGGTGGTGCAGTGGGAGGGATAGGTTTA
 CAAGGCCATCGGGCGCACACATACCTGGCAACCCAGAAAATGAGAGTGAGAT
 ATATCCAGATTGTGACGATTGCTAAAGGATTCAACGTTCCAGCAGTCGAGTG
 ACGAAGAAGAGCGAAGTCACTGCAGCAATCAAGAACAGATGCTTGAGAACCCAGGG
 CCATACTTGTGGATATCATAGTCCGCATCAGGAGCAGTGCTGCCTATGATCC
 50 CAAGCGGTGGTCTTCAGGACATGATCATGGAGGGTGTGGCAGGACCTCGT
 ACTGAAATTTCGACCTACAAGACCTACAAGTGTGACATGCGCAATCAGCATGATG
 CCCCGTGTGTATCAACTACTAGGGGTCAACTGTGAGGCCATGCGTTCTAGTT
 TGCTTGTTCATTATATAAGCTTGTATTACTTAGTCCGAACCCGTAGTTGT

GTCTATGTTCTCTTTGTAGGGATGTGCTGTCATAAGATGTCATGCAAGTTCTTG
TCCTACATATC

[0563] SEQ ID NO:299

5 TCCAAGGTTGCGGCCGCCCCAAGGAACCCTTTCTGGGCCAT:::::TACTCGGCCA
CGACTGGTAATTAAATTCAATTATT

[0564] SEQ ID NO: 326

10 tggcaggatattgtgggtaaacaaattgacgcttagacaacttaataacacattgcgga
cgttttaatgtactgaattaacgcgaattgaattcgagctcggtaccactggatttttgtt
tttaggaattagaattttattgatagaagtatttacaataacaatacataactaagggtt
tcttatatgctcaacacatgagcggaaaccctataagaaccctaattcccttatctggaaact
actcacacatttctggagaaaaatagagagagatagattgttagagagagactgggtatt
tttgcggactctttagatctggtaactggcttaactggccttggaggagctggcaactca
15 aaatcccttgccaaaaaccaacatcatgcgcattccaccatgcttgcgttatccagctgcgcgca
tgtaccccccgtgttatcccaaagcctcatgcaacctaaccatggatcggttggaaaggc
ctataacagcaaccacagacttaaaaccccttgcgcctccatagacttaagcaaatgtgttac
aatgtggatccttaggccaaccccttgcctatgtgacacgtaaacagactctcaactgt
ccaatcgtaagcggttccatgccttcaggcccagcgtaagcaataccagccacaacaccct
20 caacccatcgcaaccaaccaagggtatctatcttgcacccctcgagatcatcaatccactct
tgtgggtttgtggctctgtctaaagttactgttagacgtctcaatgtatggtaacgat
atcacaaaccgcggccatcatcgctgttagctggcttaatctcaactgtctccctccg
gagacatggcttctacactacaaaaagctccgacgaggctgcattgtcacaatcatgaa
aagaaaaactaccgatgacaaatgtgaggattcaatttacccacaaaaaagaagaaga
25 aagatctagcacatctaaggctgacgaaggcagcagaaatataaaaaatataaccatagt
ccctttcccttccatgtcttgcgtttagcatggcggaaattttaaaccctccatcatctc
ccccacaacacggcggatcgccatcatccgagagcccttcccgagatccggcc
ggatccacgcccggcgagagccctagccgcagatccgcggccctcccgccaccgatctggc
gcccacgaaagccgccttcgcccacccaaactaccacggccaaagatcgagaccgagacgga
30 aaaaaaaaaacggagaaagaaagaggagagggggcggggtggtaaccggcgccggcg
gg
ttgtgggttaggttggggggatggggggggggggggggatgttatatagtggcgatgg
ggcggttttttggaaagcgagggggggggccgcctcgctcgctcgatccctcg
tttccggccccacgaccggaccacactgtgttttttttttttttttttttttt
35 ttttttttttggctgcgagacgtgcgggtgcgtgcggacaactcacgtgatagtgggggg
gtggagactattgtccagttggctggactgggtgggtgggtgggtgggtgggtgg
cttgctatggatcgatgactttggctttaggaacttttaggggtttttttttaat
gtttttagtctaagttatcttttattttactagaaaaataccatgcgcgtcaacgggg
gaaagctatttaatcttattattgttcatgtgagaattgcctgaatataatattttctc
40 aaaaattatgtcaaatttagcatatgggtttttaaagatatttttataacaatccctcg
tatttacaaaagcaacgaacttaaaacccgactcaaatacagatatgcattccaaaagcg
aataaaactaaaaaccaattcataaaaaatgacgtatcaaagtaccgacaaaaacatcctc
aattttataatagttagaaaaagagtaatttcactttggccaccttttattaccgatattt
tactttataccacccctttaactgtatgttttactttgaccaggtaatcttacccctt
45 attttggactatccgcactcttcaagcatatgaatgacctcgagatgtctgatgt
gtcgacctgcagggtgcagcgtgaccggcgtgcggccctcttagagataatgagcattgc
tgtctaagttaaaaattaccacatattttttgtcacactgtttgaagtgcagtt
tctatcttatacatatatttaactttactctacgaataatataatctatagtactacaat
aatatcagtgttttagagaatcatataaaatgaacagtttagacatggctaaaggacaattga
50 gtattttgacaaacaggactctacagtttatcttttagtgcgtgttttcctt
ttgcaaatagcttcacctataatacttcatccatttttagtacatccatctttaggg
agggttaatgggttttagactaatttttagtacatcttatttattttagtgcctc
taaatttaagaaaaactaaaactctatttttagtttttatttaataatttagatataaaatag
aataaaataaaatgtactaaaattaaacaaatacccttaagaaattaaaaactaaggaa
55 acattttttttttcgagtagataatgccagccgtttaacgcgcgtcgacgactaaccgg

ttcaaacatttggcaataaaagttcttaagattgaatccgttgcgcgtctgcgtatgatta
 tcataataatttctgttgaattacgttaagcatgtataataatcatgtatgcgtacgtta
 tttatgagatgggtttatgatttagagtccgcattatacatttataacgcgtatagaaaa
 5 caaaatatacgccgcaaaactaggataaaattatcgccgcgcgtgtcatctatgttactagatc
 gaaagcttagcttgagcttggatcagattgtcgttccgccttcagttaaactatcagtg
 tttgacaggatattggcgggtaaac

10 [0565] SEQ ID NO: 327
 gcgaagatcc aggacaagga

15 [0566] SEQ ID NO: 328
 ctgcattaccg gcaaaagatga g

20 [0567] SEQ ID NO: 329
 ttccccccgga ccagcagcgt

25 [0568] SEQ ID NO: 330
 ccgacgagaa agaccagcaa

30 [0569] SEQ ID NO: 331
 cttaaagtgt cgatcgggac tgt

35 [0570] SEQ ID NO: 332
 tgaggcctctc gtgcgcgatc acat

40 [0571] SEQ ID NO: 333
 ccactcttgccttacacgacactgaagacacttattgattccaaacggcggccgccttcaaggac
 atgatcatggagggtatggcaggacctcgtaactgaaatttcgacactacaagacactacaagt
 gtgacatgcgcaatcagcatggtgcgcgtgttatcaactacttaggggttcaactgtga
 accatgcgtttcttagttgttgcattcatataagctgttacttagttccgaacc
 ctgtagctttagtctatgtctctttgttagggatgtgtcataagatcatgcgaag
 tttcttgcctacatataataactacttccatggaaataattctcagttctgtttgaa
 ttttgcacatcttcacaaacactgtgtcgttgccttctgttgcgtacgcgcctcctcgact
 ccatccccatggtcgcacatcagggccagggtccccgcgcgtatcggttagcgacttcgtg
 45 ggcgaggaaagccttcgttccaagggtgtccctcctcgcaatctgttggatggtaatatt
 ataaaaacgcctgccttcgcggtaagactccgcacatccaggatgaggatgaccagcct
 tttgcagttatccacttagggacaggattgcacccgcgtctggaaactgtcctttgaggaccactc
 gcctccaatttgcacggtgcgcgcagcgcacgcgtctggaaactgtcctttgaggaccactc
 cgtttgcacatgggtacctggagatcatgcacattaaaggatgaccagttcgtaaagggtcctgc
 ggtgtctattgcatttcataggttaataagtgtttgtactgtggtaaaggccaagact
 cccgcacatctctatgcggggacaagtgccacccacagtggggcaggatgaggatgac
 caaagactcccgcacatctcacttagggacaggattggcctttgcgttatctctatgc
 gggacaagtgtatccgaagtaataaaaaccatcgactctcgatataagactgtcgactcgac
 cggccgcacgcacatgggttcatggtaagctgtctattttaatgaaactcgacggtagc
 agtgtggatgagggttcatggtaactcggtactccactcgtac

50 [0572] SEQ ID NO: 334
 ccactcttgccttacacgacactgaagacgtgcgcattaccggcaagtgacccgcgcac
 atcgacggacgcgttccaggagacgcacatgtggaggatcgcgtccatcacaagca
 caactacctggccttgcgtggaggatccccgcgtcatccaggaagccttcttcatt
 catccctctggcccccggggccgtctgttatcccaaggacatccagcagcagatg
 gctgtgcgcgtctggacactccaatgatggcgcaggatcatcgccgcctgcggcaagcc
 accatctactgaatcgcttgcgtggcaggatccgtctggcgttgcgttgcggaggat
 tggatgttggatggcgtcgctgcgttgcgttgcgttgcgttgcgttgcgttgcgtt
 55 gggattccagttacaactactctgtatggccttggcaacttccccagcgcacgcaccactgtc

tctgcgcatgctggatgcacggcactgttatgcacaaattatgcagtagataaggctgacc
 tggtgcgcattgggtgcgggttatgcacattgacattgacccagctgagattggcaagaacaaggc
 acatgtctccattgtgcagatgttaagcttgcttacagggttgaatgatctattaaatg
 5 ggagcaaaggcacaacagggtctggatttggccatggcacaaggagttgatcagcagaag
 agggagtttcctctaggattcaagactttggcgaggccatccgcgcataatgcatacc
 ggtactggatgagctgacaaaaggaggcgatcattgcacttgtgttggcagcaccaga
 tggggcgctcagtattacactacaagcgccacggcagtggtgtctcgctggtttgc
 10 gggcaatgggatttgggttaccagctgcagctggcgctgtgtggccaaccagggttac
 agttgttgcattgatggtagttcctcatgaacattcaggagtggcgatcc
 gcattgagaaccccccagtgaaggatgatattgaacaaccagcatctggaaatggtggtgc
 cagtggaggataggtttacaaggccaatcgccgcacacatacctggcaacccagaaaa
 tgagagttagatataccagattttgtacgattgctaaaggattcaacgttccagcagttc
 gagtgcacgaagaagagcgaagtactgcagcaatcaagaagatgtttaggacccaggcca
 15 tacttggatcatagtcggccatcaggagcacgtgctgcctatgatccaaagcggtgg
 tgcttcaggacatgatcatggagggtatggcaggacctcgactgaaatttcgacactac
 aagacctacaagtgtacatgcgcaatcagcatggccgcgtgttgcataactactagg
 ggtcaactgtgacccatgcgtttctagttgtcttgcattcatataagcttgc
 20 ttagttccgaaccctgttagcttgcattgtactctctttgttagggatgtgc
 gatattcatgcaagttcttcctacatataataacttccatgaaataattctca
 gttctgtttgaatttgcatttcacaaacagtgtgctggccttctgttctacgccc
 cgccgtccggccgcgtccgcgtctgcgtccgcacctccggccggccaccaccc
 25 ctccgttagcgacttcgtggcgaggaaagcccttcgtccaaagggtggccctcgc
 tggatggatggtaatattataaaaggcctgccttcgcgggtgagtc
 tgactgtgactggacaggattggccttgcagttatccactaggacaggattgc
 30 atgtccttttgcggaccactccgtttgtctagaggtaactggagatcatgac
 attaaggatggccctttgcagttatctctagaaagactggagttgc
 tccgccttgcactccaggatgaggatgaccatatccgaagtaaaaaccatcg
 gactctcgactcgaccggccacgcataaggcttgcattgaagctgtctatt
 taaaatgaaactcggacggtagcagtgttatgaggcttcagcacactcg
 taactccagtcac

35 **[0573]** SEQ ID NO: 335
 tgagattggcaagaacaaggcagccacatgtctccattgtgcagatgttaagcttgc
 ttttacagggttgaatgatctattaaatggagcaaaaggcatacagggtctggat
 ttggccatggcacaaggatggcgttgcacccagggttgc
 40 ccactgggttggcagcaccagatgtggcggctc
 cagtggtgtctcgatgggttggccatggatgggttaccagctgc
 cagctggcgc
 tgctgtggccaaaccagggttgc
 acattcaggatggcgttgc
 aaccaggcatctgg
 45 cacatcacccaggaaaatgagatgttgc
 aaggattcaacgttccaggc
 aagatgcttgc
 gctgc
 50 ccgcgtgttatcaactactagggttcaactgt
 ttcattcatataaggcttgc
 ttttgc
 agtacttccatgg
 gctgg
 55 ccaggccccggccatgc
 gatcg
 gacttc
 cgtgg
 cgg
 ccagg

tggccctcctcgcaatcttggatggtaatattataaaaggcctgccttcgcgggt
 aagactcccccattccaggatgaggatgaccgccttgcagttatccactaggacag
 gattgcattcctgccaaaccctgccaagcttgcggtagcctccaatttgcggtagc
 5 catgacattaaaggatgaccagttcgtaaaggcctgcgggtctattgccttcata
 ataagtgtttgcgtactgtggtaaaggccaagactccgcctatctctatgcgg
 caagtgcacccacagtggggcaggatgaggatgaccaaagactccgcctactag
 ggacaggattggcctttgcagttatctatgcggacaagtgtatccgaagtaata
 10 aaaccatcgactctcgataagactgtcgactcgaccggcgcacgcataaggat
 gctgctattctattaaattgaaatccaaagcggtggtgcattcaaggacatgat
 ggtgatggcaggacctcgactgaaatttcgaccaagacactacaaggatgcgca
 atcagcatgatgcgcgtgttatcaactactaggggtcaactgtgagccatgcgtt
 15 ctatgcggacaactgttagttgtatcatataagctgttattacttagtcc
 agtctatgttcttttagggatgtgtcataagatgtcatgcaagttttgccta
 catatcaataataaagtactccatggaataattctcagttctgtttgaatttt
 tcacaaacagtgtgctgggccttctgttactttacatgtctgtgtcagggtctgaca
 taacgaccgatggagggtggcggcagggtttagaaggaaattttttggaa
 20 gaagtctgaataacagttggagggaaatagaagatataacttcgattaatttat
 caagccgcttgcattcaagcactagacagttaggggtttgtgttgcatttc
 gagatccggcttgcggagggagaccgaggctcggtgtgtgttgcatttc
 tttgagatagagagagatcctgcactgtgggtttgtgttgcacagcgagaga
 gacattgagagatgttagatcggttacc

[0574] SEQ ID NO:350

25 CCAGAAGGTAATTATCCAAGATGTAGCATCAAGAATCCAATGTTACGGGAAAA
 ACTATGGAAGTATTATGTAAGCTCAGCAAGAACAGATCAATATGCCGCACATA
 TGCAACCTATGTTCAAAAATGAAGAATGTACAGATACAAGATCCTATACTGCCAG
 AATACGAAGAAGAACACGTAGAAATTGAAAAAGAACCCAGGGCGAAGAAAAG
 AATCTGAAGACGTAAGCACTGACGACAACAATGAAAAAGAACAGATAAGGTCG
 30 GTGATTGTGAAAGAGACATAGAGGACACATGTAAGGTGGAAAATGTAAGGGCGG
 AAAGTAACCTTATCACAAAGGAATCTTATCCCCCACTACTTATCCTTTATTTT
 TCCGTGTCATTTCGCCCTGAGTTTCTATATAAGGAACCAAGTCGGCATTG
 TGAAAACAAGAAAAATTGGTGTAGCTATTCTTGAAGTACTGAGGATACA
 ACTTCAGAGAAATTGTAAGTTGTAGATCTCATGGCTCCAAGGAAGAGGAAGG
 35 AGTCTAACAGGGAGTCAGCTAGGAGGTCAAGGTACAGGAAGGTGGGTATCCACG
 GGGTACCCGCCGCTATGGCTGAGAGGCCCTCCAGTGTGCAATCTGCATCGTAA
 CTTCAGTCGTAGTGACAAACCTGAGCAACCACATCCGACCCACACAGGCAGAA
 GCCTTTGCCCTGTGACATTGAGGGAGGAAATTGCCACCAGCAGCAGCCGCATA
 AACCATACCAAGATACACACGGGCAGCCAAAGCCCTCCAGTGTGCAATCTGC
 40 ATGCGTAACTCAGTCGTAGTGACAAACCTGAGCGAACACATCCGACCCACACAG
 GCGAGAACGCTTGCCTGTGACATTGAGGGAGGAAATTGCCAGCAAGAC
 CCGAAAAACCATACCAAGATACACACGGGCAGAACGCCCTCCAGTGTGCAAT
 CTGCATGCGTAAGTTGCCGCTCCGACGCCCTGACCCAGCATGCCAGAGATGC
 GGACTGCGGGGATCCCAACTGTGAAATCAGAATTGAGAGAAAAAGTCTGAG
 45 CTTAGACACAAATTGAAGTACGTTCCACATGAATATATCGAACTTATCGAGATTG
 CTAGGAACACTAACACAGGACAGAATTGGAGATGAAGGTTATGGAGTTCTTAT
 GAAAGTGTACGGATATAGGGGAAAGCACCTGGTGGTCTAGGAAACCTGATGG
 TGCAATCTACACTGTGGGATCACCTATTGACTATGGTGTATCGTGGATACAAAG
 GCATACTCTGGTGGATACAATTGCCAATCGGACAAGCTGACGAAATGCAGAGA
 50 TATGTTGAAGAGAACCAAACATAGAAACAAACATATTAAATCCAATGAATGGTGG
 AAGGTGTATCCTTCATCTGTTACAGAGTTCAAATTCCCTTTGTGCTGGACACTT
 TAAGGGTAACACAAAGCACAGCTTACTAGGTTGAACCATATTACAAATTGCAAT
 GGTGCTGTGTTGTCAGTTGAAGAGCTTGTGATCGGAGGTGAAATGATTAAGGCAG
 GAACACTTACTTGGAGGAAGTTAGAAGAAAATTCAACAAACGGTGAATCAATT
 55 TTAGATCTGGCGCGGAGAGGGCAGAGGAAGTCTTCAACATCGGGTGACGTGG

AGGAGAATCCGGCCCTAGGATGGCTCCAAGGAAGAGGAAGGAGTCTAACAGGG
 AGTCAGCTAGGAGGTCAAGGTACAGGAAGGTGGGTATCCACGGGGTACCGCCG
 CTATGGCTGAGAGGCCCTCCAGTGTCAATCTGCATCGTAACCTCAGTCGTAG
 TGACACCCCTGAGCACGCACATCCGCACCCACACAGGCGAGAAGCCTTTGCCTGT
 5 GACATTGTGGGAGGAAATTGCCGACAGGAGCAGCCGCATAAACATACCAAG
 ATACACACGGGATCTCAGAAGGCCCTCCAGTGTCAATCTGCATCGTAACCTCA
 GTCGCTCCGACGACCTGTCCAAGCACATCCGCACCCACACAGGCGAGAAGCCTTT
 TGCTGTGACATTGTGGGAGGAAGTTGCCGACAACCTCCAACCGCATCAAGCAT
 GCCCAGAGATGCGGACTGCCGGATCCAACTTGTGAAATCAGAATTGGAAGAG
 10 AAAAAGTCTGAGCTAGACACAAATTGAAGTACGTTCCACATGAATAATCGAAC
 TTATCGAGATTGCTAGGAACCTAACACACAGGACAGAAATTGGAGATGAAGGTTAT
 GGAGTTCTTATGAAAGTGTACGGATATAGGGAAAGCACCTTGGTGGTTCTAGG
 AACCTGATGGTGCAACTACACTGTGGGATCACCTATTGACTATGGTGTATCG
 TGGATACAAAGGCATACTCTGGGATACAATTGCCAACGGACAAGCTGACG
 15 AAATGCAGAGATATGTTGAAGAGAACCAAACAGAAACAAACATATTAATCCAA
 ATGAATGGTGBAAGGTGTATCCTTCATCTGTTACAGAGTCAAATTCTTTGTG
 TCTGGACACTTTAAGGTAACACTACAAAGCACAGCTACTAGGTTGAACCATATTA
 CAAATTGCAATGGTGTGTTGTCAGTTGAAGAGCTTGTGATCGGAGGTGAAAT
 GATTAAGGCAGGAACACTTACTTGGAGGAAGTTAGAAGAAAATTCAACAAACGG
 20 TGAAATCAATTGATAACTCGAGCTCGGTACCAGCATAATTATTAAATGTA
 CTAATTACTGTTTGTAAATGCAATTGCTTCTCGGGATTAAATATCAAAA
 TCTATTAGAAATACACAATATTGTTGCAAGGCTGCTGGAGAAATCGATCTGCTA
 TCATAAAAATTACAAAAAAATTATTGCTCAATTATTAGGATTGGTATTAA
 GGACGCTTAAATTATTGTCGGGCACTACGCATCATTGTGATTGAGAAGATCAG
 25 CGATACGAAATATTGCTAGTACTATCGATAATTATTGAAAATTCTACAGAAAA
 GCAAACGTTACATGAATTGATGAAACAATACAAAGACAGATAAAGCCACGCACA
 TTTAGGATATTGGCCGAGATTACTGAATATTGAGTAAGATCACGGAATTCTGAC
 AGGAGCATGTCTCAATTGCCAAATGGCAGTTGAAATACTCAAACCGCCCCA
 TATGCAGGAGCGGATCATTGATTGTTGTTGGTTGCCACATGGGAGTC
 30 CAAGGTT

[0575] SEQ ID NO:351

CCAGAAGGTAATTATCCAAGATGTAGCATCAAGAATCCAATGTTACGGGAAAA
 ACTATGGAAGTATTATGTAAGCTCAGCAAGAACAGATCAATATGCCGACATA
 35 TGCAACCTATGTTCAAAATGAAGAACATGTACAGATACAAGATCCTATACTGCCAG
 AATACGAAGAAGAACATCGTAGAAATTGAAAAAGAACCCAGGGCGAAGAAAAG
 AATCTGAAGACGTAAGCACTGACGACAACAATGAAAAGAACAGATAAGGTGCG
 GTGATTGTGAAAGAGACATAGAGGACACATGTAAGGTGGAAAATGTAAGGGCGG
 AAAGTAACCTTATCACAAGGAATCTTATCCCCACTACTTATCCTTTATATT
 40 TCCGTGTCATTGCTGAGTTCTATATAAGGAACCAAGTTCGGCATTG
 TGAAAACAAGAAAAATTGGTGTAAAGCTATTCTTGAAGTACTGAGGATACA
 ACTTCAGAGAAATTGTAAGTTGTAGATCTCCATGGCTCCAAGGAAGAGGAAGG
 AGTCTAACAGGGAGTCAGCTAGGAGGTCAAGGTACAGGAAGGTGGGTATCCACG
 GGGTACCCGCCGCTATGGCTGAGAGGCCCTCCAGTGTCAATCTGCATCGTAA
 45 CTTCACTGTCAGTCCTCCGACCTGTCCCCACATCCGCACCCACACCGGGCGAGAAG
 CCTTTGCCTGTGACATTGTGGGAGGAAATTGCCAGGCCGGCAACCTGTCCA
 AGCATACCAAGAACACACGCATCCCAGGGCACCTATTCCAAGGCCCTCCAGTG
 TCGAATCTGCATCGTAAGTTGCCAGTCCGGCACCTGACCCGCCATACCAAG
 ATACACACGGCGAGAACGCCCTCCAGTGTCAATCTGCATCGTAACCTCAGTA
 50 CCTCCGGCTCCCTGTCCCCACATCCGCACCCACACCGGGCGAGAACGCCCTTGC
 CTGTGACATTGTGGGAGGAAATTGCCAGTCCGGCAACCTGGCCGCCATGCC
 CAGAGATGCGGACTGCCAACCTGTGAAATCAGAATTGGAAGAGAAA
 AAGTCTGAGCTAGACACAAATTGAAGTACGTTCCACATGAATAATCGAACCTA
 TCGAGATTGCTAGGAACCTAACACAGGACAGAACATTGGAGATGAAGGTTATGG

AGTTCTTATGAAAGTGTACGGATATAGGGAAAGCACCTGGTGGTTAGGAA
 ACCTGATGGTGCAATCTACACTGTGGGATCACCTATTGACTATGGTGTATCGTG
 GATACAAAGGCATACTCTGGTGGATACAATTGCCAATCGGACAAGCTGACGAA
 ATGCAGAGATATGTTGAAGAGAACCAAACAGAAACTAGAAACAAACATATTAATCCAAAT
 5 GAATGGTGGAGGTGTATCCTTCATCTGTTACAGAGTTCAAATTCTTTGTGTC
 TGGACACTTAAGGGTAACACTACAAAGCACAGCTTACTAGGTTGAACCATTACAA
 AATTGCAATGGTGTGTTGTCAGTTGAAGAGCTTTGATCGGAGGTGAAATGAA
 TTAAGGCAGGAACACTTACTTGGAGGAAGTTAGAAGAAAATTCAACACGGTG
 10 AAATCAATTITAGATCTGGCGCGAGAGGGCAGAGGAAGTCTCTAACATGCG
 GTGACGTGGAGGAGAATCCCAGCCCTAGGATGGCTCCAAGGAAGAGGAAGGAGT
 CTAACAGGGAGTCAGCTAGGAGGTCAAGGTACAGGAAGGTGGGTATCCACGGGG
 TACCCCGCGCTATGGCTGAGAGGCCCTCCAGTGTGCAATCTGCATGCGTAACCT
 15 CAGTACCTCCGGCTCCCTGTCCGCCACATCCGCACCCACACCGCGAGAAGCCT
 TTTGCCTGTGACATTGTGGAGGAATTGCCCTGCGCCAGACCCCTGCGCGACC
 ATACCAAGATACACACGGGCAGCCAAAGCCCTCCAGTGTGCAATCTGCATGC
 GTAACTTCAGTACCTCCGGCAACCTGACCCGCCACATCCGCACCCACACCGCGA
 20 GAAGCCTTTGCCTGTGACATTGTGGAGGAATTGCCCGACCGCTCCGCCCTG
 GCCCGCCATACCAAGATACACACGGGATCTCAGAAGCCCTCCAGTGTGCAATCT
 GCATGCGTAACCTCAGTCGCTCCGACGTGCTGTCCGAGCACATCCGCACCCACAC
 CGGCAGAAGCCTTTGCCTGTGACATTGTGGAGGAATTGCCCGCAACTTC
 TCCCTGACCATGCATGCCAGAGAGATGCGGACTGCGGGGATCCCAACTGTGAAAT
 25 CAGAATTGGAAGAGAAAAAGTCTGAGCTTAGACACACAAATTGAAGTACGTTCCAC
 ATGAATATATCGAACTTATCGAGATTGCTAGGAACCTCAACACAGGACAGAAATT
 GGAGATGAAGGTTATGGAGTTCTTATGAAAGTGTACGGATATAGGGAAAGCA
 CCTTGGTGGTTCTAGGAAACCTGATGGTCAACTACACTGTGGGATCACCTATT
 30 GACTATGGTGTATCGTGGATACAAGGCATACTCTGGTGGATACAATTGCCAA
 TCGGACAAGCTGACGAAATGCGAGAGATATGTTGAAGAGAACCAACTAGAAACA
 AACATATTAATCCAATGAATGGTGGAGGTGTATCCTCATCTGTTACAGAGTT
 CAAATTCTTTGTGCTGGACACTTAAGGGTAACTACAAAGCACAGCTTACT
 AGGTTGAACCATATTACAATTGCAATGGTGTGTTGTCAGTTGAAGAGCTT
 35 TGATCGGAGGTGAAATGATTAAGGCAGGAACACTTACTTGGAGGAAGTTAGAA
 GAAAATTCAACAACGGTGAATCAATTGATAACTCGAGCTCGGTACCCAGCA
 TAATTGTTATTAAATGTAATTACTGTTGTTAAATGCAATTGCTTCTCGG
 GATTGTTAATATCAAATCTATTAGAAATACACAATTGTTGTCAGGCTTGCTG
 GAGAATCGATGCTATCATAAAAATTACAAAAAAATTGTTATTGCCCTCAATTAT
 40 TTTAGGATTGGTATTAAAGGACGCTTAAATTATTGTCGGGTCACTACGCATCATTG
 TGATTGAGAAGATCAGCGATACGAAATATTGCTAGTACTATCGATAATTGTT
 AAAATTCTATAAGAAAAGCAAACGTTACATGAATTGATGAAACAATCAAAGACA
 GATAAAGGCCACGCACATTAGGATATTGGCCGAGATTACTGAATATTGAGTAAGA
 TCACGGAATTCTGACAGGAGCATGTCTCAATTGAGCTTGAAGAGCTT
 TGCACATGGAGTCCAAGGTT

[0576] SEQ ID NO:352

45 GCCCAAGGAACCCCTTCTGGGCCATCTCGTACTCGGCCACGACTGGTAATTAA
 ATGGATCCAACCGACAACCACTTGCAGCTTCCACTACCTGTTCAAGAGAACATTCAATAAGG
 TTAATTCTAATTGAAATCCGAAGATAAGATTCCCACACACTTGTGGCTGATATC
 AAAAGGCTACTGCCTATTAAACACATCTCTGGAGACTGAGAAAATCAGACCTCC
 AAGCATGAAGAACGCTGAGCTACTGCTACTTGTGAGAAGTCCCTCATCGAG
 50 AAGTTGATTCTGTGATCTTATGCAGCTCTGAGGGTGAGGAATCAAGAG
 CTTCTCTTCGATGTTGGTGGAGAGGATACGTTCTCAGAGTTAACTCTTGCCT
 GACGGATTCTACAAGGATAGATACGTTACAGACACTTCGCTTCAGCTGCTCTCC
 CTATCCCTGAAGTTCTGATATCGGAGAGTTCTGAGTCTCTTACCTACTGTATC
 TCAAGAAGGGCTCAGGGTGTACTCTTCAAGATCTCCTGAGACTGAGCTCCTG

CTGTTCTTCAACCTGTTGCTGAGGGTATGGATGCTATCGCTGCTGCTGATCTTCTC
5 CAAACTCTGGATTCGGACCTTCGGTCTCAGGAATCGGACAGTACACTACTT
GGAGAGATTCTATCTGCGCTATCGCTGATCCTCATGTTACCATTGGCAGACCGTT
ATGGATGATAACCGTTCTGCTTCTGCTCAAGCTCTGATGAGCTTATGCTTIG
GGCTGAGGATTGTCCTGAGGTTAGACATCTGTTACGCTGATTCGGATCTAAC
AACGTTCTCACCGATAACCGGAAGAATCACCCTGTTATCGATTGGTCTGAGGCTA
10 TGTTGGAGATTCTCAATACGAGGTGGCCAACATATCTTITGGAGGCCCTGGCTT
GCTTGTATGGAACAACAGACTAGATACTTCGAGAGAAGGCATCCTGAGCTTGTG
GATCTCCTAGACTTAGAGCTTACATGCTTAGGATCGGACTTGATCAGCTTACCA
GTCTCTCGTTGATGGAAACTTCGATGATGCTGCTTGGGCTCAGGGAAAGATGTGAT
GCTATCGTTAGATCTGGTCTGGAACTGTTGGAAGAAGACTCAAATCGCTAGAACAG
15 CTGCTGCTGTTGGACTGATGGATGTGTTGAAGTTCGCTGATTCTGAAACAG
AAGGCCTTCTACTAGACCTAGAGCCAAGAACAGTGAAGATCGCGGCAATAGCTC
TTAGCGCCATCCCCGGTTGATCCTATCTGTTGAAATAGTTCGGTGGGCAAGG
CTCTCTTCAGAAAGACAGGCGGCAAAGGAACCCAAGGTGAGGTGGCTATGG
20 CTCTCAGTTCTGTGGAAGCGCTTGGCTAAGGTGCAGAGGTGTTAGCGGGATG
AAGCAAAAGTGTCCGATTGTAACAAGATACTGTTGATCCTACGTAAGGATATTAAA
GTATGTATTCACTAAATATAATCAGTGTATTCCAATATGTTACTACGATTCCAA
TGTCTTATTGTCGCCGTATGTAATCGCGTCACAAAATAATCCCCGGTGACTTTTC
25 TTTAATCCAGGATGAAATAATATGTTATTATAATTTCGCGATTGGTCCGTTAT
AGGAATTGAAGTGTGCTGCGGTGCCACCACTCCATTCTATAATTACATGTA
TTGAAAAAATAAAAATTATGGTATTCAATTAAACACGTATACTTGATAAGAAT
GATATCTGAAAGAAATATAGTTAAATATTGATAAAATAACAAGTCAGGT
ATTATAGTCCAAGCAAAACATAAAATTATTGATGCAAGTTAAATTCAAGAAATA
TTTCAATAACTGATTATATCAGCTGGTACATTGCCGTAGATGAAAGACTGAGTGC
GATATTATGGTGTAAATACATAGCGGCCGCCCCAAGGAACCCCTTCTGGGCCAT
CTTCGTACTCGGCCACGACTGGTAATTAAAT

[0577] SEQ ID NO:353

5 aattatttaggattggattaaaggacgcttaaattttgtcggtcaactacgcatcattgtgattgagaagatcagcgatacgaatattcg
tagtactatcgataatttttggaaaattcataagaaaagcaacgttacatgaatgtgaaacaatacacaagacagataaagccacgc
catttaggatattggccgagattactgaatattgagtaagatcacggaattctgacaggagcatgtctcaattcagccaaatggcagtt
gaaactactaaaccgcacccatgtcaggagcggatcattgttgggttgccaaatcatgggaggtccaagggtgcggccgc
gcaccaaggaaacccttctggccatctcgactcgccacactgtaatttaat

[0578] SEQ ID NO:354

10 GCCCAAGGAACCCTTCTGGGCCATCTCGTACTCGGCCACGACTGGTAATTAA
ATGGATCCAACCGACAACCACCTTGCGGACTTCCTTCAAGAGAATTCAATAAGG
TTAATTCTAATTGAAATCCGAAGATAAGATTCCCACACACTTGTGGCTGATATC
AAAAGGCTACTGCCTATTAAACACATCTCTGGAGACTGAGAAAATCAGACCTCC
AAGCATGAAGAAGCCTGAGCTACTGCTACTTCTGTTGAGAAGTCCCTCATCGAG
AAGTCGATTCTGTGTCTGATCTTATGCAGCTCTGAGGGTGAGGAATCAAGAG
CTTCTCTTCGATGTTGGAGAGAGGATACGTCTCAGAGTTAATCTTGCCT
15 GACGGATTCTACAAGGATAGATACGTGACAGACACTTCGCTTCAGCTGCTCTCC
CTATCCCTGAAGTTCTGATATCGGAGAGTTCTGAGTCTCTTACCTACTGTATC
TCAAGAAGGGCTCAGGGTGTACTCTCAAGATCTCCTGAGACTGAGCTCCTG
CTGTTCTCAACCTGTTGCTGAGGCTATGGATGCTATCGCTGCTGATCTTCT
CAAACCTCTGGATTGGACCTTCGGCTCAGGGAAATCGGACAGTACACTACTT
20 GGAGAGATTCATCTGCGCTATCGCTGATCCTCATGTTACCAATTGGCAGACCGTT
ATGGATGATACCGTTCTGCTCTGTTGCTCAAGCTCTGATGAGCTTATGCTTTG
GGCTGAGGATTGTCCTGAGGTTAGACATCTGTTACGCTGATTGGATCTAAC
AACGTTCTCACCGATAACGGAAGAATCACCGCTGTTATCGATTGGTCTGAGGCTA
TGTTCGGAGATTCTCAATACGAGGTGGCCAACATATTCTTGGAGGCCTGGCTT
25 GCTTGTATGGAACACAGACTAGATACCTCGAGAGAAGGCATCCTGAGCTGCTG
GATCTCCTAGACTAGAGCTTACATGCTTAGGATCGGACTTGATCAGCTTACCA
GTCTCTCGTTGATGGAACACTCGATGATGCTGCTGGCCTCAGGGAAAGATGTGAT
GCTATCGTTAGATCTGGTCTGGAACTGTTGAGAAGACTCAAATCGCTAGAAGAT
CTGCTGCTGTTGGACTGATGGATGTTGAAGTTCTGCTGATTCTGAAACAG
30 AAGGCCTTCTACTAGACCTAGAGCCAAGAAGTGAAGATCGGCGCAATAGCTTC
TTAGCGCCATCCCGGGTTGATCCTATCTGTGTTGAAATAGITCGGGTGGCAAGG
CTCTCTTCAGAAAGACAGGCAGGCCAAAGGAACCCAAGGTGAGGTGGCTATGG
CTCTCAGTTCTTGTGGAAGCGCTTGGTCTAAGGTGAGGGTGTAGCGGGATG
AAGCAAAAGTGTCCGATTGTAACAAGATATGTTGATCCTACGTAAGGATATTAAA
35 GTATGTATTCACTACTAAATATCAGTGTATTCCAATATGTAACGATTCCAA
TGTCTTATTGTCGCCGTATGTAATCGCGTCACAAAATAATCCCGGTGACTTTC
TTTAATCCAGGATGAAATAATATGTTATTATAATTGGTCAATTTCATAATTACATGTA
AGGAATTGAAGTGTGCTTGCAGGCCACACTCCCATTCTGAAATTTACATGTA
40 TTTGAAAAATAAAATTATGGTATTCAATTAAACACGTATACTTGTAAAGAAT
GATATCTGAAAGAAATATAGTTAAATTATGATAAAATAACAAGTCAGGT
ATTATAGTCCAAGCAAAAACATAAATTATTGATGCAAGTTAAATTCAAGAAATA
TTTCAATAACTGATTATATCAGCTGGTACATTGCCGTAGATGAAAGACTGAGTGC
GATATTATGGTGTAAACATAGCGGCCAGCGAGAGAAAGCTTATTGCAACTTC
AACTACTGCTGGTCGATCGTGTGGCCACTC
45

[0579] SEQ ID NO:355

10 GCCCAAGGAACCCTTCTGGGCCATCTCGTACTCGGCCACGACTGGTAATTAA
ATGGATCCAACAGCGCCAGTGTGCTGGAATTGCCCTCGTCGACCTG
CAGGTCAACGGATCAGGATATTCTGTTAAGATGTTGAACTCTATGGAGGTTG
50 TATGAACTGATGATCTAGGACGGATAAGTCCCTCTCATAGCGAACTTATT
AAAGAATGTTGTGATCATTCTGTTACATTGTTATTAAATGAAAAAAATTATT
GGTCATTGGACTGAACACGAGTGTAAATATGGACCAGGCCAAATAAGATCC
ATTGATATATGAATTAAATAACAAGAATAATCGAGTCACCAAAACCAACTGCCTT
TTTAACGAGACTTGTACCAACTGATACAAAAGTCATTCTATGCAAATC

AATAATCATACAAAAATATCCAATAACACTAAAAAATTAAAAGAAATGGATAAT
 TTCACAATATGTTACGATAAAGAAGTTACTTTCCAAGAAATTCACTGATT
 TAAGCCCACCTGCATTAGATAAATGGCAAAAAAAACAAAAAGGAAAAGAAATA
 AAGCACGAAGAATTCTAGAAAATACGAAATACGCTCAATGCAGTGGGACCCAC
 5 GGTCAATTATTGCCAATTTCAGCTCCACCGTATTTAAAAAATAAAACGATA
 ATGCTAAAAAAATATAATCGTAACGATCGTTAAATCTCAACGGCTGGATCTTAT
 GACGACCGTTAGAAATTGTGGTGTGACGAGTCAGTAATAAACGGCGTCAAAG
 TGGTTGCAGCCGGCACACACGAGTCGTGTTATCAACTCAAAGCACAAATACTT
 TCCTCAACCTAAAAATAAGGCAATTAGCCAAAACAACTTGCGTGTAAACAAAC
 10 GCTCAATACACGTGTCATTTATTAGCTATTGCTTCACCGCCTAGCTTCTC
 GTGACCTAGTCGTCCTCGTCTTCTCTCTCTATAAAACAATACCCAAAG
 AGCTCTTCTTCTCACAAATTCAAGATTCAATTCTCAAATCTTAAACTTCT
 CAATTCTCTCACCGTGATCAAGGTAATTCTGTGTTCTTATTCTCTCAAATC
 TTCGATTTGTTCGTGTGATCCAAATTCTGATATGTTCTTGGTTAGATTCTG
 15 TTAATCTTAGATCGAAGACGATTTCTGGGTTGATCGTAGATATCATCTTAATT
 CTCGATTAGGGTTCATAGATATCATCCGATTGTTCAAATAATTGAGTTGTC
 GAATAATTACTCTCGATTGATTTCTATCTAGATCTGGTGTAGTTCTAGTT
 GTGCGATCGAATTGTCGATTAATCTGAGTTCTGATTAACAGATGAGAGGAT
 CTGGATCTGAGTCTGATGAGTCTGGACTCCTGCTATGAAATCGAGTGTAGAAT
 20 CACTGGAACCTTAACGGTGTGAGTTGAGCTGTTGGAGGTGGTGGAGGAACT
 CCTGAGCAGGGAAAGAATGACTAACAAAGATGAAGTCTACCAAGGGTGCTTAC
 TTCTCTCCATACCTTCTTCACGTTATGGGATACGGATTCTACCACTTCGGAAC
 TTACCCATCTGGATACGAGAACCCCTTCATGCTATCAACAAACGGTGGATAC
 ACCAACACTAGGATCGAGAAGTACGAGGATGGTGGTGTCTCACGTTAGCTTCT
 25 CTTACAGATACGAGGCTGGAAGAGTGATCGGAGATTCAAGGTTATGGGAACTG
 GATTCCCTGAGGATTCTGTTATCTTCACCGACAAGATCATCAGGTCTAACGCTACT
 GTGAGCATCTCATCCTATGGGAGATAACGATCTCGATGGATCTTACCCAGAA
 CCTTCTCACTTAGAGATGGTGGTTACTACTCTCTGTGGGATTCTCACATGCAC
 TTCAAGTCTGCTATCCACCCCTCTATCCTCAAAACGGTGGACCTATGTTGCTT
 30 CAGAAGAGTTGAGGAAGATCACTCTAACACCGAGCTTCCAATCGTTGAGTACCA
 ACATGCTTCAAGACCCCTGATGCTGATGCTGGTGGAGGAATGATAATATCAAAT
 CTATTAGAAATACACAATATTGTTGCAGGCTGCTGGAGAATCGATCTGCTAT
 CATAAAAATTACAAAAAAATTATTGCTCAATTATTAGGATTGGTATTAAG
 GACGCTTAAATTATTGTCGGGTCACTACGCATCATTGATGAGAAGATCAGC
 35 GATACGAAATATTCGTAGTACTATCGATAATTATTGAAAATTCTATAAGAAAAG
 CAAACGTTACATGAATTGATGAAACAATACAAAGACAGATAAACGCCACGCACAT
 TTAGGATATTGGCCGAGATTACTGAATATTGAGTAAGATCACGGAATTCTGACA
 GGAGCATGTCTCAATTCAAGCCAAATGGCAGTTGAAATACTCAAACCGCCCCAT
 ATGCAGGAGCGGATCATTGTTGTTGGTGCCTTGCACATGGGAGTCC
 40 AAGGTTGCGGCCGCGAGCAGAGAGAAAGCTTATTGCAACTCAACTTGCTGGTC
 GATCGTGTGGCCACTC

[0580] SEQ ID NO:375
 TTCTGGCCTCTTATTGGGCCGCCAAGGAACCTTTCTGGGCCATCTCGTACT
 45 CGGCCACGACTGGTAATTAAATGGATCCACTAGTAA

[0581] SEQ ID NO:376
 TTCTGGCCTCTTATTGGGCCGCCAAGGAACCTTTCTGGGCCATCTCGTACT
 CGGCCACGACTGGTAATTAAATGGATCCACTAGTAA

[0582] SEQ ID NO:377
 TTCTGGCCTCTTATTGGGCCGCCAAGGAACCTTTCTGGGCCATC:CAGTCGT
 GGCGAGTACGAAGATGGCCAGA:::TACTCGGCCACGACTGGTAATTAAATGGA
 TCCACTAGTAA

[0583] SEQ ID NO:378
 TTCTGGCCTCTTATTGGGCCGCCAAGGAACCCTTCCTGGGCCATC:::GTACTCG
 GCCACGACTGGTAATTAAATGGATCCACTAGTAA

5 [0584] SEQ ID NO:379
 [0585] TTCTGGCCTCTTATTGGGCCGCCAAGGAACCCTTCCTAGG:::::T
 ATCTCAGTTGGTGTAGGTCGTCGCTCCAAGCTGGGCTGCGTCACGAAC:CGTA
 CTCGGCCACGACTGGTAATTAAATGGATCCACTAGTAA

10 [0586] SEQ ID NO:380
 TTCTGGCCTCTTATTGGGCCGCCAAGGAACCCTTCCTGGGCCA:::::::::::GAC
 TGGTAATTAAATGGATCCACTAGTAA

15 [0587] SEQ ID NO:381
 TCCAAGGTTGCGGCCGCGGCCAAGGAACCCTTCCTGGGCCATCTCGTACTCGG
 CCACGACTGGTAATTAAATTCAATTATT

20 [0588] SEQ ID NO:382
 TCCAAGGTTGCGGCCGCGGCCAAGGAACCCTTCCTGGGCCAT:T:::TACTCGGCCA
 CGACTGGTAATTAAATTCAATTATT

[0589] SEQ ID NO:383
 :CGTACTCGGCCACGACTGGTAATTAAATTTCATTTATT

25 [0590] SEQ ID NO:384
 TCCAAGGTTGCGGCCGCGGCCAAGGAACCCTTCCTGGGCCATCTCGTACT
 CGGCCACGACTGGTAATTAAATTCAATTATT

30 [0591] SEQ ID NO:385
 TCCAAGGTTGCGGCCGCGGCCAAGGAACCCTTCCTGGGCCATCTC:::::::::::TG
 GTAATTAAATTCAATTATT

35 [0592] SEQ ID NO:386
 TCCAAGGTTGCGGCCGCGGCCAAGGAACCCTTCCTGG:::::TAGCGGTGGTT
 TTTTGTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGA
 TCGTACTCGGCCACGACTGGTAATTAAATTCAATTATT

40 [0593] SEQ ID NO:387
 TCCAAGGTTGCGGCCGCGGCCAAGGAACCCTTCCTGGGCCATCTACGAGC
 GTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACATATCCC
 AGCCACGACT::::::::::GGTAATTAAATTTCATTTATT

45 [0594] SEQ ID NO:388
 TAGTTTATTGCCCCAAGCGAGAGAAAGCTTATTGCAACTCAACTCGTACTCGG
 CCACGACTGGTAATTAAATGGATCCACTAGTAA

50 [0595] SEQ ID NO:389
 TAGTTTATTGCCCCAAGCGAGAGAAAGCTTATTGCAACTCAACT

[0596] SEQ ID NO:390
 TAGTTTATTGCCCCAAGCGAGAGAAAGCTTATTGCAACTCAACG

[0597] SEQ ID NO:391
 TAGTTTATTGCCCAAGCGAGAGAAAGCTTATTGCAACTTCAACTCGTACTCG
 GCCACGACTGGTAATTAAATGGATCCACTAGTAA

5 [0598] SEQ ID NO:392
 TAGTTTATTGCCCAAGCGAGAGAAAGCTTATTGCAACTTCAACTAT:::GTACTCG
 GCCACGACTGGTAATTAAATGGATCCACTAGTAA

10 [0599] SEQ ID NO:393
 TAGTTTATTGCCCAAGCGAGAGAAAGCTTATTGCAACTTCA::::TACTCGGCCA
 CGACTGGTAATTAAATGGATCCACTAGTAA

15 [0600] SEQ ID NO:394
 AGGTAATTAAATGGATCCACTAGTAA

[0601] SEQ ID NO:395
 TCCAAGGTTGCGGCCGCAGCGAGAGAAAGCTTATTGCAACTTCAACTACTGCTG
 GTCGATCGTGTG
 GCCACTCTGTTATCTATCA

20 [0602] SEQ ID NO:396
 TCCAAGGTTGCGGCCGCAGCGAGAGAAAGCTTATTGCAACTTCA:::ACTTGCTGGT
 CGATCGTGTG
 GCCACTCTGTTATCTATCA

25 [0603] SEQ ID NO:397
 TCCAAGGTTGCGGCCGC:::::::::::GCGCCGACCCAGCTTCTTGTACAAA
 GTTGGCATTATAAGAAAGCATTGCTTATCAATTGTTGCAACGAACAGGTCACTA
 TCAGTC
 AAA:ACTTGCTGGTCGATCGTGTG
 GCCACTCTGTTATCTATCA

30 [0604] SEQ ID NO:398
 TCCAAGGTTGCGGCCGCAGCGAGAGAAAGCTTATTGCAA:::::CTTC:ACTTGCTG
 GTCGATCGTGTG
 GCCACTCTGTTATCTATCA

35 [0605] SEQ ID NO:399
 TCCAAGGTTGCGGCCGCAGCGAGAGAAAGCTTATTGCAACTTCA:::GATAAAAGTT
 GCTCGCCTGTGTGGGTGTGGATGCT:ACTTGCTGGTCGATCGTGTG
 GCCACTCTT
 GTTTATCTATCA

40 [0606] SEQ ID NO:400
 TCCAAGGTTGCGGCCGCAGCGAGAGAAAGCTTATTGCAACTTCAACTACACT
 TGCTGGTCGATCGTGTG
 GCCACTCTGTTATCTATCA

45 [0607] SEQ ID NO:401
 TCCAAGGTTGCGGCCGCAGCGAGAGAAAGCTTATTGCAACTTCAACTACTGCTG
 GTCGATCGTGTG
 GCCACTCTGTTATCTATCA

[0608] SEQ ID NO:402
 CTTACATGCTTAGGATCGGACTTG

50 [0609] SEQ ID NO:403
 AGTTCCAGCACCAGATCTAACG

[0610] SEQ ID NO:404
 CCCTGAGGCCAAGCAGCATCATCG

[0611] SEQ ID NO:405
CGGAGAGGGCGTGGAAAGG

5 [0612] SEQ ID NO:406
TTCGATTGCTACAGCGTCAAC

[0613] SEQ ID NO:407
AGGCACCATCGCAGGCTTCGCT

10 [0614] SEQ ID NO:408
TTCTGGCCTCTTATTGGGCCGCCAAGGAACCCTTCTGGGCATCTCGTACT
CGGCCACGACTGGTAATTAAATGGATCCAACCGACAACCACTT

15 [0615] SEQ ID NO:409
TTCTGGCCTCTTATTGGGCCGCCAAGGAACCCTTTACTCGGCCACG
ACTGGTAATTAAATGGATCCAACCGACAACCACTT

[0616] SEQ ID NO:410
TTCTGGCCTCTTATTGGGCCGCCAAGGAACCCTTCTGG:::TCGTACTCGGC
20 CACGACTGGTAATTAAATGGATCCAACCGACAACCACTT

[0617] SEQ ID NO:411
TTCTGGCCTCTTATTGGGCCGCCAAGGAACCCTTCTGG

25 [0618] SEQ ID NO:412
TTCTGGCCTCTTATTGGGCCGCCAAGGAACCCTTCTGGGCATCT:::CGGC
CACGACTGGTAATTAAATGGATCCAACCGACAACCACTT

30 [0619] SEQ ID NO:413
TTCTGGCCTCTTATTGGGCCGCCAAGGAACCCTTCTGGGCAT:::TCGTACTC
GGCCACGACTGGTAATTAAATGGATCCAACCGACAACCACTT

[0620] SEQ ID NO:414
TTCTGGCCTCTTATTGGGCCGCCAAGGAACCCTTCTGGGCATCT

35 [0621] SEQ ID NO:415
GTAATACATAGCGGCCGCGCCCAAGGAACCCTTCTGGGCATCTCGTACTCG
GCCACGACTGGTAATTAAATTCAATTATTCTCAACTTCTTA

40 [0622] SEQ ID NO:416
GTAATACATAGCGGCCGCGCCCAAGGAACCCTTCTGGGCAT:::GCCACG
ACTGGTAATTAAATTCAATTATTCTCAACTTCTTA

[0623] SEQ ID NO:417
45 GTAATACATAGCGGCCGCGCCCAAGGAACCCTTCTGGGCATCT:::GAC
TGGTAATTAAATTCAATTATTCTCAACTTCTTA

[0624] SEQ ID NO:418
50 GTAATACATAGCGGCCGCGCCCAAGGAACCCTTCTGGGCATCT:::GAC
TGGTAATTAAATTCAATTATTCTCAACTTCTTA

[0625] SEQ ID NO:419
GTAATACATAGCGGCCGCGCCCAAGGAACCCTTCTGGGCAT:::GTACTCGGC
CACGACTGGTAATTAAATTCAATTATTCTCAACTTCTTA

5 [0626] SEQ ID NO:420

::GTACTCGGCCACGACTGGTAATTAAATT:::TCTTTCAACTTCTTA

10 [0627] SEQ ID NO:421

GTAATACATAGCGGCCGCGCCCAA:::::::::::::TACTCGGCCACGACTGGTAA
TTAATTTCATAATTATTTCTCAACTTCTTA

15 [0628] SEQ ID NO:422

TGTAATACATAGCGGCCGCGCCCAAAGGAACCCTTACTCGGCCA:::::::::::::TA
ATTAAATTTCATAATTATTTCTCAACTTCTTA

20 [0629] SEQ ID NO:423

tnantgattc ccaatggcgg cgcttcaag gacatgatca tggagggtga tggcaggacctcgactgaa atggtccgaa
15 ggtccacgccc gccaactacg ag

25 [0630] SEQ ID NO:424

cnantactcg tagttggcgg cgtggacattt cggaccattt cagtagcagg tctgccatcacccctccatg atcatgtcct
tggaaagcggcc gccattggga at

30 [0631] SEQ ID NO:425

tnantgattc ccaatggcgg cgcttcaag gacatgatca tggagggtga tggcaggacc tcgtactgaa atttgcaggt
acaag

35 [0632] SEQ ID NO:426

angngtcttg tacctgcaaa ttccatgtacg aggtccctgcc atcacccctcc atgatcatgtcctgaaagc gccgcattt ggaat

40 [0633] SEQ ID NO:427

GTTTACCCGCCAATATATCCTGTCAAACACTGATAGTTAAACTGAAGGCAGGGAA

45 ACGACAATCTGATCATGAGCGGAGAATTAAAGGGAGTCACGTTATGACCCCCCGCC

GATGACGCGGGACAAGCCGTTTACGTTGGAAGTGCACAGAACCGCAACGTTGA

AGGAGCCACTCAGCAAGCTTACTAGTAGCGCTGTTAAACGCTCTCAACTGGAA

GAGCGGTTACCCGGACCGAAGCTTGCATGCCTGCAGTGCAGCGTGAACCGGGTCTG

GCCCCCTCTAGAGATAATGAGCATTGCATGTCTAAGTTATAAAAAATTACCA

55 TATTTTTTTGTCACACTGTTGAAGTGCAGTTATCTATCTTATACATATATT

AAACTTTACTCTACGAATAATAATCTATAGTACTACAATAATCAGTGTGTTA

GAGAATCATATAAAATGAACAGTTAGACATGGCTAAAGGACAATTGAGTATTTG

ACAACAGGACTCTACAGTTATCTTGTGCTGATGTGTTCTCTTTTTTG

60 CAAATAGCTTCACCTATATAACTTCATCCATTATTAGTACATCCATTAGGG

TTTAGGGTTAATGGTTTATAGACTAATTTTTAGTACATCTATTATTCTATT

TTAGCCTCTAAATTAAGAAAACCTAAACTCTATTAGTTTTTATTAAATAATT

TAGATATAAAATAGAATAAAATAAGTGAACCTAAACAAATACCCCTTA

AGAAAATTAAAAAAACTAAGGAAACATTCTTGTGAGTAGATAATGCCAGC

CTGTTAAACGCCGTCGACGAGTCTAACGGACACCAACCAGCGAACCGAGCGT

65 CGCGTCGGGCCAAGCGAACCGAGACGGCACGGCATCTCTGCGCTGCCCTGGACC

CCTCTCGAGAGTCCGCTCCACCGTTGGACTTGCTCCGCTGCGCATCCAGAAA

TTGCGTGGCGGAGCGGGCAGACGTGAGCCGGCACGGCAGGGCGGCTCCTCCTC

TCACGGCACCGGCAGCTACGGGGATTCTTCCCACCGCTCCTCGCTTCCCT

70 CCTCGCCCGCCGTAATAAAATAGACACACACACACACACACACACACACAC

GTTGTTCGGAGCGCACACACACACACACACACACACACACACACACACACAC

ACCTCCGCTTCAAGGTACGCCGCTCGTCCCTCCCCCCCCCCCCCTCTACCTTCTC

TAGATCGGCAGTCCGGCCATGGTTAGGGCCCGGTAGTTCTACTCTGTTATGTT

TGTGTTAGATCCGTGTTGTGTTAGATCCGTGCTGCTAGCGTACACGGATGC

GACCTGTACGTAC

ATCCTGGGATGGCTAGCCGTCGCAGACGGGATCGATTCATGATTTTTTG
 TTCGTTGCATAGGGTTGGCTTGCCTTCTTCAATATGCCGTGCAC
 TTGTTGTCGGGTCACTTTCATGCTTTTGTCTGGTTGTGATGATGGTC
 TGTTGGCGGTCGTTAGATCGGAGTAGAATTCTGTTCAAACACTGGTGG
 5 ATTATTAATTGGATCTGTATGTGTGCCATACATATTCAAGTTACGAATTG
 AAGATGATGGATGAAATATCGATCTAGGATAGGTACATGTTGATGCCGGTT
 TACTGATGCATATACAGAGATGCTTTGTCGCTGGTTGTGATGATGGTGTG
 GTTGGCGGTCGTTCTAGATCGGAGTAGAATACTGTTCAAACACTACCT
 GGTGTATTATTAAATTGGAACTGTATGTGTGTCATACATCTTCATAGTTACG
 10 AGTTAAAGATGGATGAAATATCGATGTAGGATAGGTACATGTTGATGGGT
 TTTACTGATGCATATACATGATGGCATATGCAGCATCTATTCAATGCTCTAACCT
 TGAGTACCTATCTATTATAATAAAACAAGTATGTTTATAATTATTGATCTGAT
 AACTGGATGATGGCATATGCAGCTATGTTGATTTAGCCCTGCCT
 15 CATACTGCTATTATTGCTTGGTACTGTTCTTGTGATGCTCACCTGTTGTT
 GGTGTTACTTCTGCAGGTCGACTCTAGAGGATCCACACGACACCAGTCCGCCCG
 CGAGGTGCACATCGACGTGAACAACAAGACGCCACACCCCTCCAGCTGGAGGA
 CAAGACCAAGCTCGACGGCGCAGGTGGCGCACCTCCCCGACCAACGTGGCAA
 CGACCAAGATCAAGACCTCGTGGCGAATCCAACGGCTCATGACCGGACCGA
 GGGCACCCTACTACTCAATTAAATGGCGAGGCGAGATCAGCCTACTTCGAC
 20 AACCGTTCGCCGGCTCCAACAAATACGACGGCCACTCCAACAAAGTCCCAGTACG
 AGATCATCACCCAGGGCGCTCCGGCAACCAGTCCCACGTGACCTACACCCTCCA
 GACCACCTCCTCCCGCTACGGCCACAAGTCCCTGAGTCATGAGTCAGTCAGTT
 AACCTAGACTTGTCCATCTTCTGGATTGGCCAACCTAATTAAATGTATGAAATAAA
 AGGATGCACACATAGTACATGCTAATCACTATAATGTGGGATCAAAGTTGTG
 25 GTTATGTGTAATTACTAGTTATCTGAATAAAAGAGAAAGAGATCATCCATATTTC
 TTATCTAAATGAATGTCACGTGCTTATAATTCTTGATGAACAGATGCATT
 CATTAAACCAATCCATATACATATAAAATTAAATCATATATAATTAAATATCAATT
 GGGTTAGCAAAACAAATCTAGTCTAGGTTGTTTGCATGCCGGACCG
 AATTGGGGATCTGCATGAAAGAAACTGTCGCACTGCTGAACCGCACCTGTC
 30 TTCATCGAACACGACCTGTGCCAAGATGACGGTGTGCGGTCTAAGTGAGGCTG
 AATTGCCCTGGACAGAACGGACTCCCTACAATTAGTTAGGCCAACGGTGCATC
 CATGTTAGCTCCGGCTCGGGCTGTATGCCATCTGCAATAGCATCCATGGAGC
 TCGTCCATGTAGTTGGAGATGAACCAATGATCGGGCTGTGGACGTATGTTCT
 GTGTACTCCGATAGTAGAGTACGTGTTAGCTTTCATGGTCAAGTGAAATTG
 35 TGTTGGTTAATTACCCCTACGTTAGTTGCCGGACAGGAGACACATCATGAATT
 AAAGGCATGATGTCCTCTCTGTAATGTTATTCTTGATGTGATGAATCAAAT
 GTCATATAAAACATTGTTGCTTTAGTTAGGCCTGATCGTAGAACGAAATGCT
 CGTGTAGCGGGCTACGAGCCTATGACGCAATAACACTGGTTGCCGGCCCGA
 GTCGCTGACAAAAAAAGCATGTTAAGTTATTACAATTCAAACCTAACATA
 40 TTATATTCCCTCAAAGCAGGTTACGATCACACCTGTACCTAAAAAAACATGAA
 GAATATATTACTCCATTATTATGAGATGAACCACTGGCAAGAGTGGTAAGCTAT
 ATAAAAAAATGAACATTATTACGAGATGTTATATGCCATTATATTGATTGCAAGA
 TATATGTTCTTCTCCACGGCACCTAACGGATACATGATAAGGCCAACGGCAG
 ATCACGGGAAATTATCGAACATGTTACGCCATTGCCGGAAAAAAATGCA
 45 GGGCAGGTGTTGCCGTAGCAGTTAACGACTTAAGCTGGAGGTTGCCACACTTG
 GATGCAAGCGTCTGACCCCTCTAAAAAAATCGGCGGCTTGTCCGTATCCGTATCC
 CCTATCCAACATCTAGCTGGCACACGACGGGGCTGGCAGATCGTGGATGCCG
 GGTCGACGTCGATCGTCAAGCCATCATGACCAATGACCCATCTGTTATGGATGCT
 TGCTAGCTAGACTAGTCAGACATAAAATTGGATACTTCTCCCAACTGGGAGAC
 50 GGGGACTGATGTGCAGCTGACGTGAGCTAAATTTCCTATAAATATGCATGA
 AATACTGCATTATCTGCCACAGCCACTGCCACAGCCAGATAACAAAGTGCAGCTG
 GTAGCACGCAACGCATAGCTGGACTGTTAGCTAGGTAGCCAACCGGATCCACA
 CGACACCATGCTCGACACCAACAAGGTGTACGAGATCAGCAACCACGCCAACGG
 CCTCTACGCCGCCACCTACCTCTCCCTGACGACTCCGGCGTGTCCCTCATGAACA

AGAACGACGACGACATCGACGACTACAACCTCAAGTGGTCTCTTCCCACATCGA
 CGACGACCAGTACATCATCACCTCCTACGCCGCCAACAACTGCAAGGTGTGGAAC
 GTGAACAAACGACAAGATTAATGTGTCAACCTACTCCTCCACCAACTCCATCCAGA
 AGTGGCAGATCAAGGCCAACGGCTCTTACGTGATCCAGTCCGACAACGGCA
 5 AGGTGCTACCGCCGGCACCGGCCAGGCCCTCGGCCTCATCCGCTCACCGACGA
 GTCCTCCAACAACCCGAACCAGCAATGAAACCTGACGTCCTGCAAGACCATCCA
 GCTCCCGAGAAGCCGATCATCGACACCAAGCTCAAGGACTACCCGAAGTACTC
 CCCGACCGGCAACATCGACAACGGCACCTCCCCGAGCTCATGGCTGGACCCCTC
 GTGCCGTGCATCATGGTGAACGACCCGAACATCGACAAGAACACCCAGATCAAG
 10 ACCACCCGTACTACATCCTCAAGAAGTACCAAGTACTGGCAGAGGGCCGTGGGCT
 CCAACGTCGCCTCCGCCGCACGAGAAGAAGTCCTACACCTACGGAGTGGGCA
 CCGAGATCGACCAGAACGACCAACATCAACACCCCTGGCTCCAGATCAACAT
 CGACAGCGGCATGAAGTTCGACATCCCGAGGTGGCGCGGTACCGACGAGAT
 CAAGACCCAGCTCAACGAGGAGCTCAAGATCGAGTATTACATGAGACGAAGAT
 15 CATGGAGAAGTACCAAGGAGCAGTCCGAGATCGACAACCCGACCGACCAAGTCCAT
 GAACTCCATCGGCTTCCCTACCATCACCTCCCTGGAGCTTACCGCTACAACGGC
 TCCGAGATCCGCATCATGCAGATCCAGACCTCCGACAACGACACCTACAACGTGA
 CCTCCTACCCGAACCACCAAGCAGGCCCTGCTGCTGACCAACCACCTACGA
 GGAGGTGGAGGAGATACCAACATCCGAAGTCCACCCCTCAAGAACAGCTCAAGAA
 20 GTACTACTTCTGAGTCATGAGTCATGAGTCAGTTAACCTAGACTTGTCCATCTTCT
 GGATTGGCAACTTAATTAAATGTATGAAATAAAGGATGCACACATAGTACAT
 GCTAATCACTATAATGTGGGCATCAAAGTTGTGTTATGTGAATTACTAGTTAT
 CTGAATAAAAGAGAAAGAGATCATCCATATTCTTATCCTAAATGAATGTACGT
 GTCTTATAATTCTTGATGAACCAAGATGCATTCACTAACCAACATCCATATACAT
 25 ATAATATTAAATCATATATAATTAAATATCAATTGGGTTAGCAAACAAATCTAGT
 CTAGGTGTGTTTGCATCCATGGAGTCAAAGATTCAAATAGAGGACCTAAC
 AGAACTCGCCGTAAAGACTGGCGAACAGTTCATACAGAGTCTCTACGACTCAAT
 GACAAGAAGAAAATCTCGTCAACATGGTGGAGCAGCACGCTTGTCTACTCCA
 AAAATATCAAAGATACTCAGTCAGAACGACCAAGGGCAATTGAGACTTTCAAC
 30 AAAGGGTAATATCCGAAACCTCCTCGGATTCCATTGCCAGCTATCTGTCACTT
 TATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCTACAAATGCCATATTGCGAT
 AAAGGAAAGGCCATCGTTGAAGATGCCCTGCGACAGTGGTCCAAAGATGGA
 CCCCCACCCACGAGGAGCATCGTGGAAAAAGAACAGCTTCAACCACGTCTTCA
 AAGCAAGTGGATTGATGTGATATCTCCACTGACGTAAGGGATGACGCACAATCCC
 35 ACTATCCTCGCAAGACCCCTCCTCTATATAAGGAAGTTCTATTGAGGAGG
 ACAGGGTACCCGGGATCCACCATGTCTCCGGAGAGGGAGACCAGTTGAGATTAG
 GCCAGCTACAGCAGCTGATATGGCCGGTTGTGATATCGTAAACCATTACATT
 GAGACGTCTACAGTGAACTTAGGACAGAGCCACAAACACCACAAAGAGTGGATT
 GATGATCTAGAGAGGTTGCAAGATAGATACCCTGGTTGGTGTGAGGTTGAGG
 40 GTGTTGTGGCTGGTATTGCTTACGCTGGGCCCTGGAAGGCTAGGAACGCTTACGA
 TTGGACAGTTGAGAGTACTGTTACGTGTCACAAGGCATCAAAGGTTGGCCTA
 GGATCCACATTGTACACACATTGCTTAAGTCTATGGAGGCGCAAGGTTTAAGT
 CTGTGGTTGCTGTTATAGGCCTTCAAACGATCCATCTGTTAGGTTGATGAGGCT
 TTGGGATACACAGCCCCGGGTACATTGCGCGCAGCTGGATACAAGCATGGTGG
 45 TGGCATGATGTTGGTTTGGCAAAGGGATTGAGTTGCCAGCTCCTCCAAGGC
 CAGTTAGGCCAGTTACCCAGATCTGAGTCGACCTGCAGGCATGCCGCTGAAATC
 ACCAGTCTCTCTACAAATCTATCTCTCTATAATAATGTGTGAGTAGTTCCA
 GATAAGGGAAATTAGGGTTCTTATAGGGTTGCTCATGTGTTGAGCATATAAGAA
 ACCCTAGTATGTATTGTATTGTAAAATCTCTATCAATAAAATTCTAATT
 50 CTAAAACAAATCCAGGGCGAGCTCGTACCCGGGATCCTCTAGAGTCGACC
 TGCAGGCATGCCCGCGGATATCGATGGGCCCGGCCGAAGCTTGGTCCGGGCC
 ATCGTGGCCTCTGCTCTCAGGATGAAGAGCTATGTTAACGTGCAAGCGCTC
 AATTGCCCTATAGTGAAGTCGTATTACAATCGTACGCAATTCAAGTACATTAAAAA

CGTCCGCAATGTGTTAAAGTTGCTAACCGTCAATTGTTACACCACAATATA
TCCTGCCA

[0634] SEQ ID NO:428

5 GAGGCCGACACGGCACACACGGCAGACATTCACCGCCGGCTCCTCCGTGCCACT
CGGCACAAGGCTCATCAGTCGCCGATGCCGATGCGATCAACGGAAGCGGATGG
CCCCTCTTAAAGATTGGCACAGGAACACTGGCCACTGCCCTGATGTGCAATT
ATGCCTGCGAAAGCCTAGGCAACACACCGCAATAAACGAGCGAATGACACGGAA
AGCTGATGTGGTATGAATTATAACATTGGGCCAAATATTATTCTATCCAC
10 CATTGTGTAGCCACAGCATCGGTATTGAGTTGCGAGGACAAATCCCTCGTGA
GGTCAAAAACAGCAAATAATAACCCATCTCCTGAAGACACCAAAAAAGGAG
CAGCTCCTCGTCAATGAACAAGCGTCACAAGAAAAGGGAGCACGTAATAAC
CTCTCAATTGCTTCAGCATGAAAAGAACGGGAAGAAATGCAAGTCTACAGAGG
AAAGTGCAGCTGTTCCGGCTGCCATGCCAAGTTCCATGGCGAGGAAAGCT
15 GAACTGGATTCCAGTCTCGCGCTGTCACTGCTCAGCTGCTTGTAGGATGCGGCAA
TAGTTCACCTGGATGAAAAGATAACAGTTAGTCTGAAGCAGTCGAGTGGACAT
CCAAAGTATCAAATCGAAAGCTGTAAATGGGAAGGAAATACCTCTACCC
GGAAAAGTTGGTAGGCAAAATAATCCCAACGCCAGCAGAGCTC

20 [0635] SEQ ID NO:429

CGTCAAGCGCTCAATTGCCCTATAGTGAGTCGTATTACAATCGTACGCAATT
AGTACATTAAAAACGTCCGCAATGTGTTAAAGTTGCTAACCGTCAATTGTT
ACACCAAGAGGCCGACACGGCACACACGGCAGACATTCACCGCCGGCTCCTCCGTC
GCCACTCGGACAAGGCTCATCAGTCGCCGATGCCGATGCGATCAACGGAAAGC
25 GGATGGCCCCTTAAAGATTGGCACAGGAACACTGCCACTGCCCTGATGTC
GCAATTATGCCCTCGAAAGCCTAGGCAACACACCGCAATAACGAGCGAATGAC
ACGGAAAGCTGATGTGGTATGAATTATAACATTGGGCCAAATATTATTCT
ATCCACCAATTGTTAGGCCACAGCATCGTATTGAGTTGCGAGGACAAATCCC
TCGTGAGGTCAAAAACAGCAAATAATAACCCATCTCCTGAAGACACCAAAAAA
30 AAGGAGCAGCTCCTCGTCAATGAACAAGCGTCACAAGAAAAGGGAGCACGTA
AATAACCTCTCAATTGCTTCAGCATGAAAAGAACGGGAAGAAATGCAAGTCTA
CAGAGGAAAGTGCAGCTGTTCCGGCTGCCATGGCAAGTTCCATGGCGAGG
AAAAGCTGAACGGATTCCAGTCTCGCGCTGTCACTGCTCAGCTGCTTGTAGGAT
GCCGAATAGTTCACCTGGATGAAAAGATAACAGTTAGTCTGAAGCAGTCGA
35 GTGGACATCCAAAGTATCAAATCGAAAGCTGTAAATGGGAAGGAAATATAC
CTCTACCCGGAAAAGTTGGTAGGCAAAATAATCCCAACGCCAGCAGAGCTC

[0636] SEQ ID NO:430

AGTTGGGAAGGCAAAACGAATATAAGTGCAATTGGATTACTGTTAGTCGAGTC
40 TATTAAAGGAATTCTATTGTAATGTTCTAACCTAACCTAACGCTAT
GGCTGATATGGATCTGATTGGACTTGATTATCCATGATAAGTTAACGAGCAACT
CAAAGAGGTTAGGTATATATGGTTTGAAAGGTAATTAGTTAACATTAGAAA
AAAAAAAGTGTATCCAATAGGCTCTATAAACAACTCTCAAATTAGGGCTTCT
ATCCATCCACCTTGCTCTATTGGATAGCCTGATTACTCTCTATTAGTCC
45 GTAGGTTAATGAGTCTGTTGGATTAGCCTACACTTTCTGTAAAATCTATT
GATAGTAGCTAAATCAGTAATTGGCTAGTATTAGCTATTCTCTGGAGTT
GCTATAAGACCAAGAACATGTAATTGGAAGTTGCGACGGACGAGAACGCA
TGACAAATCCAGAGTATTGATGATGGAATTACCTATTACCCACTCTCCATT
GTGTCCATTCTCATCATCCCCGGCGCTTCTGCATCCGGTACAGCTGACATGAC
50 ACGTTCACCGCGTTACATGGCTGATGGCTCACAAAGTCACCCCCACATGTCTAGTGT
TCGCCCCAGGCAAGATCGTCCTCGGCCCTGCGCTGCCGTGCTCTGCCGCCGCTGCTT
GGGCCCTGCTGGCGCCCGCTGCCGATCACACGGCCTACGCCGTGCAAGGCAGCGC
CACCGAACCCGCACTGTTGCGATAGGTGGCAGTGGCAGTGGCACTGGCA
CGGCACCGCAGTCGATCGCTCCGCTCATCTGCTGACAGTGGATAGAGCAGCGTTGG

CCGTTGGGGCCGGATCTCGTGAAGCGTCGTCCCTGCTGTACTGTGCCGTATG
 GCGTGTGCGCTTCGCCATGTTTCTTTCTTTCTTTCTTTGCTAGGG
 CGGTTCTCGCTGGTAACAGGGACCACTCGGTTGATCCGTTGAATTACTG
 AAAGAGATGGGAATGGTCGCTGTGCCCGGGACATTGAATGAGATGTTGTAAG
 5 TGAATATGGCTTACGCTTTGCGAGTGGGAATGGATGCTAAACGAACACAAACC
 GGGTTAAACCAGAGGCCGACACGGCACACACGGCGACATTCACCGCCGGCTC
 CTCCGTCGCCACTCGCACAAAGGCTCATCAGTCGCCGATGCCGATGCGATCAAC
 GGAAGCGGAATGGCCCGCTTCTTAGAATTGGCACAGGAACACTGCCACTGCCCT
 TGATGTGCAATTATGCCTGCGAAAGCCTAGGCAACACACCGAATAAACGAGCG
 10 AATGACACGGAAAGCTGATGTGGTATGAATTATACAACATTATGGGCCAAAATA
 TTATTCTATCCACCATTGTGTAGCCACAGCATCGGTATTGAGTTGTGCGAGGAC
 AAATCCCTCGTGAGGTCAAAAACAGCAAATAATAAAACCCATCTCCTGAAGACAC
 CAAAAAAAAGGAGCAGCTCCTCGTCAATGAACAAGCGTCACAAGAAAAGGGA
 GCACGTAAATAACCTCTCAATTGCTTCAGCATGAAAAGAACGGGAAGAAATGC
 15 AAGTCTACAGAGGAAAGTCAGCTGTTCCGGCTGCCATGGCAAGTCTTACATGG
 GCGAGGAAAAGCTGAACTGGATTCCAGTCTCGCGCTGTCATGCTCAGCTGCTT
 TAGGATGCGGCAATAGTCACCTGGATGAAAAGATAACAAGTTAGTCTTGAAGC
 AGTCGAGTGGACATCCAAAGTATCAAATCGAAAGCTTGTAAATGGGAAGGAA
 ATATAACCTCTACCCGGAAAAGTTGGTAGGCAAAATAATCCCAACGCCAGCAGA
 20 GCTCCGGAACGTTGCCGAAATTAGCAGAACGGAAAAGTTCTTGTACTCACCTCC
 GACAGTTCGAAGGTTCCAGCAGTAAGGAATGCGTGGCCATGGATTCCAGCGT
 CTCTGAATATCTTGAGGGGCAGATCAAAGAAAGGTAGCGAAGGCAGACACGG
 CCAGATCACCTCCCAAGTAATCCCTCCAGGGTCAGCCAGCCACTCTCCGAGTT
 ATTAAGGACATGCCTCCGCGCCTCTGTTGGCCAACTCCCCCTTAATCTGAAACCC
 25 AGCAGAGATGACGGTCCGCCAAGCTGCACACTGGAGAAGAACCTTCAAGA
 TAAAACCTCTGGCACTGATGAAGTCGAATTGATGAATCCCCCTGCAAGCGGT
 AAATGACACCCGCTCTACACCAACGTTGAGAGCAGCACTATAAAATCCCAAG
 GCACAGCACACGTACATCGAACTCCTGAGAGCAAACCCAGCGAATATTTCT
 30 AACGGACACCAACCAGCGAACCGAGCAGCGTCGCGTGGCCAAGCGAAGCAGAC
 GGCACGGCATCTGTGCGCTGCCCTGGACCCCTCTGAGAGTCCGCTCACCCT
 TGGACTTGCTCCGCTGCGCATCCAGAAATTGCGTGGCGGAGCGGAGACGTGA
 GCCGGCACGGCAGGCGGCCCTCCTCCTCACGGCACCGGAGCTACGGGGG
 ATTCCCTTCCCACCGCTCCTCGCTGCCCTCGCCC

35 [0637] SEQ ID NO:431
 AGTTGGGAAGGAAACGAATATAAGTCGATTGATTACTGTTAGTCGAGTC
 TATTTAAGGAATTGATGAAATGTTCTAACCTAACCTAACGCTAT
 GGCTGATATGGATCTGATTGGACTTGATTATCCATGATAAGTTAAGAGCAACT
 CAAAGAGGTTAGGTATATGGTTGAAAGGAAATTAGTTAATATTAGAAA
 40 AAAAAAGTGTATCCAATAGGCTCTATAAACAACTCTTCAAATTAGTGGCTTCT
 ATCCATCCACCTTGCTCTATTGGATAGCCTGATTACTCTCTATTAGTCC
 GTAGGTTAATGAGTCGTTGGATTAGCCTACACTTTCTGTAATCTATT
 GATAGTAGCTAAATCAGTAATTGGCTAGTATTAGCTATTCTCTGGAGTT
 GCTATAAGACCAAGAACATGAAATTGGAAGTTGAGGACCGGAGAGAACATGCA
 45 TGACAAATCCAGAGTATTGATGATGAAATTACCTATTACCGACTCTCCATT
 GTGTCCATTCTCATCATCCCCGGCGCTTCTGCATCCGGTACAGCTGACATGAC
 ACGTTCACGCGTTACATGGCTGATGGCTCACAAAGTCACCCCCACATGTCTAGTGT
 TCGCCAGGCAGATCGCTCGGCCCTGCGCTGCCGTCTTGCCGCCGCTGCTT
 GGGCCCTGCTGGCGCCGCTGCCGATCACACGGCCTACCGCGGTGCAGGCAGCGC
 50 CACCGAACCCGAGTCTTGTGCGATAGGTGGCAGTGGCAGTGGCACTGGCA
 CGGCACGCGATCGATCGCTCCGCTCATCTGCTGACAGTGGATAGAGCAGCGTTGG
 CGGTTGGGGCCGGATCTCGTGAAGCGGTGTCGCTGCTGTACTGTGCCGTATG
 GCGTGTGCGCTTCGCCATGTTCTTTCTTTCTTTCTTTGCTAGGG
 CGGTTCTCGCTGGTAACAGGGACCACTCGGTTGATCCGTTGAATTACTG

AAAGAGATGGGAATGGTCGCTGTGCCCGGACATTGAATGAGATGTTGTGAAG
 TGAATATGGCTTAGCCTTGCAGTGGGGCGCAATGCACGGCATGAACATATA
 ATTTCCGGTCAAACCTTTGTGTGGAAATGGATGCTAACAGAACACAAACCGGGTT
 TAAACCAGAGGCCGACACGGCACACACCGCGACATTCACGCCGCTTCCTCCGT
 5 CGCCACTCGGCACAAGGCTCATCAGTCGCCATGCCGATGCCGATGCGATCAACCGGAAG
 CGGATGGCCGCTTCTTACAATTGGCACAGGAACACTGGCCACTGCCCTGATG
 TGCAATTATGCCTGCAGAACAGCTTAGGCAACACACCGCAATAAACGAGCGAATGA
 CACGGAAAGCTGATGTGGTATGAATTATACAACATTATGGGCCAAATATTATTC
 TATCCACCATTGTGTAGCCACAGCATCGGTATTGAGTTGTGCGAGGACAAATCC
 10 CTCGTGAGGTCAAAACAGCAAATAATAAACCCTCTCCTGAAGACACCAAAAA
 AAAGGAGCAGCTCCTCGTCAATGAACAAGCGTCACAAGAAAAGGGAGCACGT
 AAATAACCTCTCAATTGCTTCAGCATGAAAAGAACGGGAAGAACATGCAAGTCT
 ACAGAGGAAAGTGCAGCTGTTCCGGCTGCCATGGCAAGTCCTACATGGCGAG
 15 GAAAAGCTGAACGGATTCCAGTCTCCGGCTGTCATGCTCAGCTTGCTTAGGA
 TGCGGCAATAGTCACCTGGATGAAAAGATAACAAGTTAGTCTTGAAGCAGTCG
 AGTGGACATCCAAAGTATCAAATCGAAAGCTTGTAAATGGGGAGGAAATATA
 CCTCTACCCGAAAAGTTGGTAGGCAGAACATAATCCAACGCCAGCAGAGCTCC
 GGAACGTTGCCGAAATTCAAGAGCCAAAAGTTCTGTACTCACCCCTCCGACAG
 TTCGCAAGGTTCCAGCAGTAAGGAATGCGTGGCCATGGATTCCAGCGTCTTG
 20 AATATCTGAGGGGAGATCAAAAGAAAGGTCAAGCAAGGAGACACGGCCAG
 ATCACCTCCCAAGTAATCCCTCCAGGGTCAGCCGAGCCACTCTCGAGTTATA
 AGGACATGCCTCCGCCCTGTTGGGCCAACTCCCCCTTAATCTGAAACCCAGCA
 GAGATGACGGTCCGCCAAGCTGCACACTGGAGAAGAACATACCTCAAGATAAA
 ACCTCTCTGGCACTGATGAAGTCGAATTGATGAAATCCCCCTGCAAGCGGTAAAAT
 25 GACACCCGCTCCTACACCAACGTTGAGAGCAGCACTATAAAATCCAAAGGCAC
 AGCACCAACGTACATCGAACTCCTGAGAGCAACCCAAACGGCAATATTTGTAAAT
 AGTGTGGTCAGAACTGAGAAGATCAGATAAAATTATACACTGATGCAATTATTT
 CATAGTTGCCCATGAACTGTAAGGGTAGACAAAGCAAAAGTAAGACATGA
 AGGGCAAGAGAATAACCTGCCGAAATATCTCAATCCTTGCTATTCCATAGACC
 30 ACCAACTTGAGAAGTTGACTGAAACGCATATCCTTCGTTGGCCTAAGATGTGAA
 TCCCTTTATCAATCTGTATGTGACTTCAATGCAAGAAAGAAGGTTATGCCCTAA
 CTGCCTCCTATGGCCTTGATGAGACACGTGATGGATCAGTTAAGGTACGCCAC
 GCAAGGTTGTATGACAAGTCATGGTCTTGTGACAGCAAACCAAATGAAAGGC
 CAAGTAGGCCTCTGTATGATGAAAACCTCAGCAATCTGTGATGACAAAGA
 35 TGCCCGAGCCATCAATGGTGTGGTATTGATTTAACCTCGGTAGGCAGACTCCA
 ACACCAACCTCTGTTGGTCCCAACCAAAGGATCCTGATGCATCCCAGATGT
 CACCATAGCCAAACAAAGTTCTCAACTTAAGTGACCCCTCCAGCGACCAAGATCT
 TGCTACAAGAGTGGCAAGCACAGTC
 40 [0638] SEQ ID NO:464
 GTGCATTCGGATTACTGTTAGTCGAGTCATATTAAAGGAATTCAATTGAAATGTT
 CTAACCTAACCTAAAGTATTAGGCAGCTATGGCTGATATGGATCTGATTGGACTTG
 ATTTATCCATGATAAGTTAAGAGCAACTCAAAGAGGTTAGGTATATATGGTTTT
 GTAAAGGTAAATTAGTTAATATTAGAAAAAAAAGTGTATCCAATAGGCTCTAT
 45 AAACA

[0639] All patents, patent applications and publications mentioned herein are
 hereby incorporated by reference in their entirety.

[0640] Although disclosure has been provided in some detail by way of
 50 illustration and example for the purposes of clarity of understanding, it will be
 apparent to those skilled in the art that various changes and modifications can be

practiced without departing from the spirit or scope of the disclosure. Accordingly, the foregoing descriptions and examples should not be construed as limiting.

CLAIMS

What is claimed is:

1. A double-stranded polynucleotide comprising:

5 an exogenous nucleic acid sequence;
a sequence comprising one or more target sites for one or more nucleases,
wherein the one or more target sites are not within the exogenous nucleic acid
sequence and further wherein the double-stranded polynucleotide does not comprise
homology arms.

10

2. The double-stranded polynucleotide of claim 1, wherein the exogenous
nucleic acid sequence is at least 1 kb in length.

15

3. The double-stranded polynucleotide of claim 1 or claim 2, wherein the
exogenous nucleic acid sequence polynucleotide is a plasmid.

4. The double-stranded polynucleotide of any of claims 1 to 3, wherein the
exogenous nucleic acid sequence comprises a transgene.

20

5. The double-stranded polynucleotide of any of claims 1 to 4, wherein the
sequence comprises two target sites and a spacer of at least 5 nucleotides between the
two target sites.

25 5.

6. A cell comprising the double-stranded polynucleotide of any of claims 1 to
30 5.
7. A method of integrating an exogenous nucleic acid sequence
polynucleotide into an endogenous locus of a cell, the method comprising;
introducing a double-stranded polynucleotide according to any of claims 1 to 5
into the cell;
introducing one or more nucleases into the cell, wherein the nucleases cleave
the double-stranded polynucleotide and cleave the endogenous locus such that the

exogenous nucleic acid sequence polynucleotide is integrated into the endogenous locus.

8. The method of claim 7, wherein the exogenous nucleic acid sequence
5 polynucleotide is integrated in a forward orientation.

9. The method of claim 7, wherein the exogenous nucleic acid sequence
polynucleotide is integrated in a reverse orientation.

10 10. The method of any of claims 7 to 9, wherein the same nucleases cleave
the endogenous locus and the donor polynucleotide.

11. The method of any of claims 7 to 9, wherein different nucleases cleave the
endogenous locus and the donor polynucleotide.

15 12. The method of any of claims 7 to 11, wherein the exogenous nucleic acid
sequence polynucleotide is integrated into the endogenous locus via homology-
independent mechanisms.

20 13. The method of any of claims 7 to 12, wherein the nucleic acid comprising
the two target sites and the spacer is non-naturally occurring such that the target sites
are not re-created following integration of the exogenous nucleic acid sequence
polynucleotide.

25 14. The method of any of claims 7 to 13, wherein the nucleases generate a
deletion in the endogenous locus and the exogenous nucleic acid sequence
polynucleotide is integrated into the deletion.

15. The method of any of claims 7 to 14, wherein the cell is a eukaryotic cell.

30 16. The method of claim 15, wherein the cell is plant or mammalian cell.

17. The method of claim 16, wherein the plant cell is a dicotyledonous or a monocotyledonous plant cell.

18. A transgenic organism comprising an exogenous nucleic acid sequence 5 polynucleotide integrated according to the method of any of claims 7 to 17.

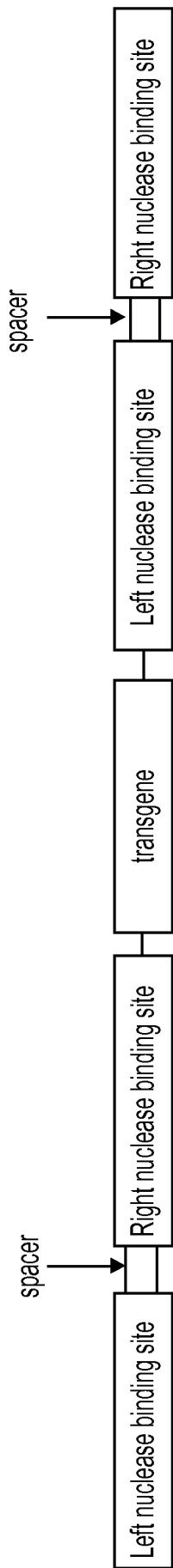
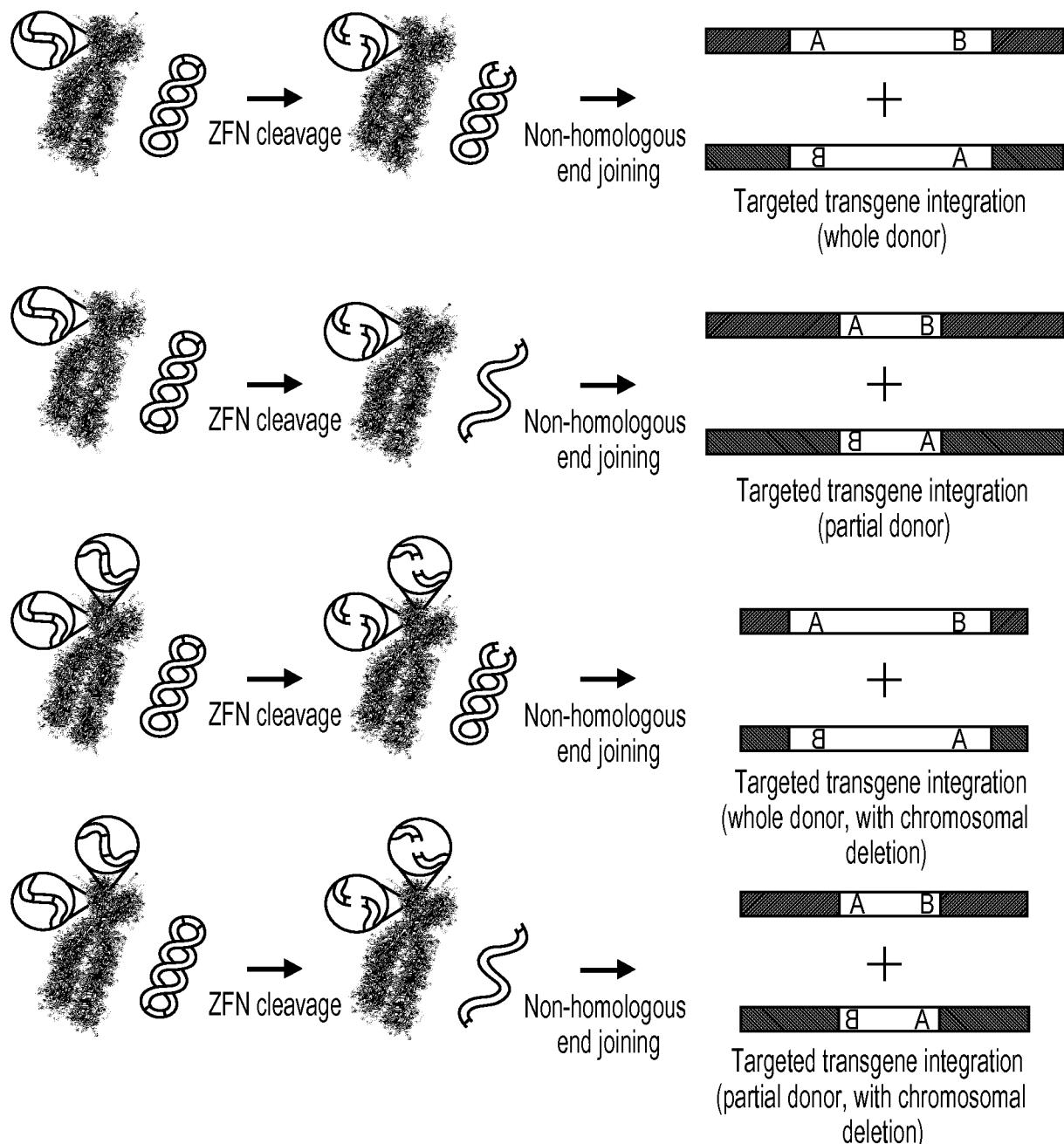
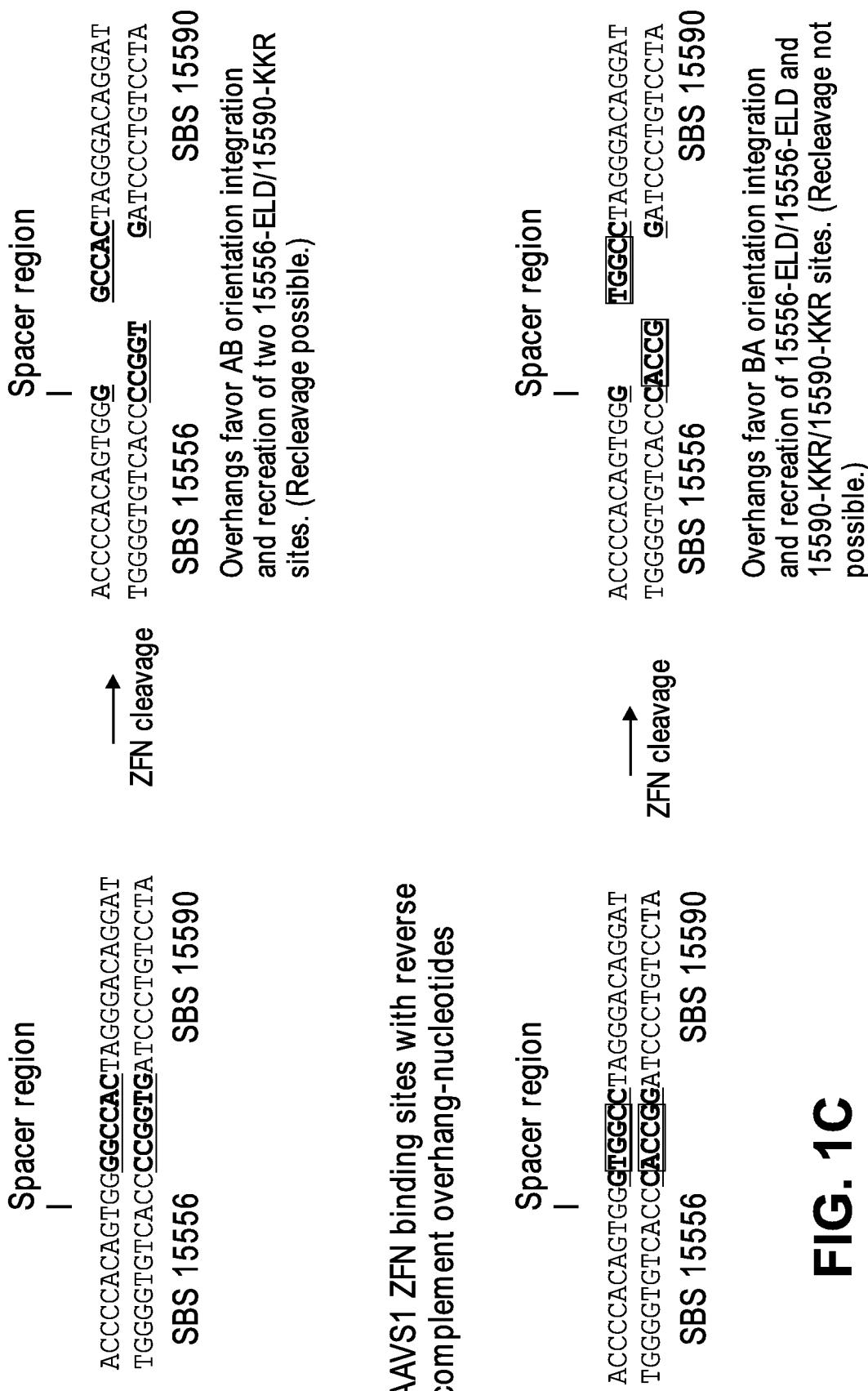


FIG. 1A

**FIG. 1B**

AAVS1 ZFN binding sites and spacer

**FIG. 1C**

Junction:	BA	BA	AB	AB
Overhang:	GCCA (w.t.)	TGGC (r.c.)	GCCA (w.t.)	TGGC (r.c.)

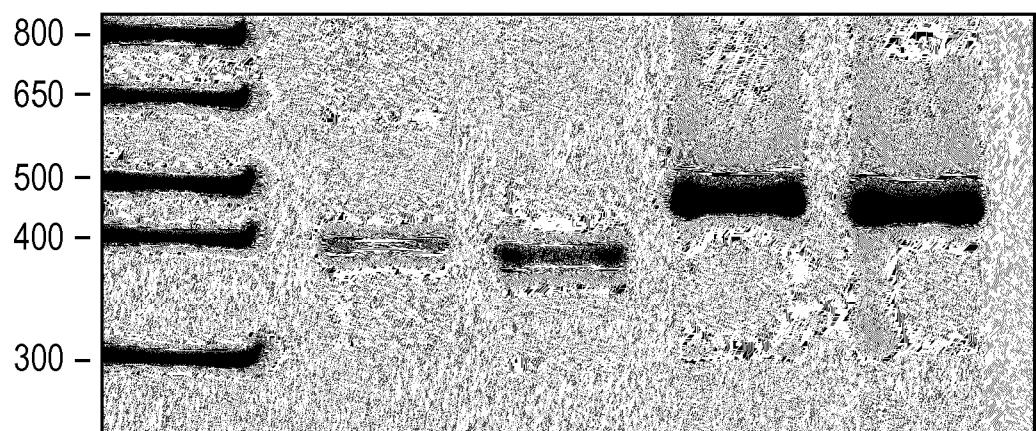


FIG. 1D

5/61

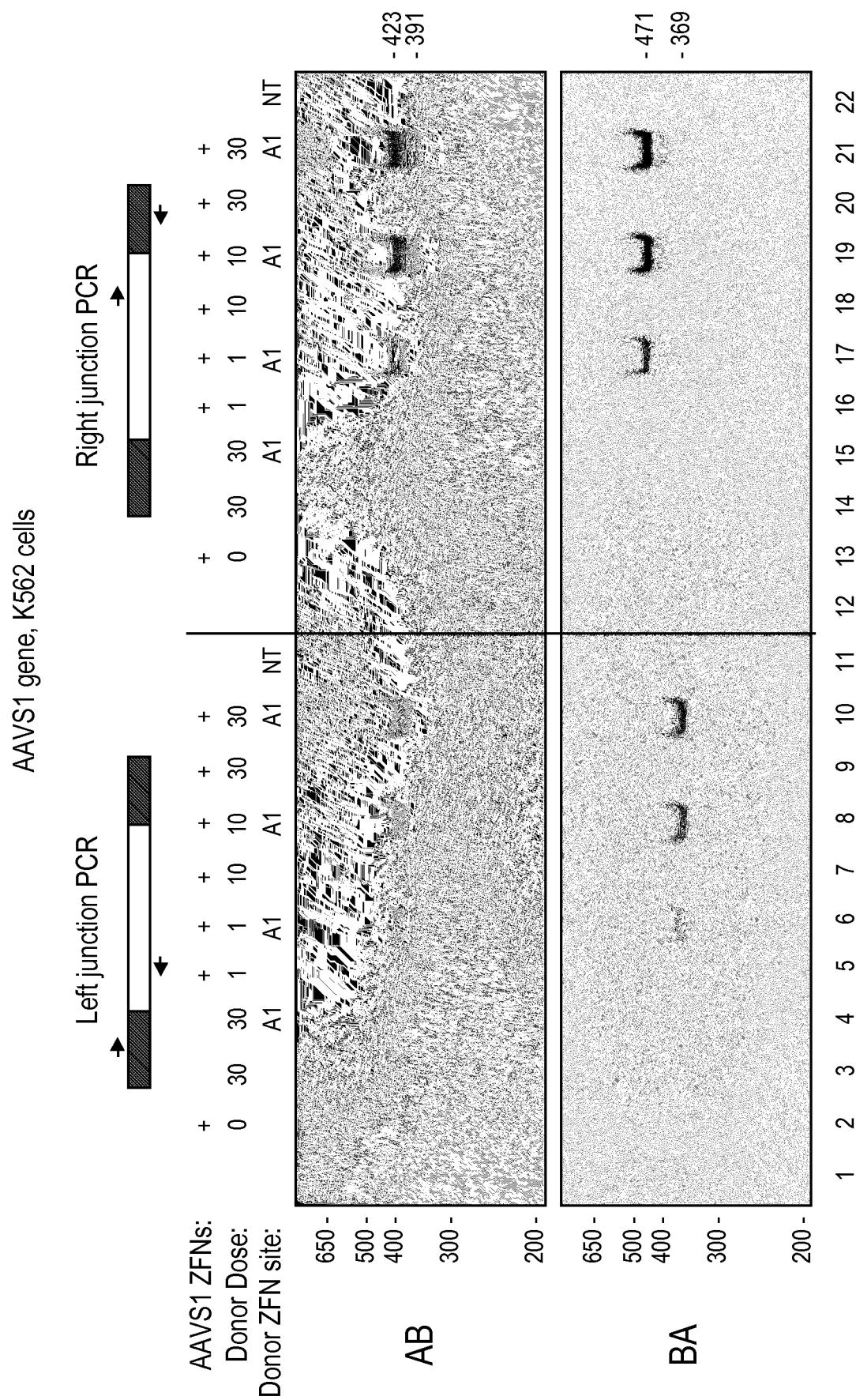
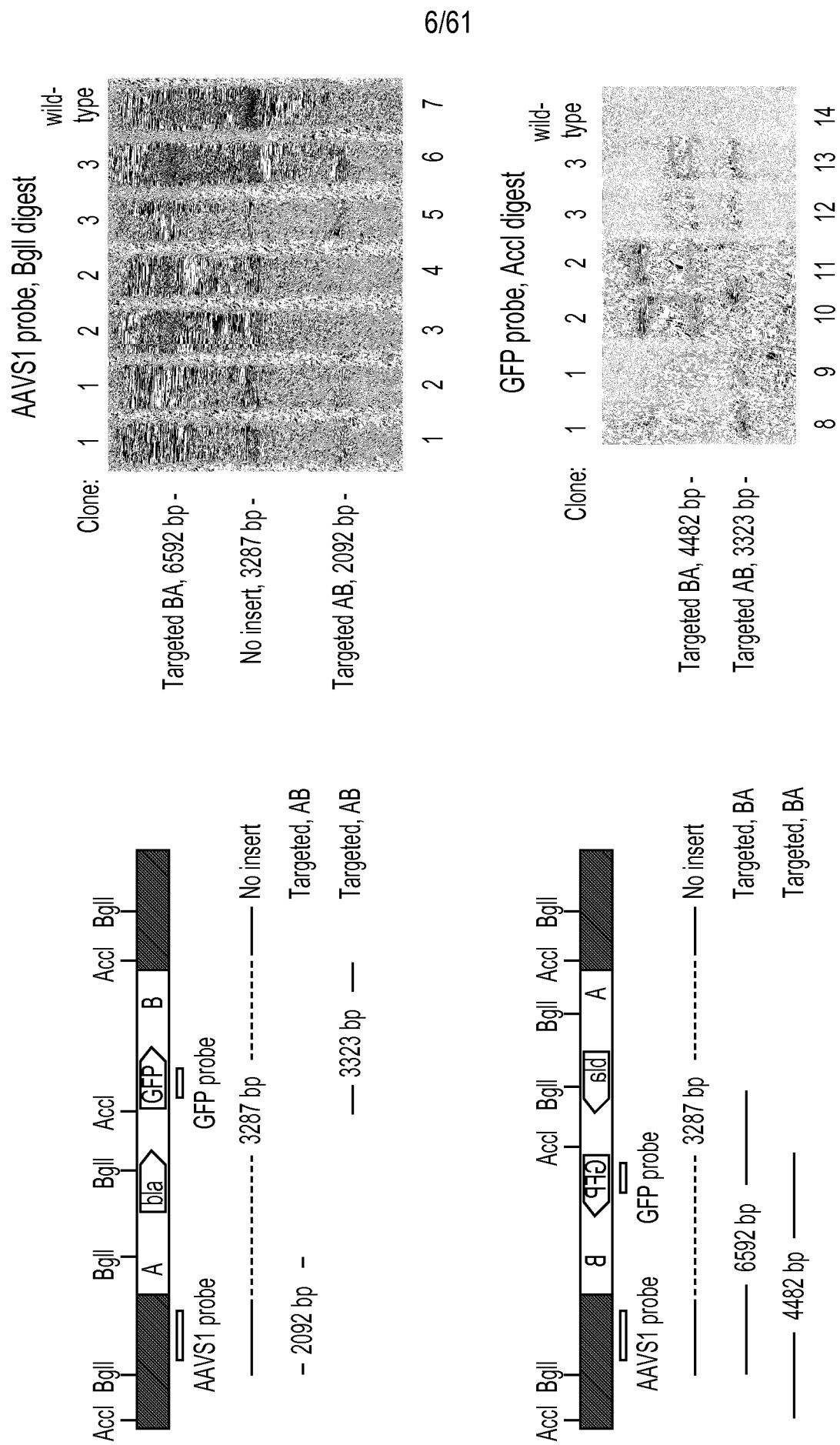


FIG. 1E



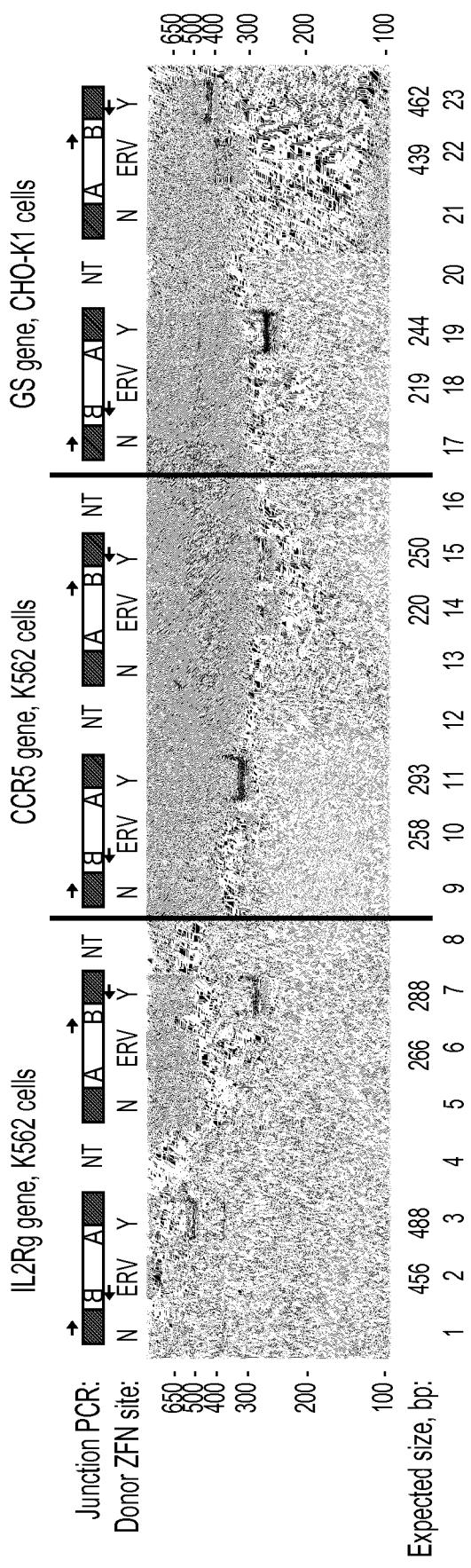


FIG. 2A

8/61

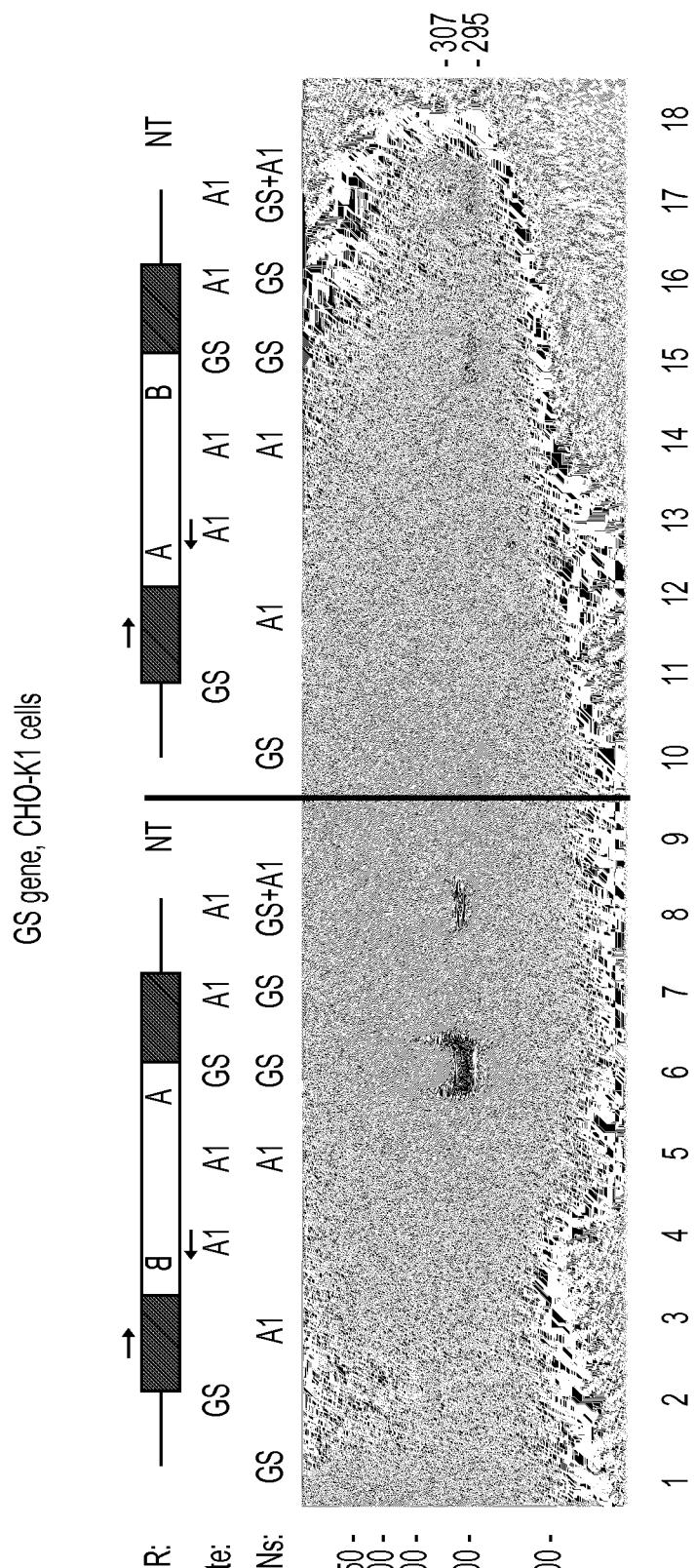
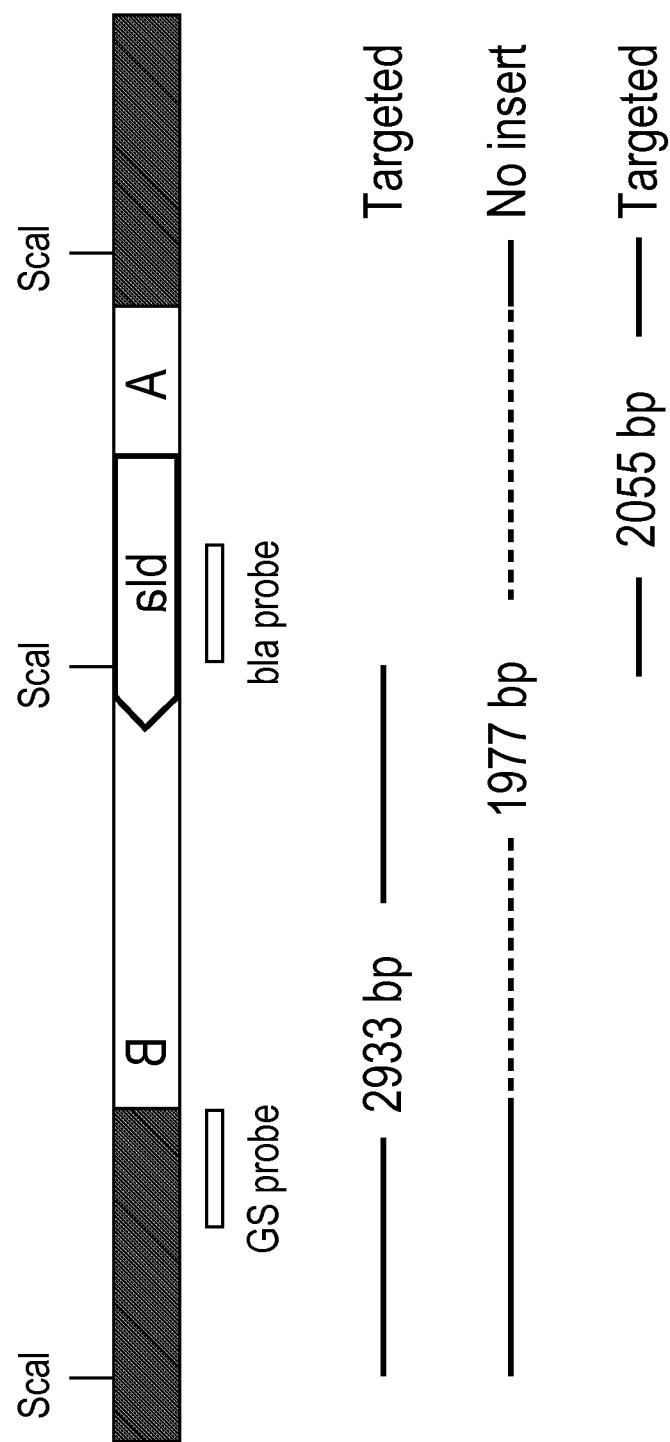
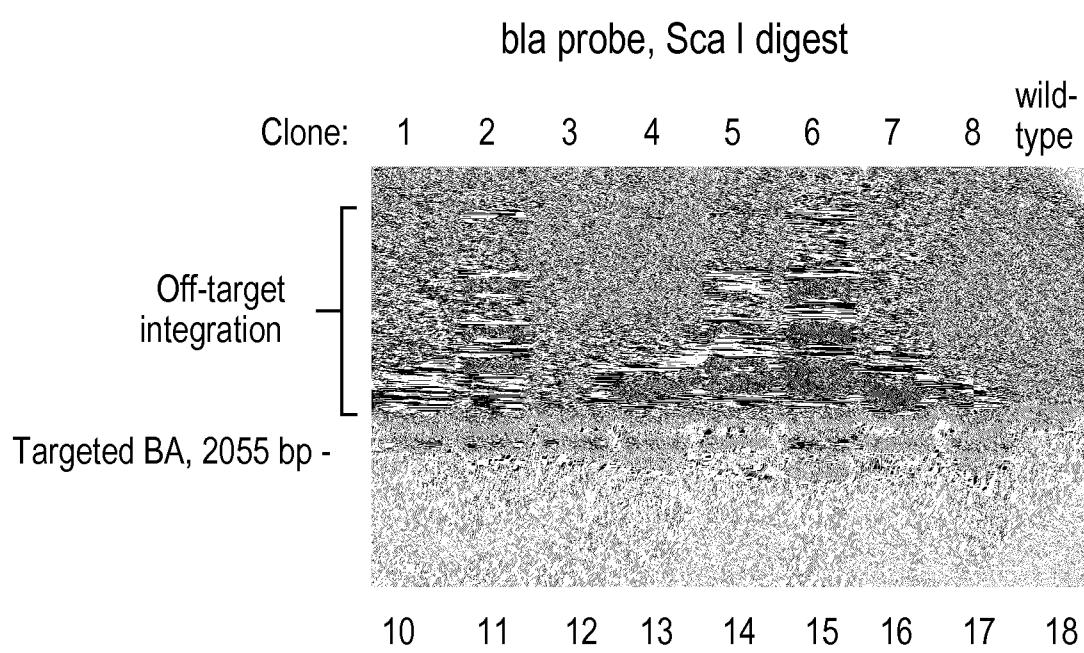
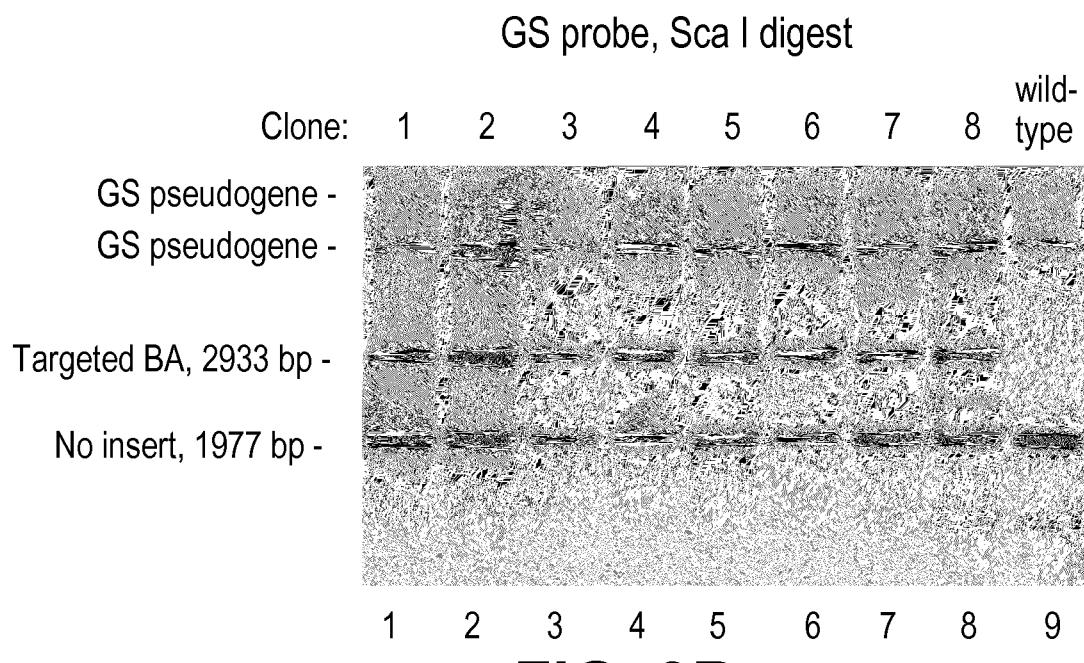


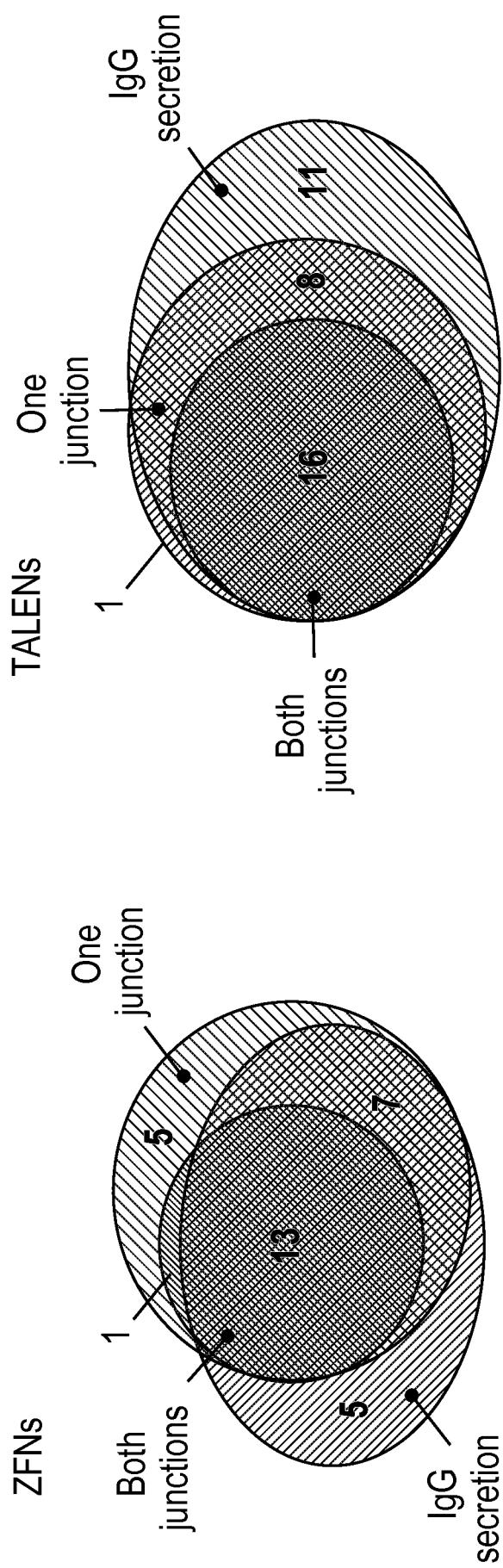
FIG. 2B

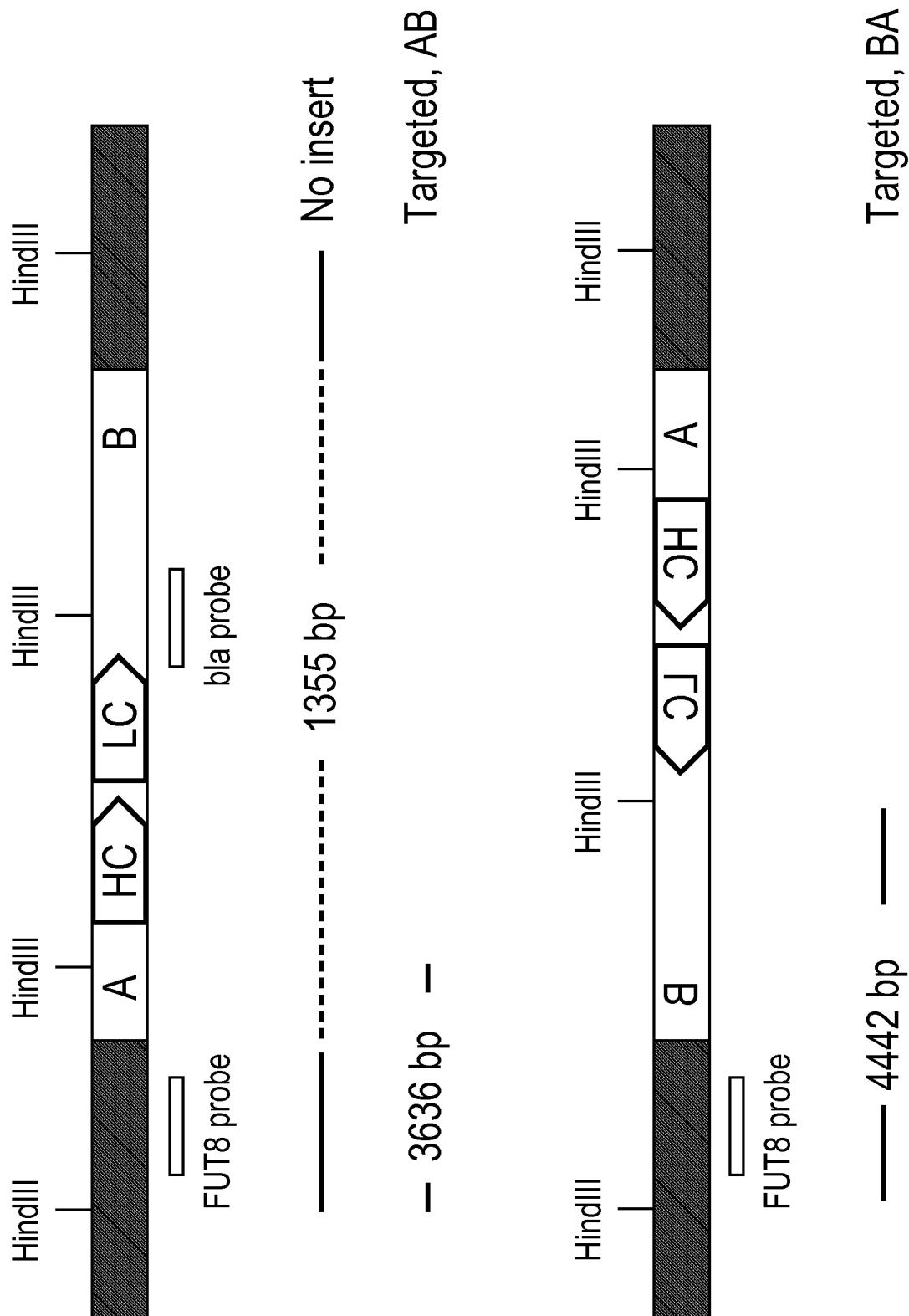
**FIG. 3A**

10/61



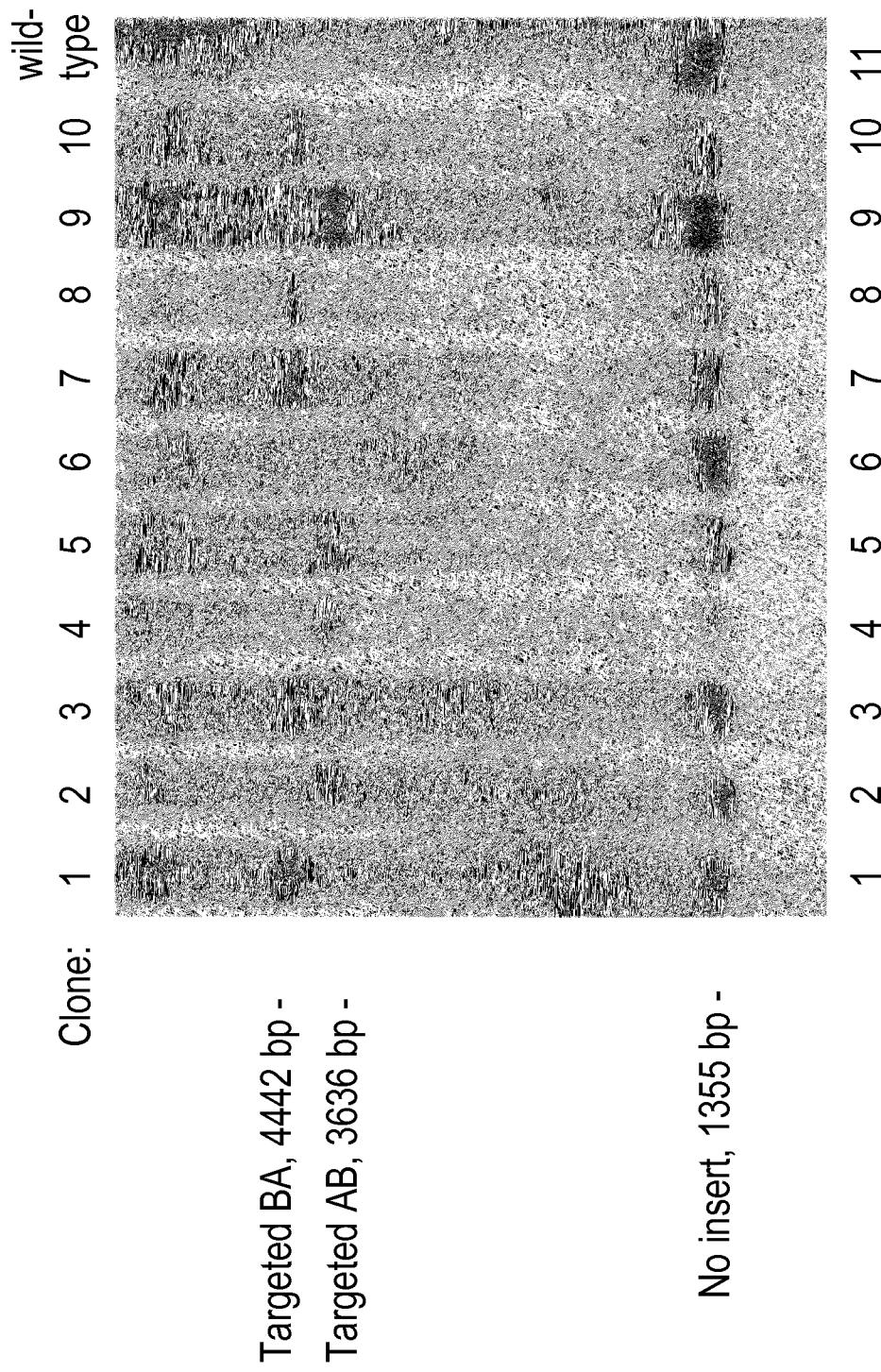
11/61

**FIG. 4A**

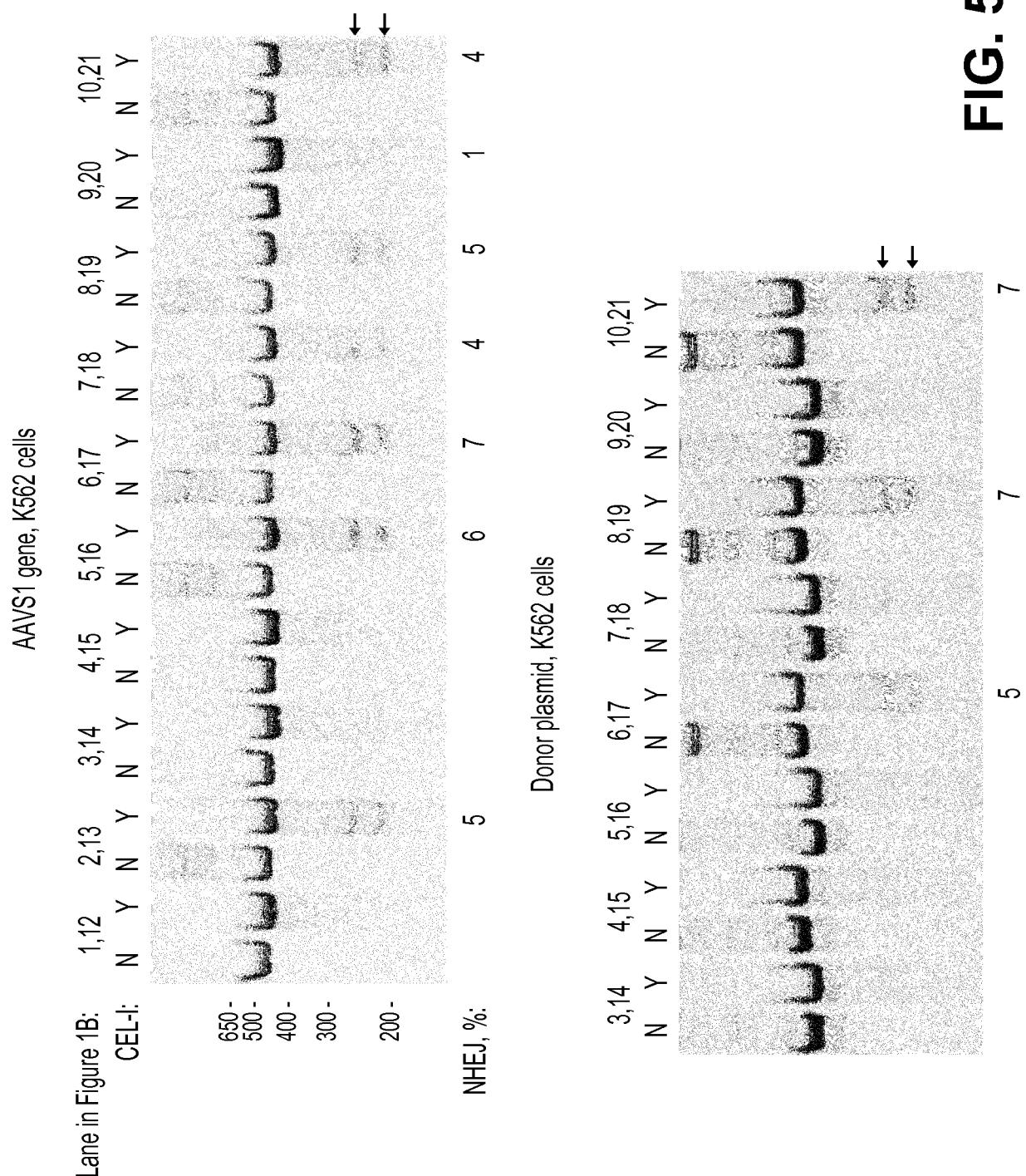
**FIG. 4B**

13/61

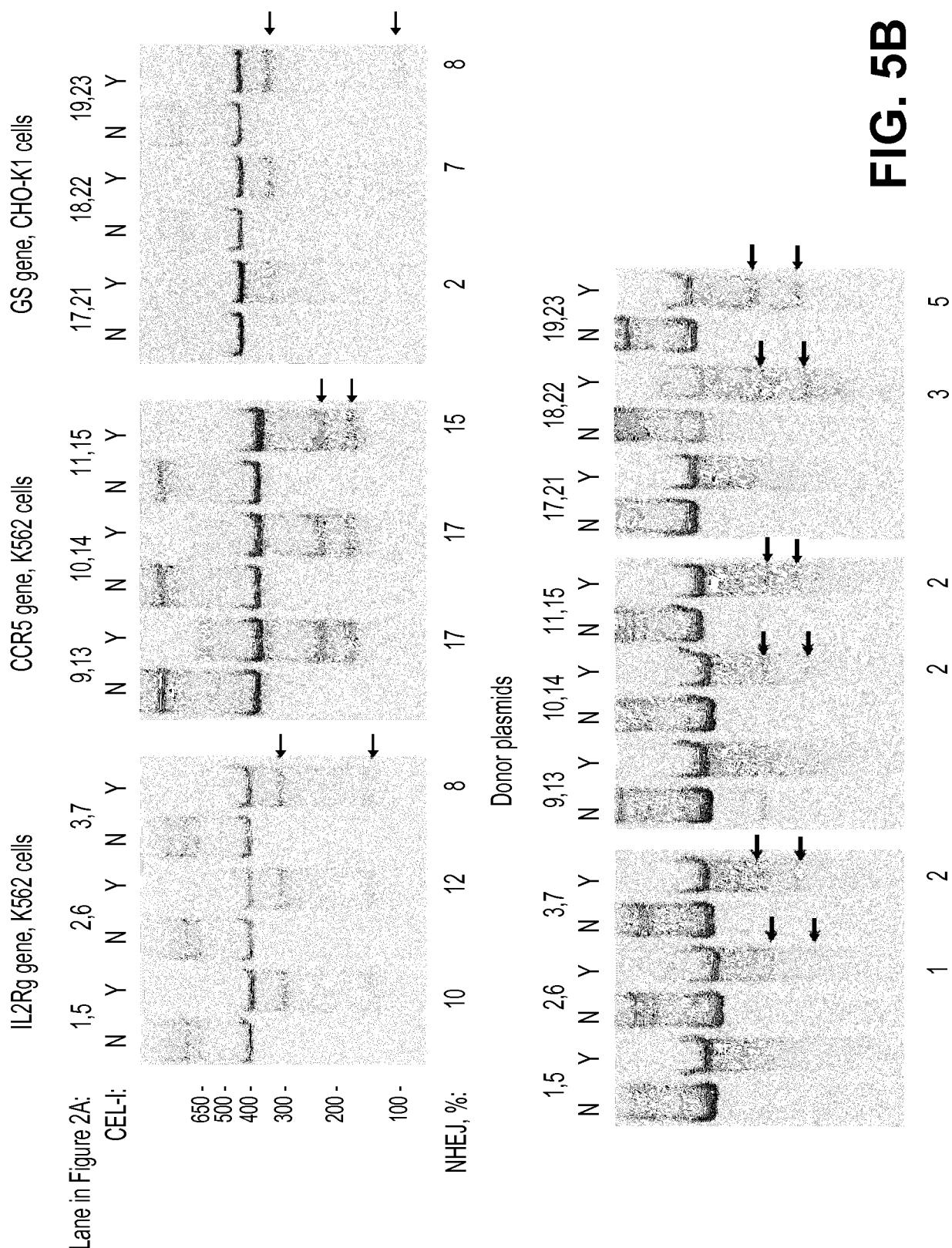
FUT8 probe, Hind III digest

**FIG. 4C**

14/61



15/61

**FIG. 5B**

1 2 2 3 5

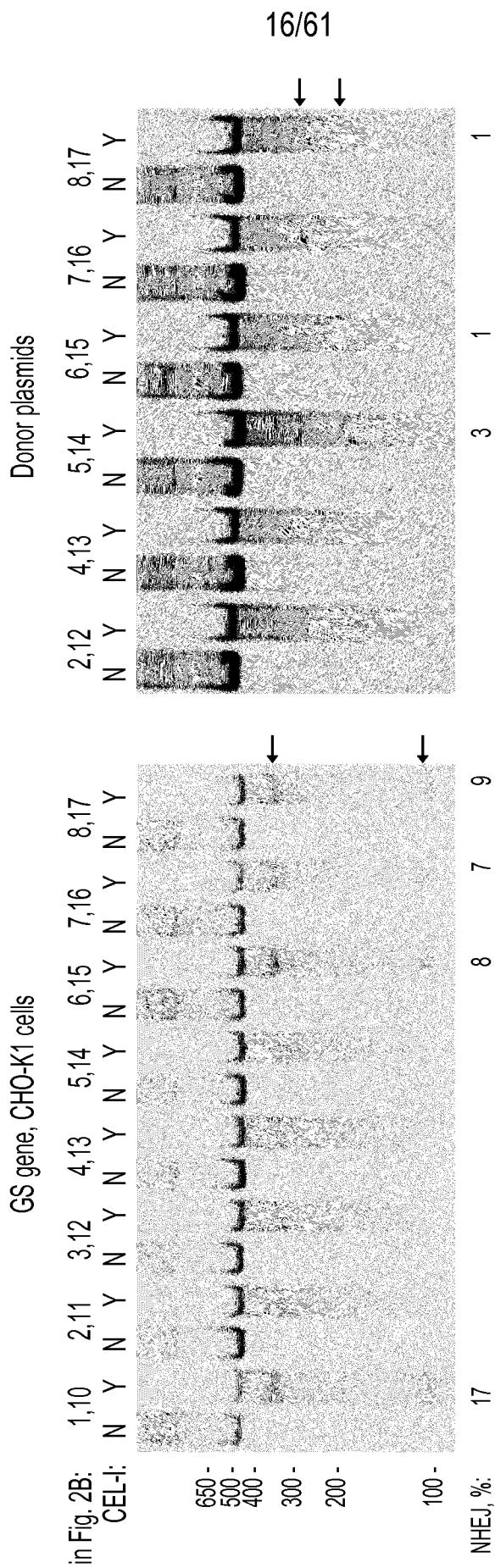
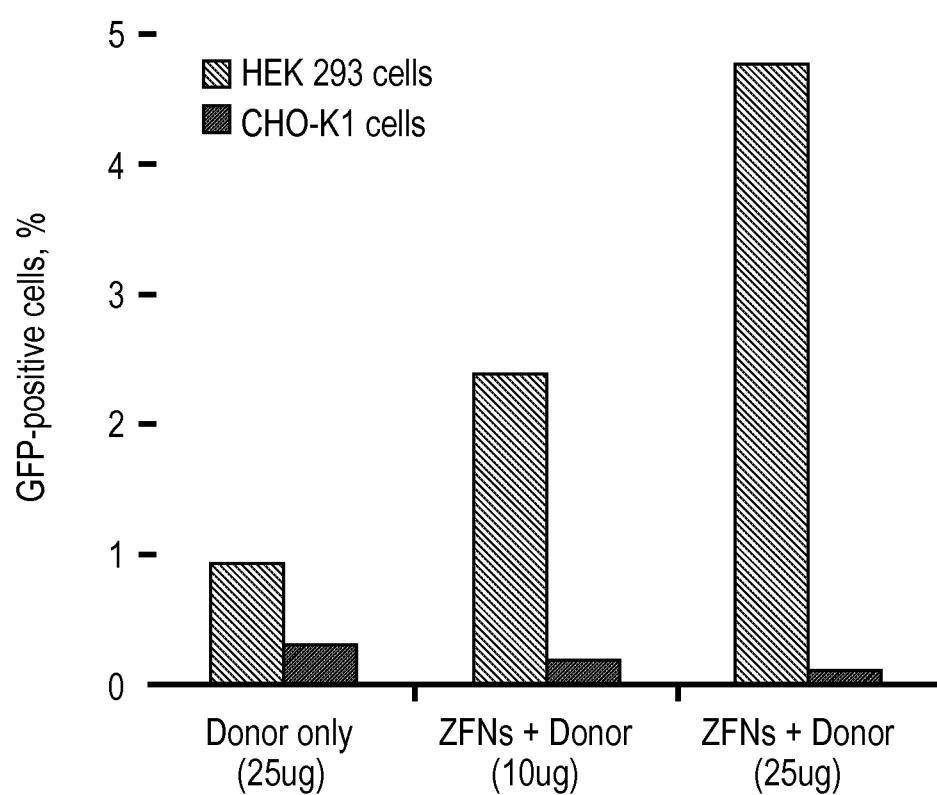


FIG. 5C

17/61

Targeted integration in HEK 293 and CHO-K1 cells

**FIG. 6**

<i>AAVS1</i> single cell-derived clones, AB insertions		
	SEQ ID NO:	
Left (5') junction		
Expected	TTATCTGTCCCCTCCACCCACAGTGG	28
Clone 1	TTATCTGTCCCCTCCACCCACAGTGG	28
Clone 3	TTATCTGTCCCCTCCACCCACAGTGG	28
Right (3') junction		
Expected	CGAATTCTGCAGAT <u>ACCCCCACAGTGGCCACTAGGGACAGGAT</u> TGGTGACAGAAAAGCC	29
Clone 1	CGAATTCTGCAGAT <u>ACCCCCACAGTGGCCACTAGGGACAGGAT</u> TGGTGACAGAAAAGCC	29
Clone 3	CGAATTCTGCAGAT <u>ACCCCCACAGTGGCCACTAGGGACAGGAT</u> TGGTGACAGAAAAGCC	29
<i>AAVS1</i> single cell-derived clones, BA insertions		
Left (5') junction		
Expected	TTATCTGTCCCCTCCACCCACAGTGG <u>CCCACTGTGGGTGATCTGCAGAATT</u> CG	30
Clone 2	TTATCTGTCCCCTCCACCCACAGTGG <u>.....GTGATCTGCAGAATT</u> CG	31
Clone 3	TTATCTGTCCCCTCCACCCACAGTGG <u>.....GTGATCTGCAGAATT</u> CG	31
Right (3') junction		
Expected	CCAGTGTGATGGATA <u>ATCC'TGTCCTA</u> GCTAGGGACAGGATGGTGACAGAAAAGC	32
Clone 2		
Clone 3		

FIG. 7

19/61

FIG. 8A

FIG. 8B

SEQ ID NO	GS pool alleles, BA insertions	III2Rg pool alleles, AB insertions
54	Left (5') junction	Right (3') junction
55	AGAACACCTTGACCCAAAGCCCATTCTGGAA: :: :: <u>TTCCCCAGGAATGGGCTTGGGT</u> CATGAAATT	TCGATGTTCTGTGT <u>GGAGCCGCTTAACCC</u> : : <u>ACTCTGTGGAAGT</u> GCTCAGCATGGAG
56	AGAACACCTTGACCCAAAGCCCATTCTGGAAACTGG <u>TTCCCCAGGAATGGGCTTGGGT</u> CATGAAATT	TCGATGTTCTGTGT <u>GGAGCCGCTTAACCC</u> : : <u>ACTCTGTGGAAGT</u> GCTCAGCATGGAG
56	AGAACACCTTGACCCAAAGCCCATTCTGGG: :: :: <u>TTCCCCAGGAATGGGCTTGGGT</u> CATGAAATT	TCGATGTTCTGTGT <u>GGAGCCGCTTAACCC</u> : : <u>ACTCTGTGGAAGT</u> GCTCAGCATGGAG
57	AGAACACCTTGACCCAAAGCCCATTCTGGGAAAC: :: <u>TTCCCCAGGAATGGGCTTGGGT</u> CATGAAATT	TCGATGTTCTGTGT <u>GGAGCCGCTTAACCC</u> : : <u>ACTCTGTGGAAGT</u> GCTCAGCATGGAG
58	AGAACACCTTGACCCAAAGCCCATTCTGGGAAACTAG <u>TTCTAGGAATGGGCTTGGGT</u> CATGAAATT (3X)	TCGATGTTCTGTGT <u>GGAGCCGCTTAACCC</u> : : <u>ACTCTGTGGAAGT</u> GCTCAGCATGGAG
59	AGAACACCTTGACCCAAAGCCCATTCTGGGAAACAG <u>TTCCCCAGGAATGGGCTTGGGT</u> CATGAAATT	TCGATGTTCTGTGT <u>GGAGCCGCTTAACCC</u> : : <u>ACTCTGTGGAAGT</u> GCTCAGCATGGAG
60	AGAACACCTTGACCCAAAGCCCATTCTGGGAAAG: :: <u>TTCCCCAGGAATGGGCTTGGGT</u> CATGAAATT	TCGATGTTCTGTGT <u>GGAGCCGCTTAACCC</u> : : <u>ACTCTGTGGAAGT</u> GCTCAGCATGGAG

8C
FIG

GS single cell-derived clones, BA orientation		SEQ ID
Left (5')	junction	
Expected	AGCAACCTTTGACCCCAAGCCCCATTCTGGAA::: :	72
Clone 1	AGCAACCTTTGACCCCAAGCCCCATTCTGGGAACCAAG	73
Clone 2	AGCAACCTTTGACCCCAAGCCCCATTCTGGAAAC: TG	74
Clone 3	AGCAACCTTTGACCCCAAGCCCCATTCTGGGAACCAAG	75
Clone 4	AGCAACCTTTGACCCCAAGCCCCATTCTGGGAACCAAG	76
Clone 5	AGCAACCTTTGACCCCAAGCCCCATTCTGGAAAC: TG	77
Clone 6	AGCAACCTTTGACCCCAAGCCCCATTCTGGGAAC: TG	77
Clone 7	AGCAACCTTTGACCCCAAGCCCCATTCTGGGAACCAAG	78
Clone 8	AGCAACCTTTGACCCCAAGCCCCATTCTGGGAACCAAG	78
Right (3')	junction	
Expected	TTGATTGGTATGGCAGCCTGCACCAT::: :AATGGTGCAGGCCATACCAACTTTAGCACCA	79
Clone 1		
Clone 2		
Clone 3		
Clone 4	TTGATTGGTATGGCAGCCTGCACCAT::: :AATGGTGCAGGCCATACCAACTTTAGCACCA	80
Clone 5		
Clone 6		
Clone 7		
Clone 8		

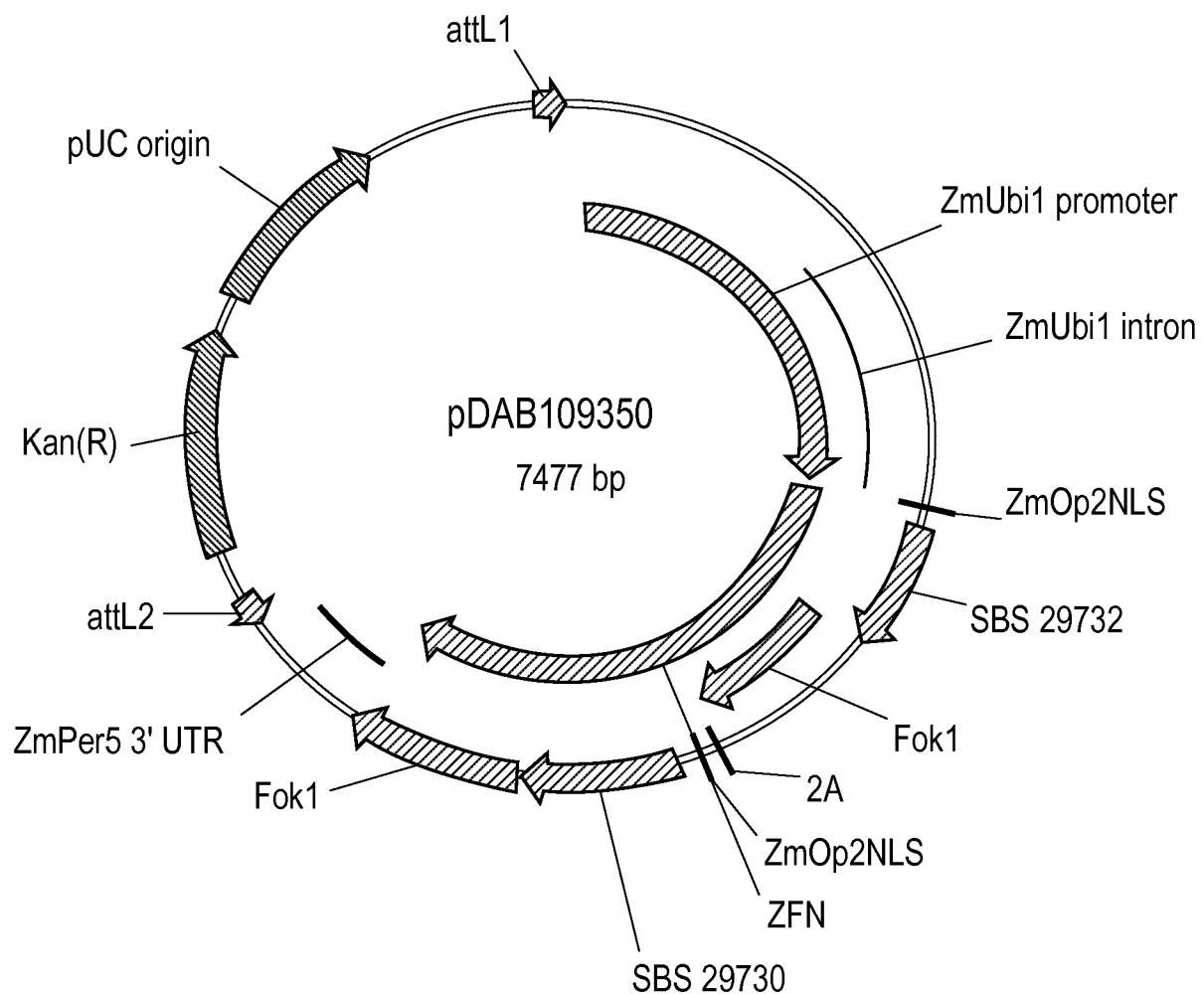
23/61

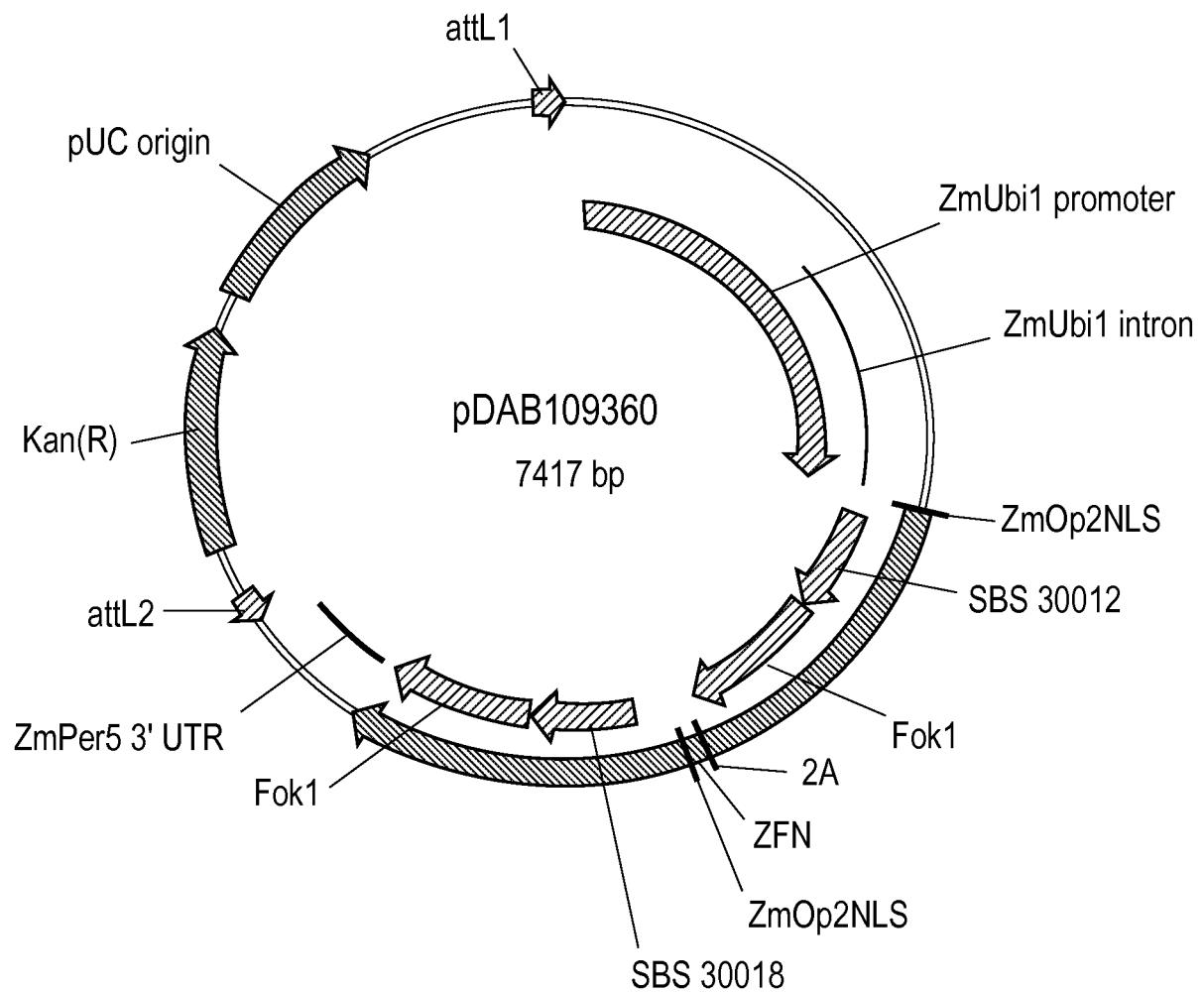
FIG. 10A

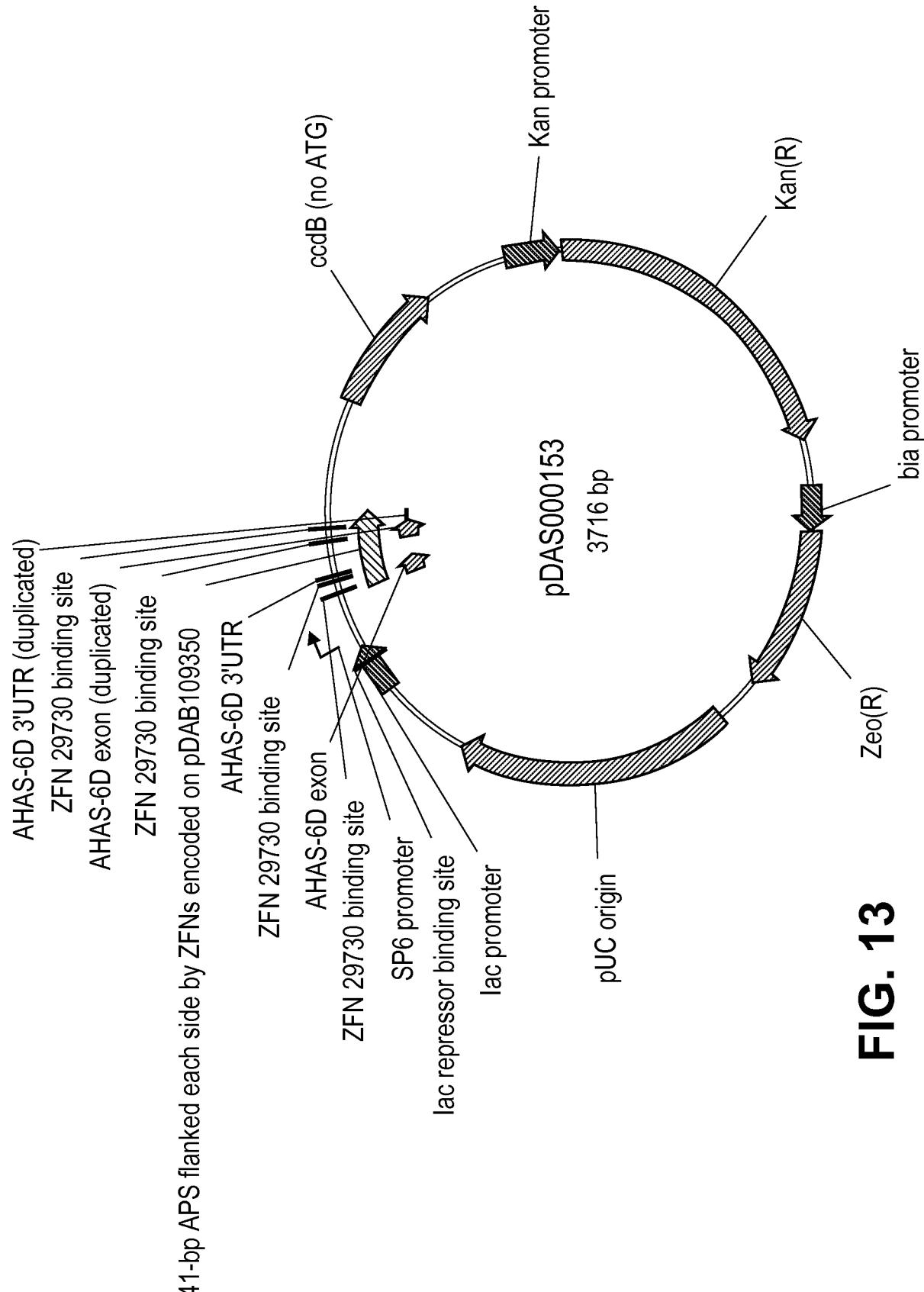
24/61

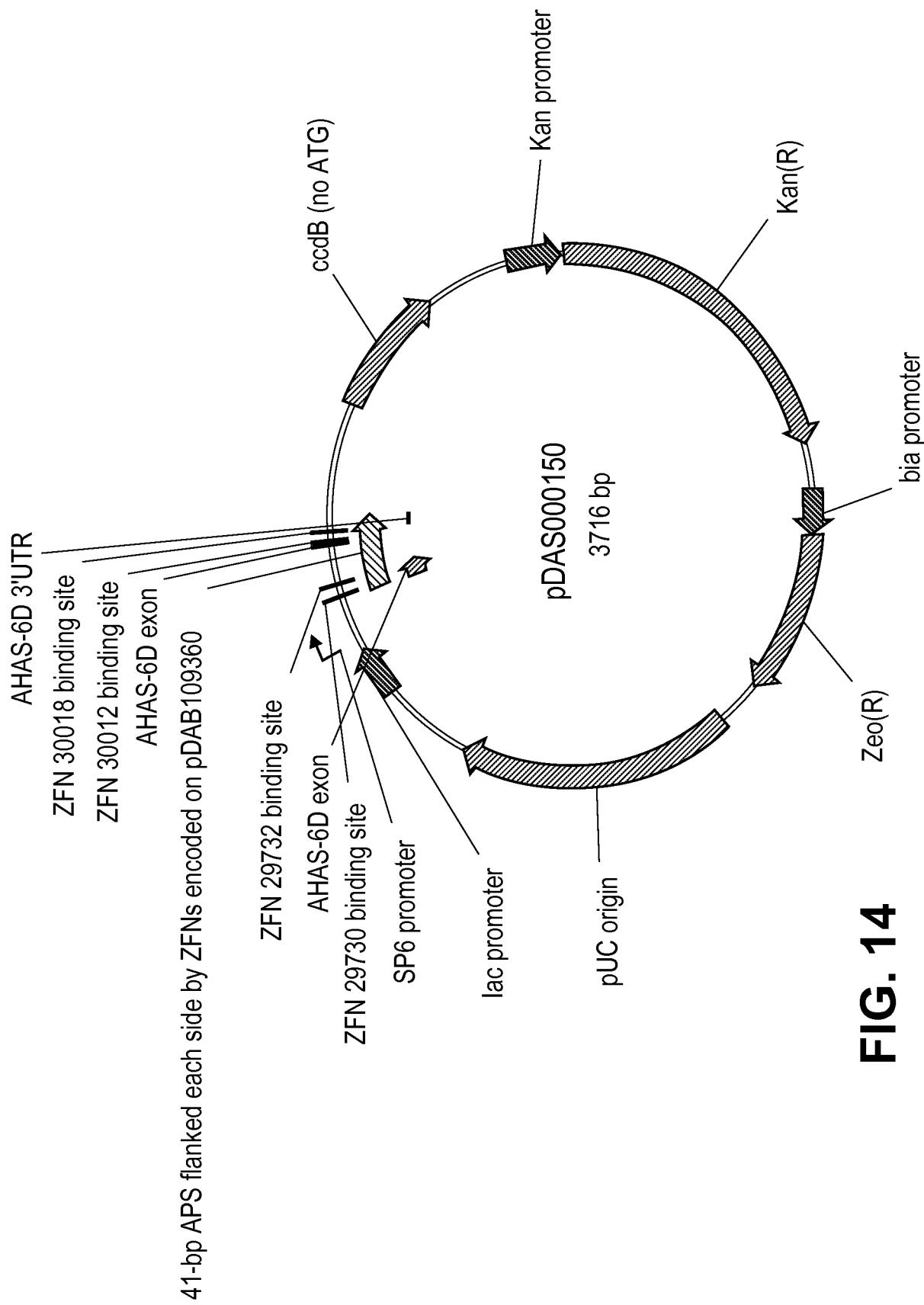
		SEQ ID
<i>FUT8</i>	single cell-derived clones, AB orientation, cleaved by TALENS	
Left (5') junction		
Expected	TGGATAAAAAAGAG <u>TGTATCTGCCCACTGATGACCCTTATTGTTAAAGGAGGCCAAGACAAAGTAAAGGCCGGAATT</u> C	97
(X2)	TGGATAAAAAAGAG <u>TGTATCTGCCCACTGATGACCCTT:::TTAAAGGAGGCCAAGACAAAGTAAAGGCCGGAATT</u> C	98
(X2)	TGGATAAAAAAGAG <u>TGTATCTGCCCACTT:::CTTGTAAAGGAGGCCAAGACAAAGTAAAGGCCGGAATT</u> C	99
(X2)	TGGATAAAAAAGAG <u>TGTATCTGCCCACTGATG:::TTAAAGGAGGCCAAGACAAAGTAAAGGCCGGAATT</u> C	100
Right (3') junction		
Expected	TTAGGGGCCGTGT <u>ATCTGCCCACTGATGACCCTTCTTTGTAAAGGAGGCCAAGACAAAGTAAGTAGACCAACAAGTG</u>	101
<i>FUT8</i>	single cell-derived clones, BA orientation, cleaved by TALENS	
Left (5') junction		
Expected	TGGATAAAAAAGAG <u>TGTATCTGCCCACTGATGACCCTTGAAGGGTCAATCAGTGGCCAGATACACGGGCCGCTAAATTCAA</u>	102
(X10)	TGGATAAAAAAGAG <u>TGTATCTGCCCACTGATGACCCTTCA:::TACA CGGGCGCTAAATTCAA</u>	103
Right (3') junction		
Expected	GAATTGGGGCCTT <u>TACTTTGTCTTTGCCTTAAACAAATCTTGTAAAGGAGGCCAAGACAAAGTAAGTTAGACCAA</u>	104
(X2)	GAATTGGGGCCTT <u>TACTTTGTCTTTGCCTTAA:::TTTGTAAAGGAGGCCAAGACAAAGTAAGTTAGACCAA</u>	105
(X8)	GAATTGGGGCCTT <u>TACTTTGTCTTTGCCTT:::CTTTGTAAAGGAGGCCAAGACAAAGTAAGTTAGACCAA</u>	106

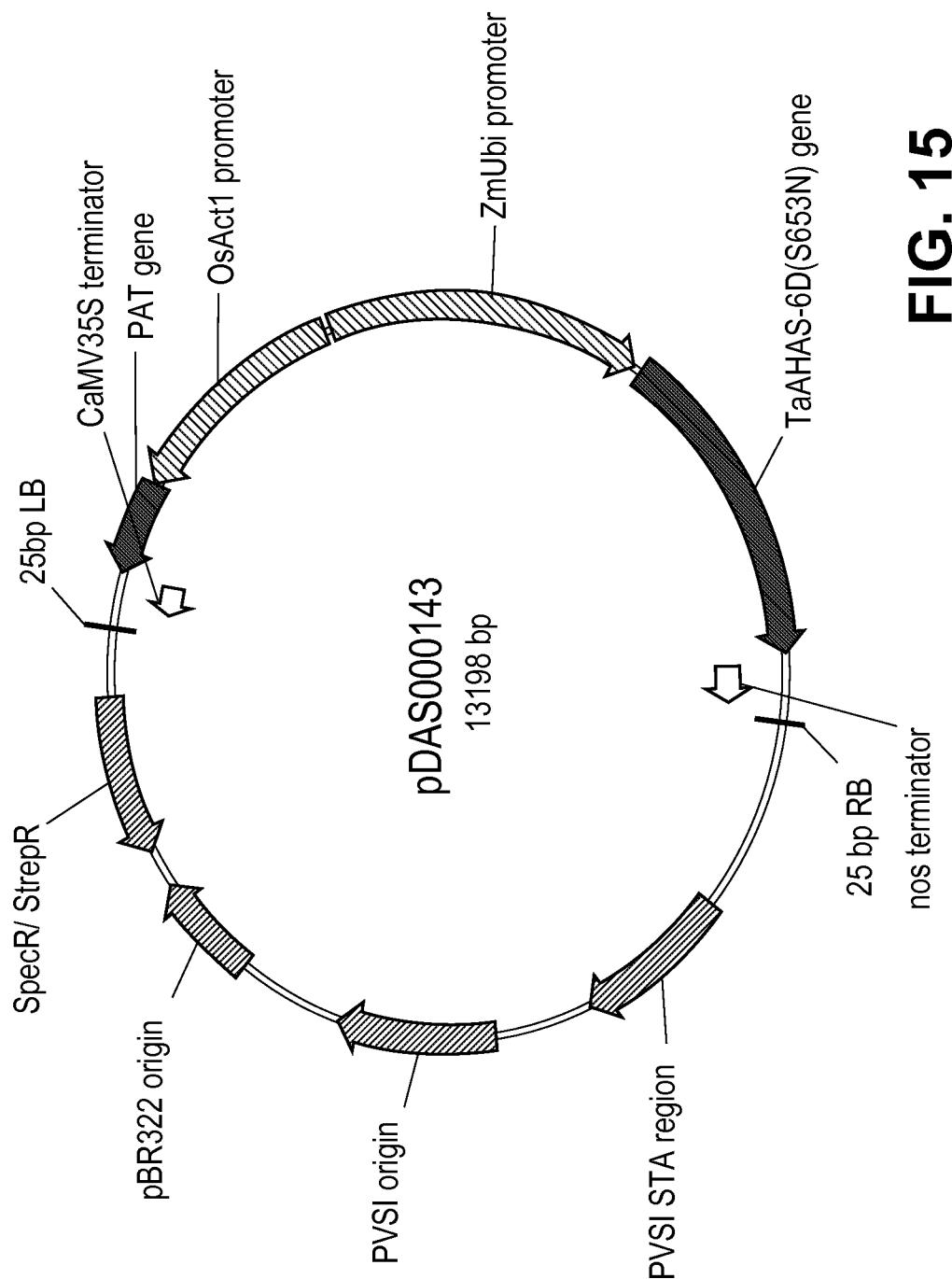
FIG. 10B

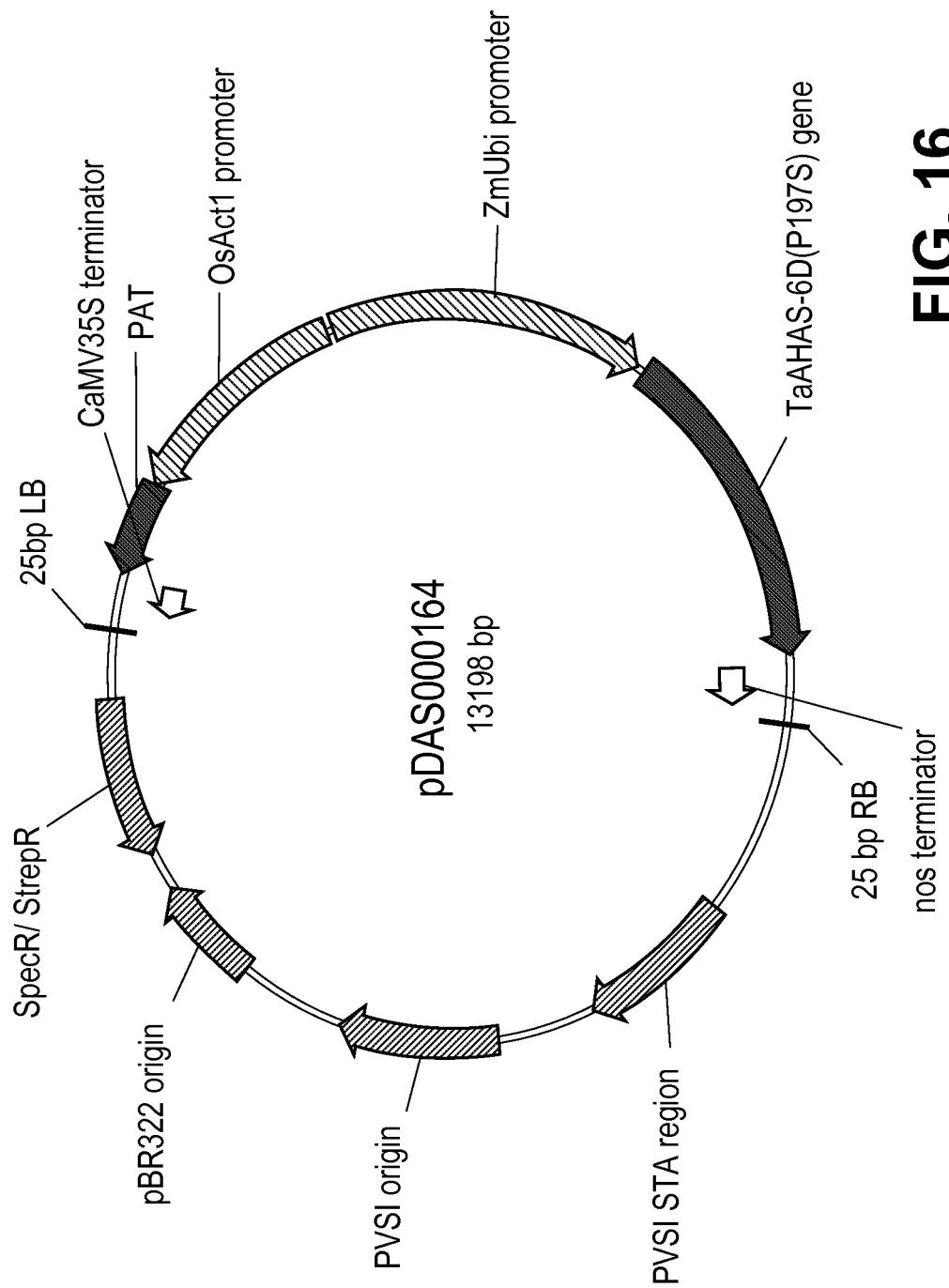
**FIG. 11**

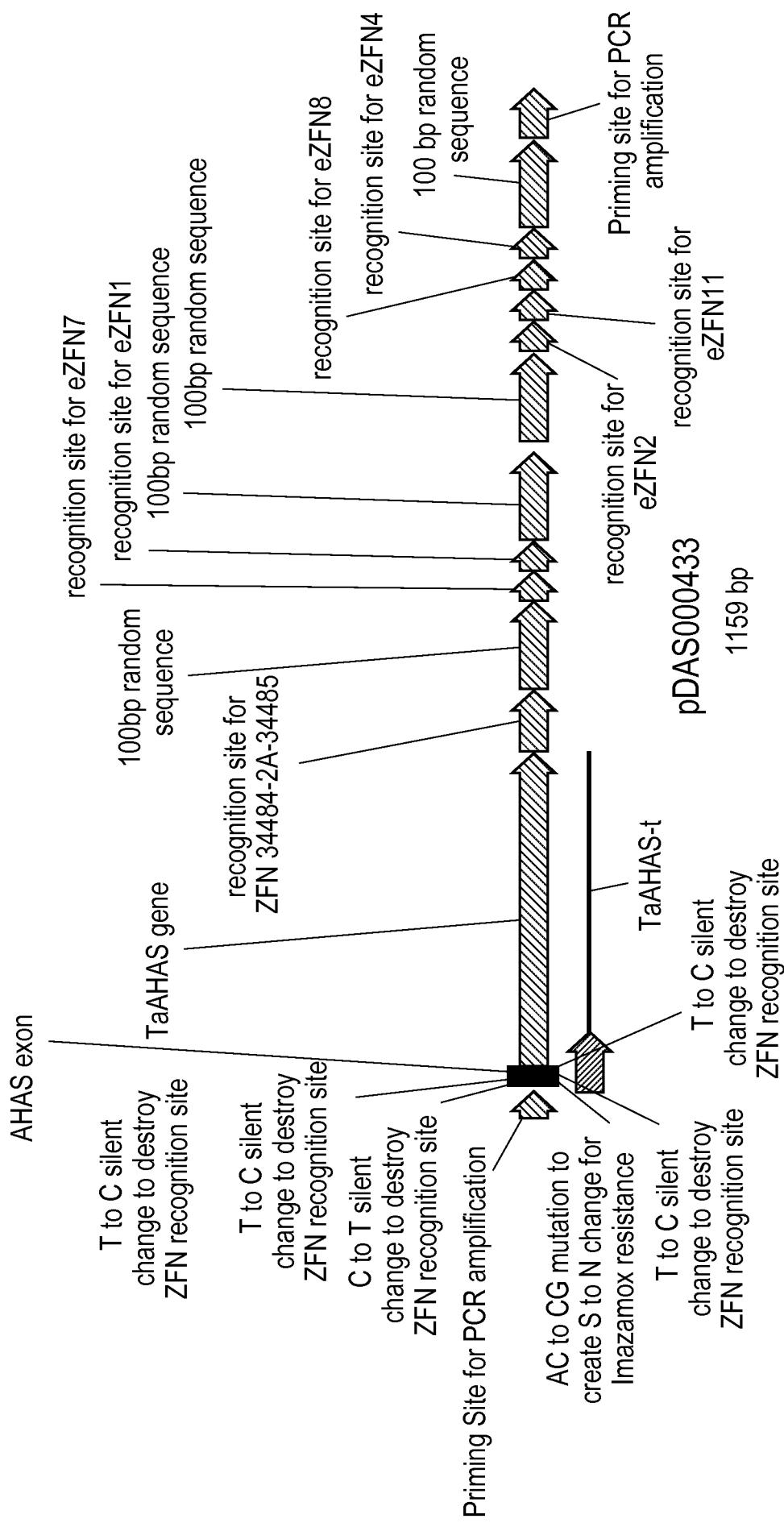
**FIG. 12**

**FIG. 13**

**FIG. 14**

**FIG. 15**

**FIG. 16**

**FIG. 17**

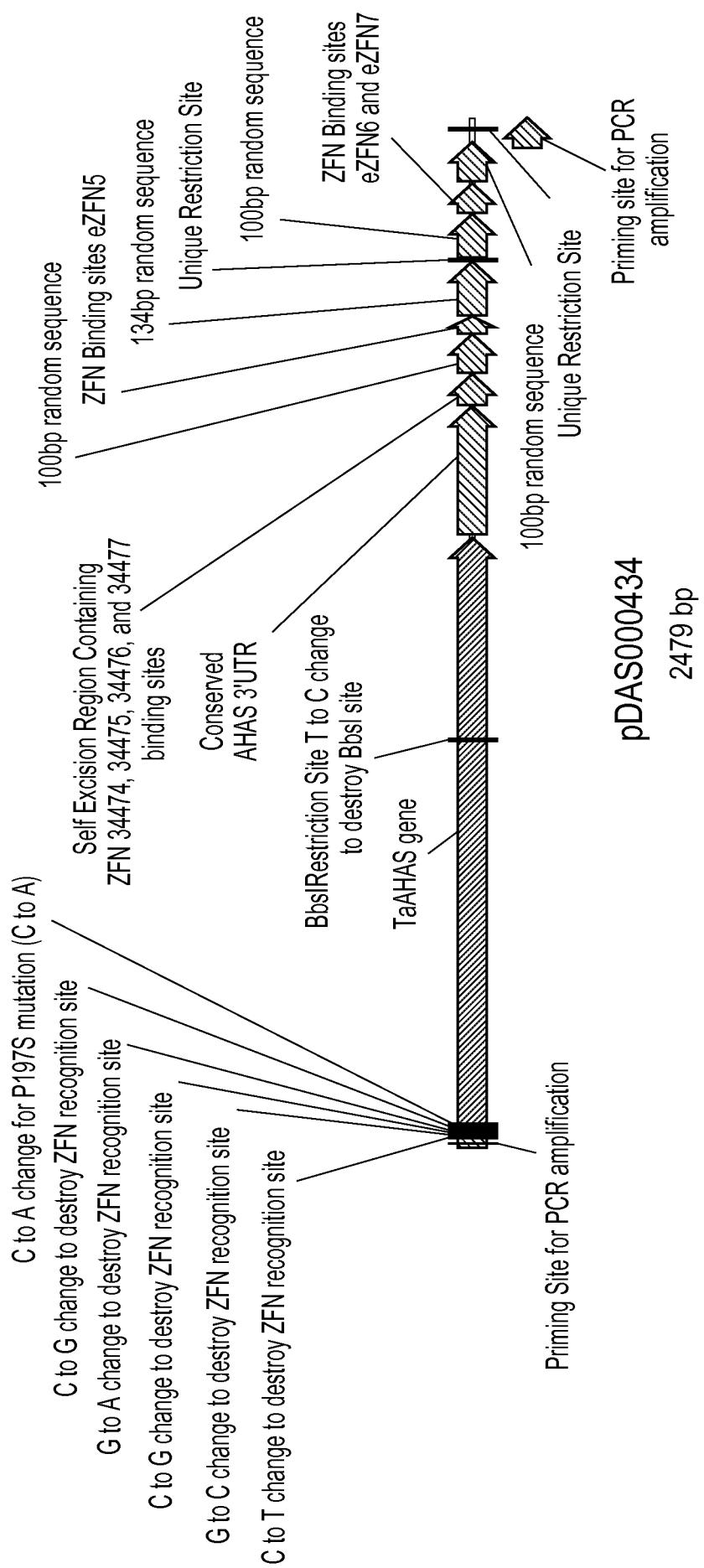
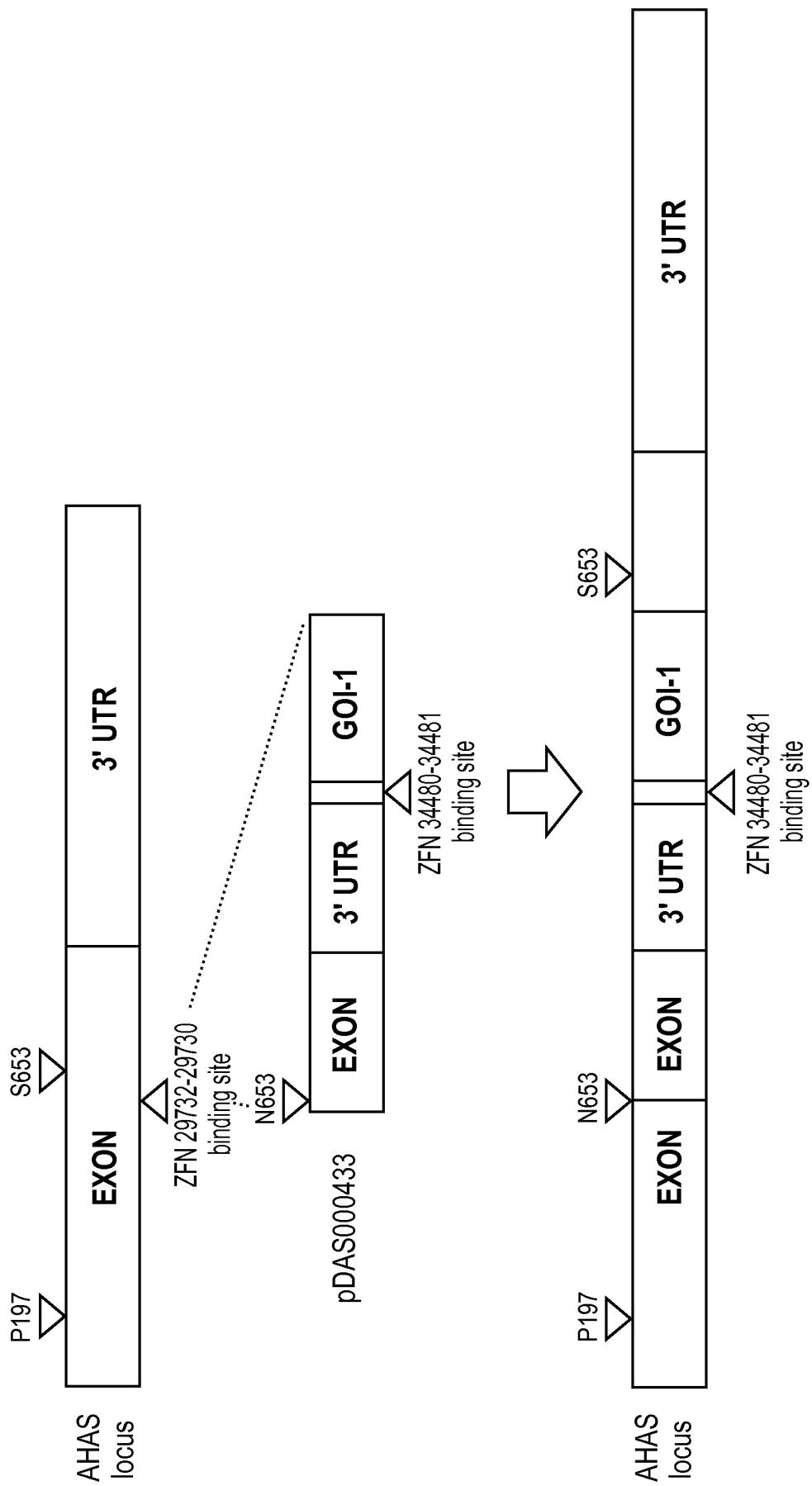
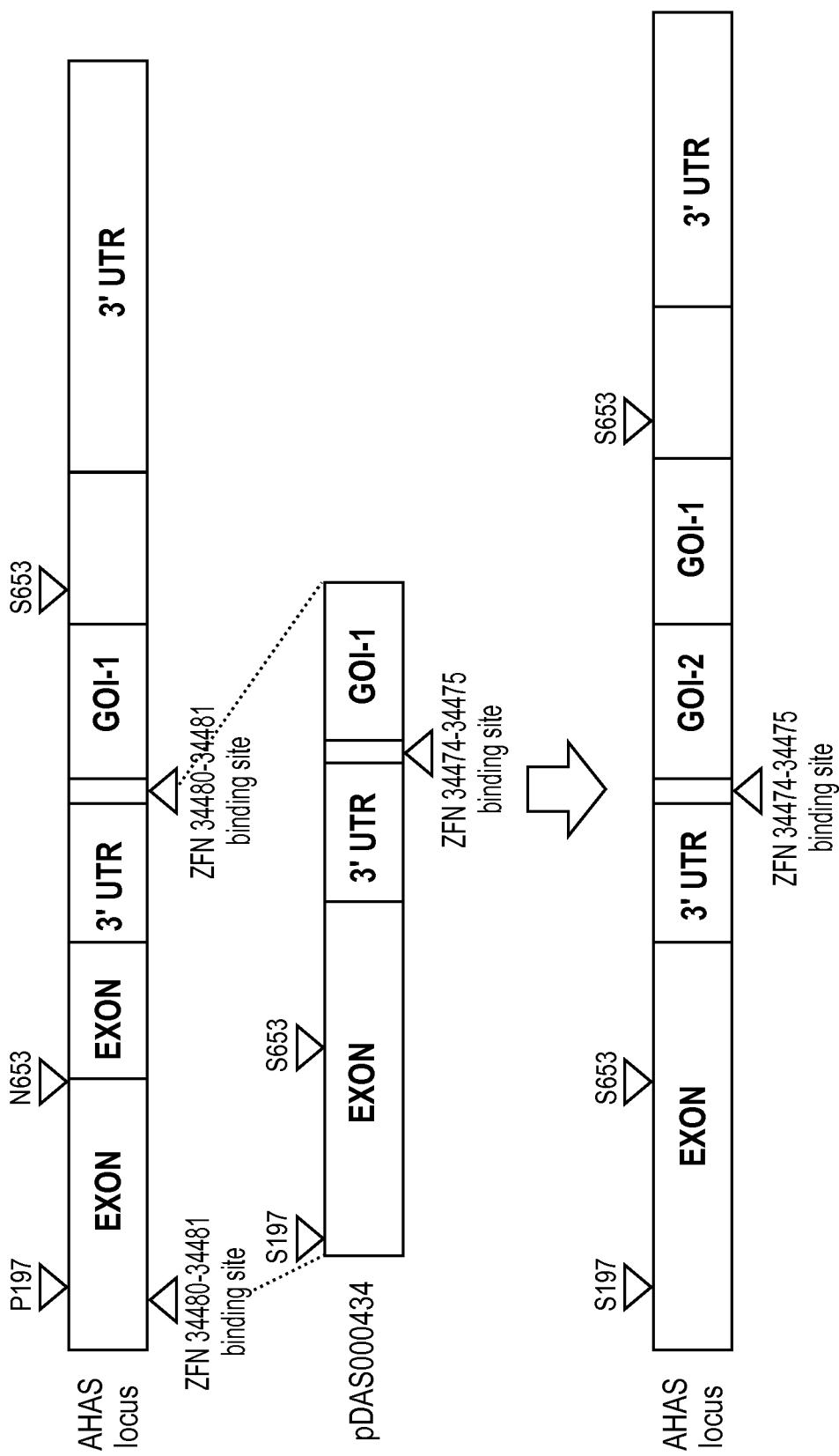
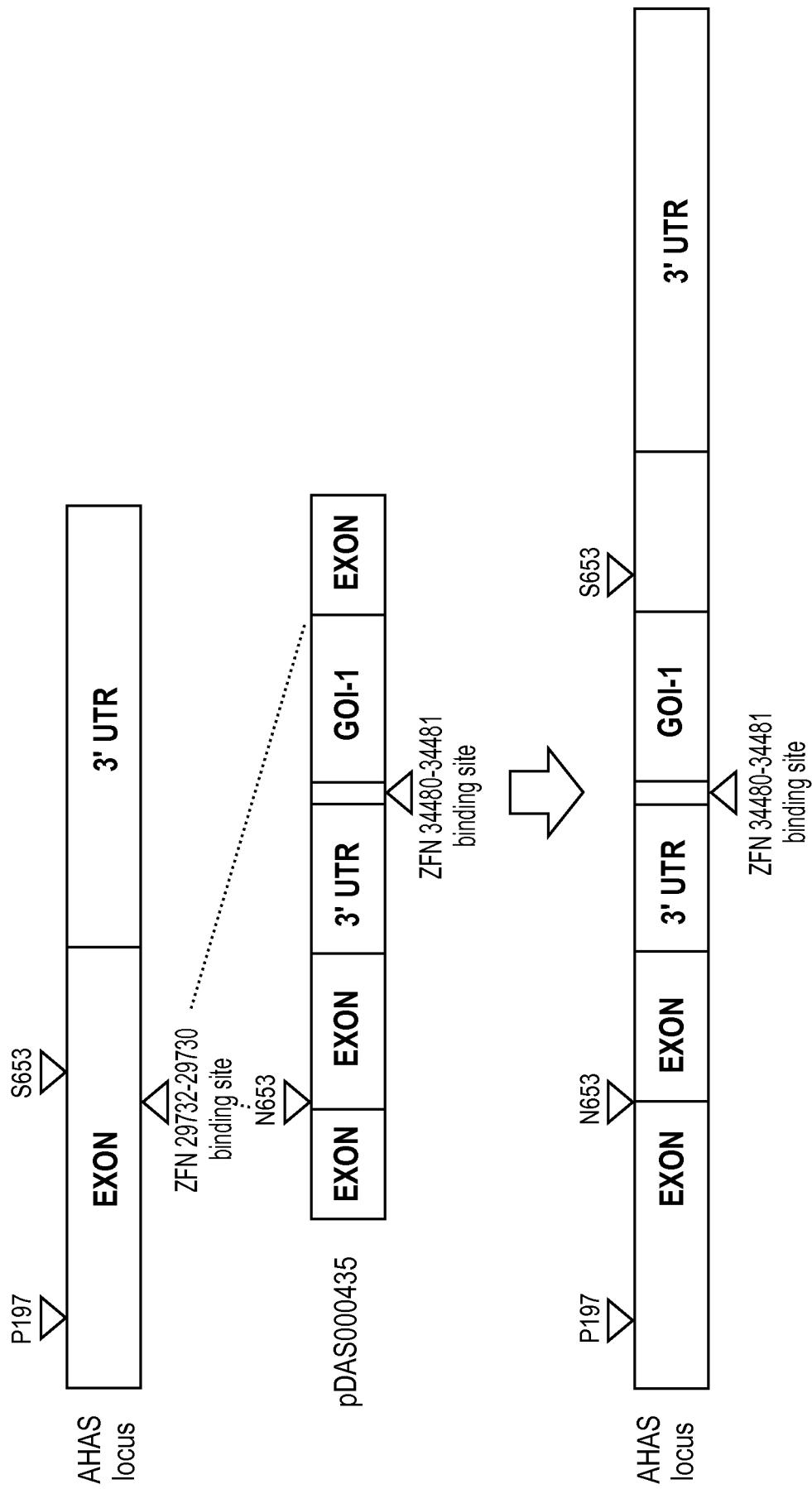
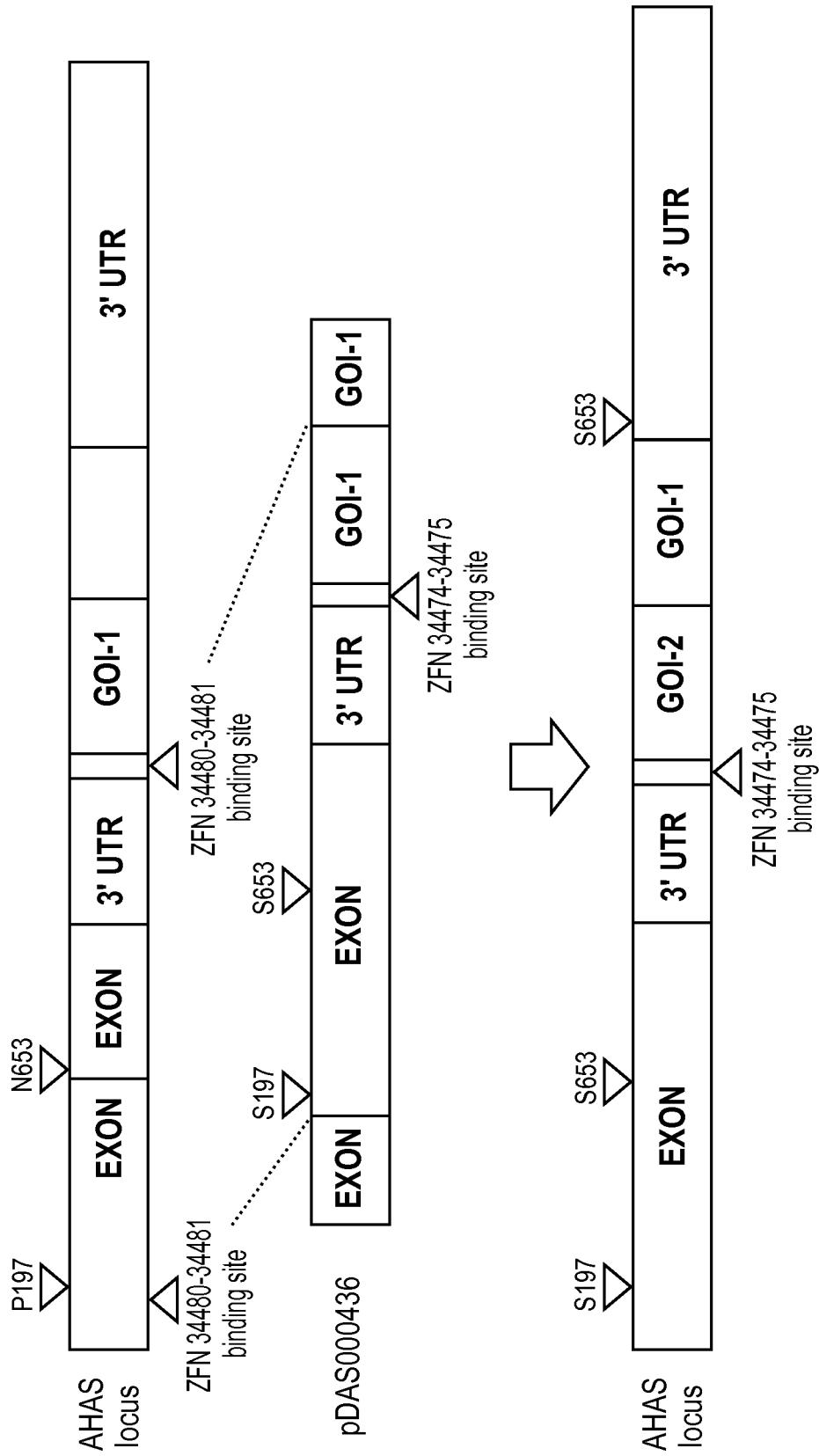


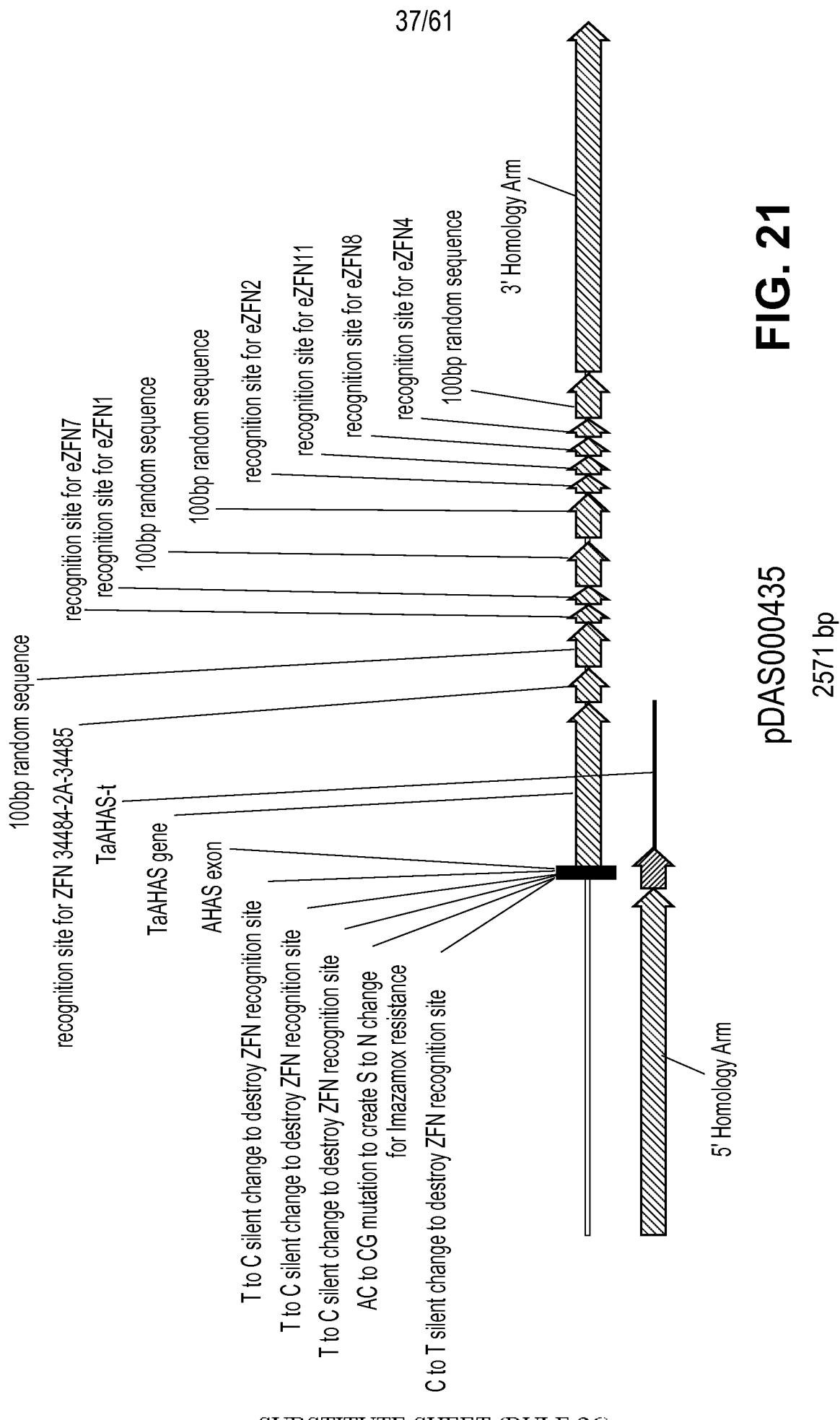
FIG. 18

**FIG. 19A**

**FIG. 19B**

**FIG. 20A**

**FIG. 20B**



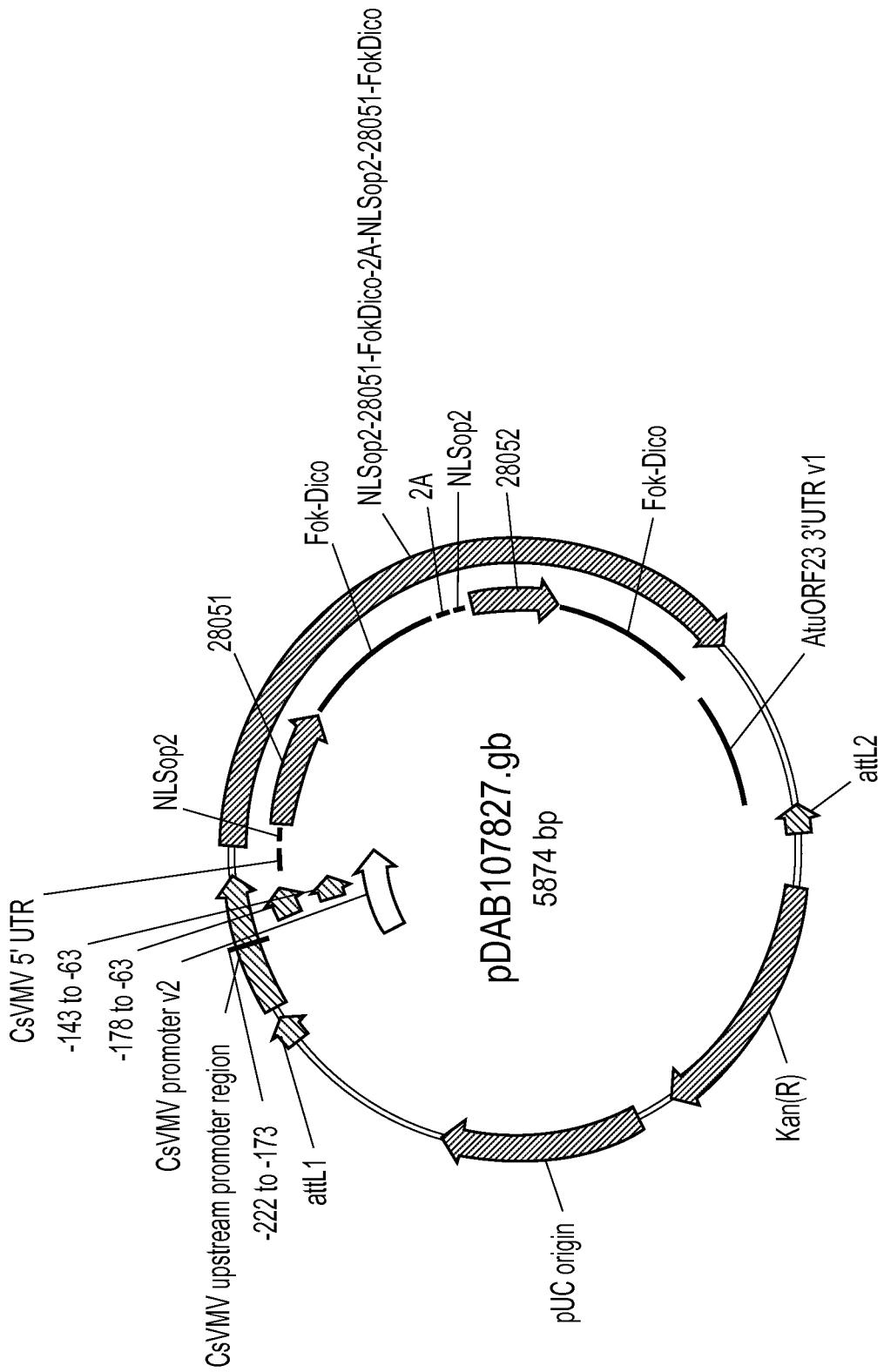
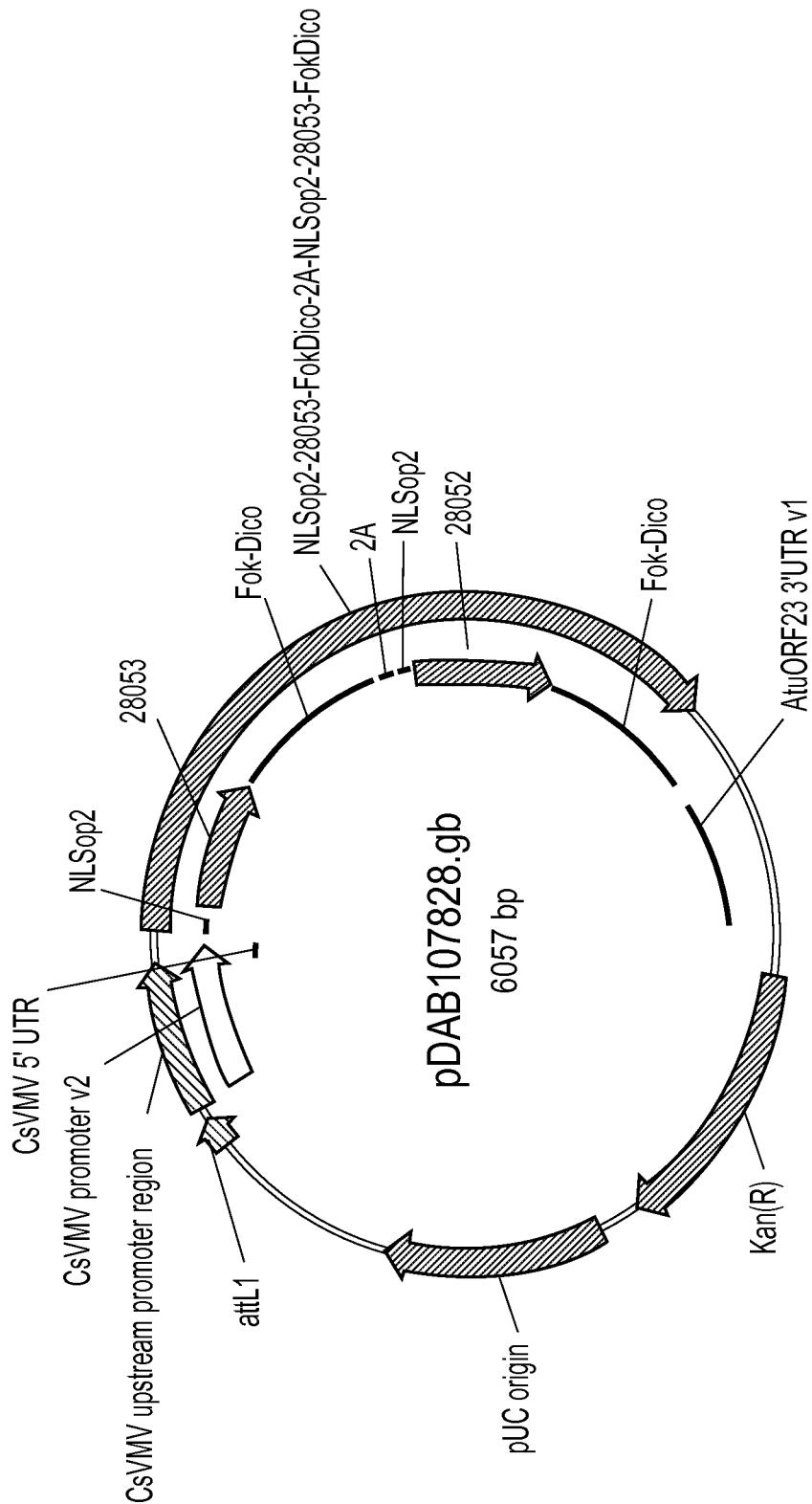
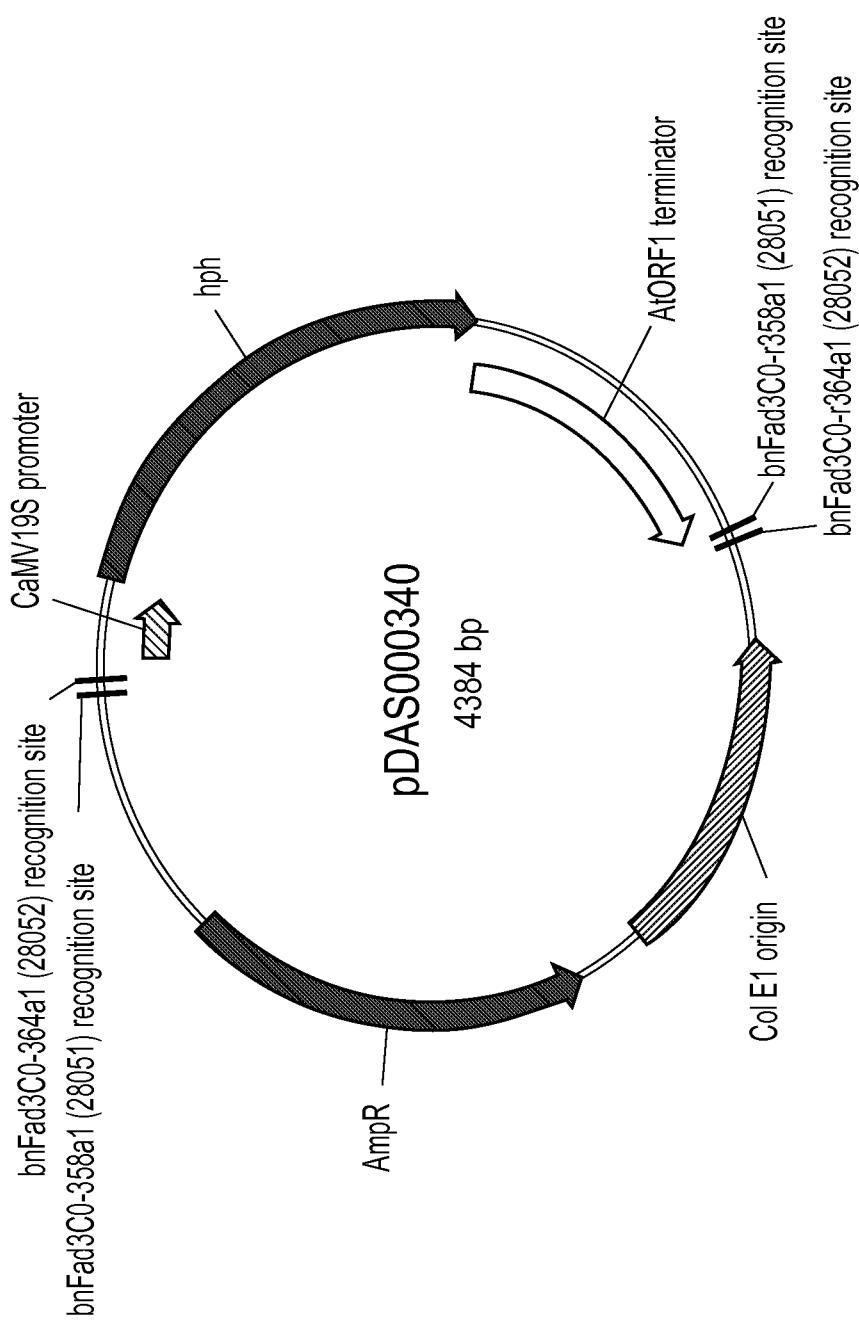
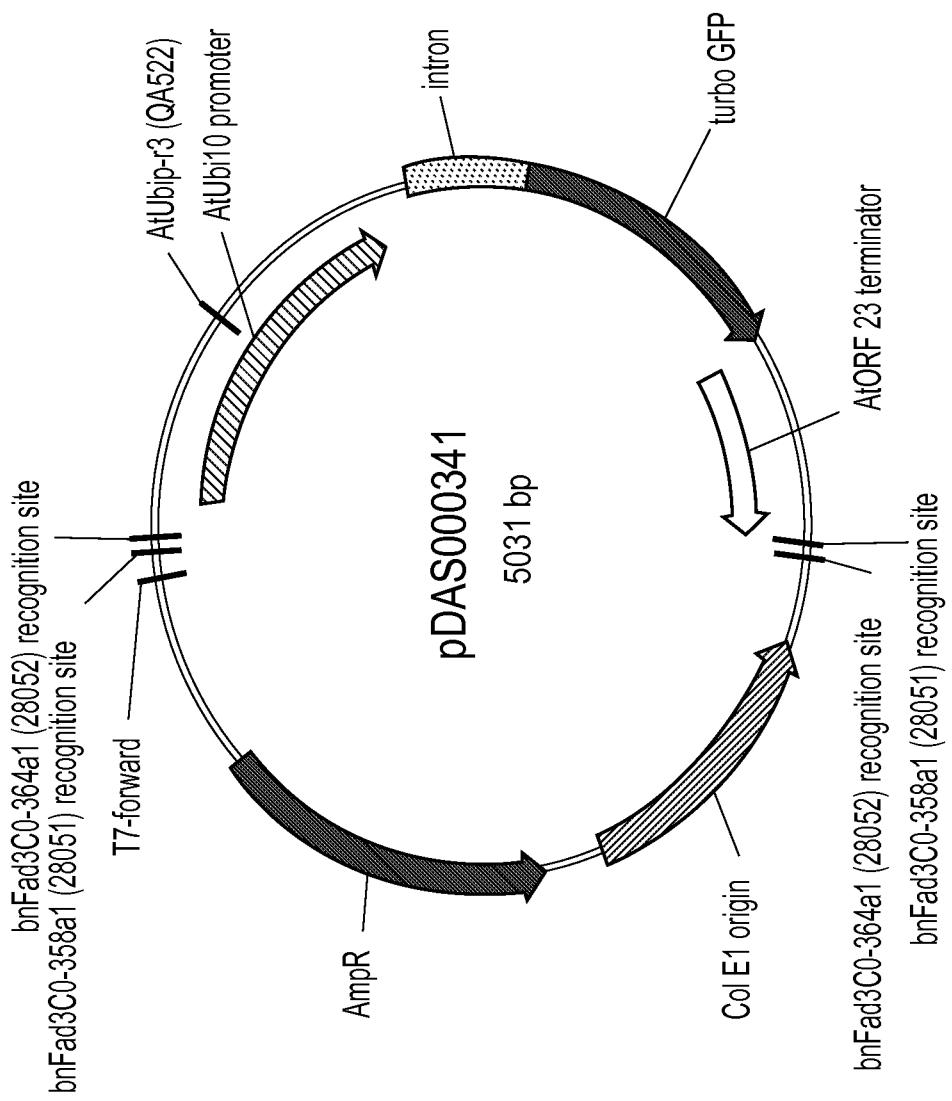
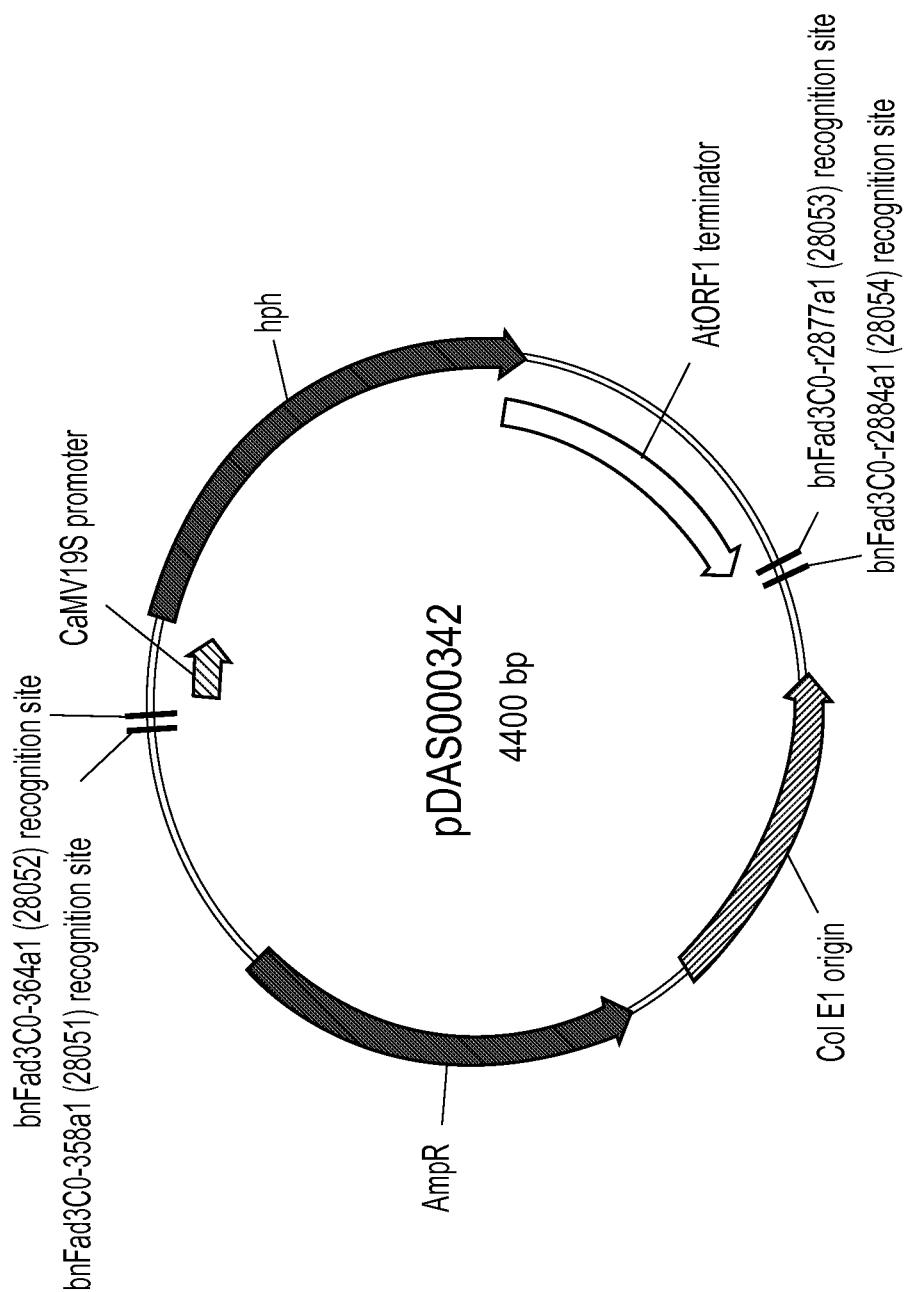


FIG. 22

**FIG. 23**

**FIG. 24**

**FIG. 25**

**FIG. 26**

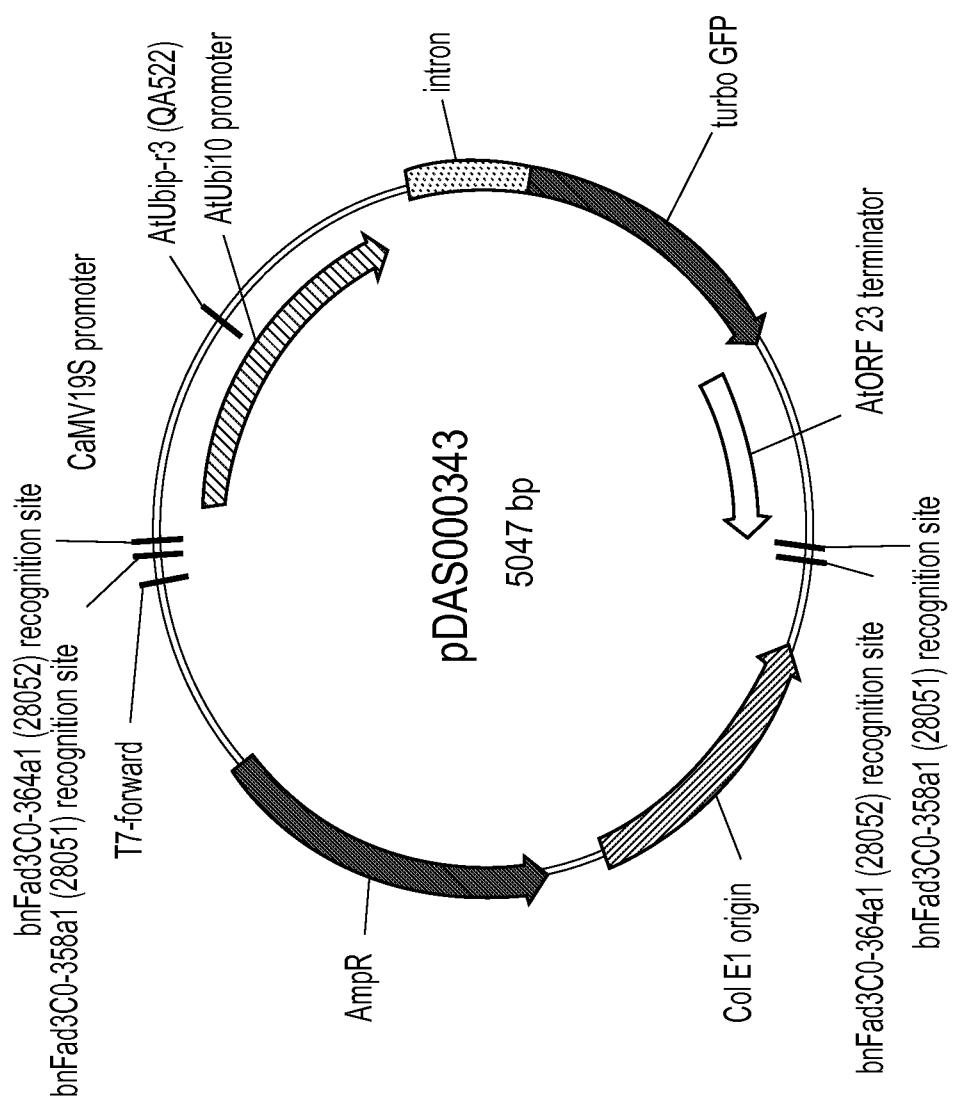
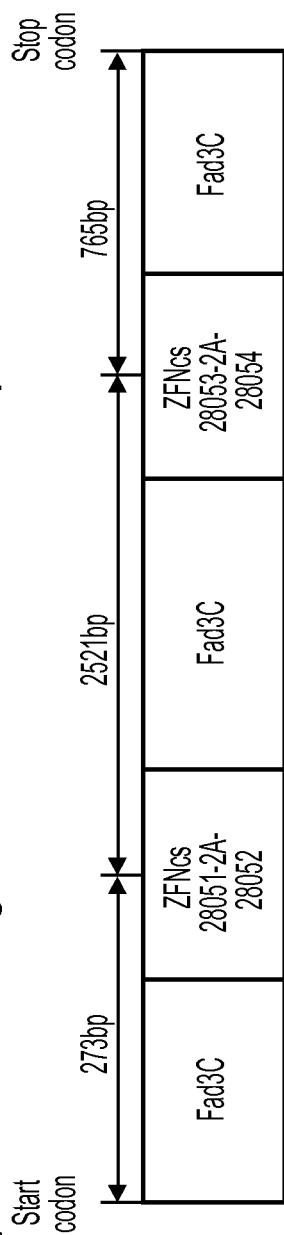
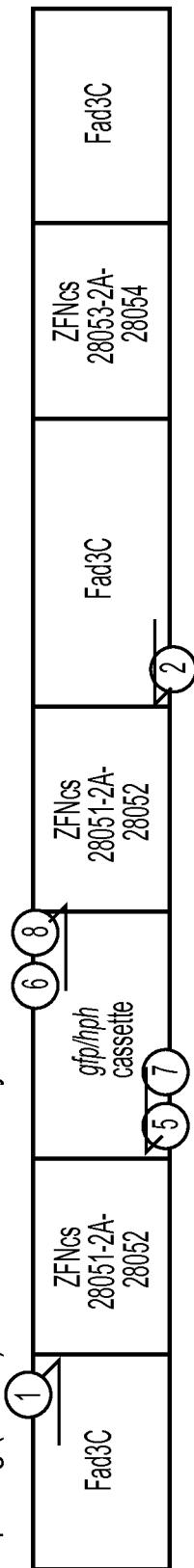
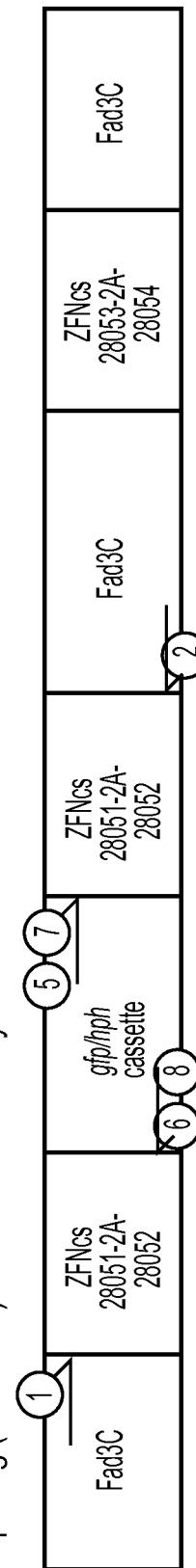
**FIG. 27**

FIG. 28**(A) Position of ZFN recognition sites relative to start and stop codons in FAD3C****(B) Position of primers (1-8 from Table 2) on spliced or edited FAD3C locus**

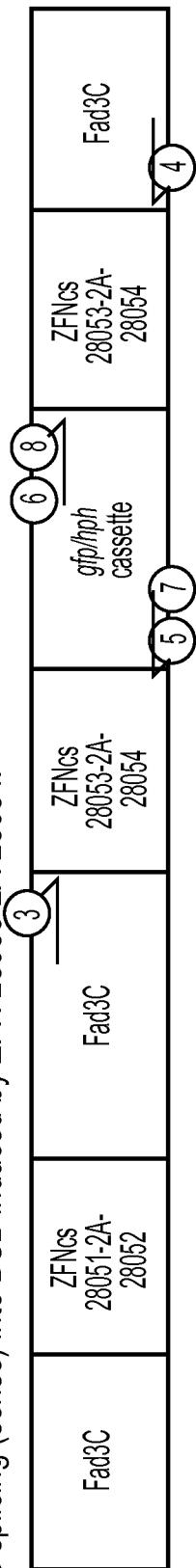
Gene splicing (sense) into DSB induced by ZFN 28051-2A-28052:



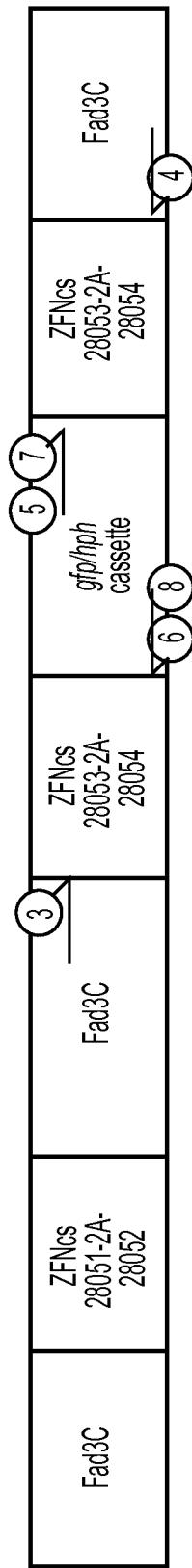
Gene splicing (antisense) into DSB induced by ZFN 28051-2A-28052:



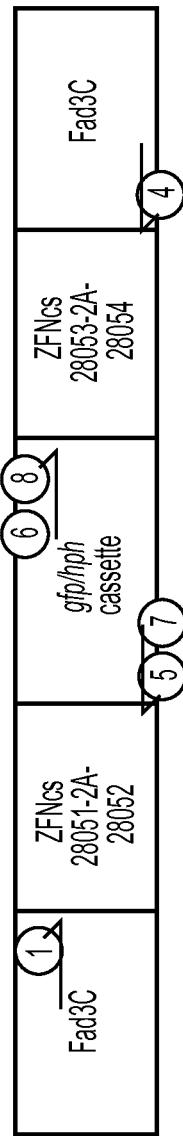
Gene splicing (sense) into DSB induced by ZFN 28053-2A-28054:



Gene splicing (antisense) into DSB induced by ZFN 28053-2A-28054:



Gene splicing (sense) into DSB induced by ZFN 28051-2A-28052 and 28053-2A-28054:



Gene splicing (antisense) into DSB induced by ZFN 28051-2A-28052 and 28053-2A-28054:

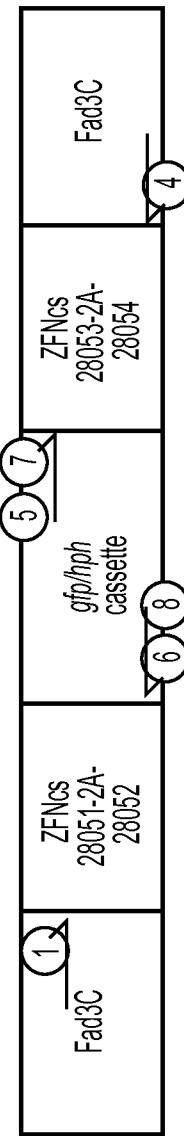


FIG. 28 (CONT.)

(A) Sequences amplified from the junction of the tGFP cassette from pDAS000341 with Fad3C at the DSB recognized by ZFN 28051-2A-28052

"." indicates deletions at cut-site

5' junction of tGFP cassette with FadC			
Fad3	ZFN recognition site 20051	Inserted Bases	ZFN recognition site 20052
TTCTGGCTCTTATTGGGCCAAAGGAACCTTTCTGGCCATC	TCGTACTCGGCCACGACTGGTATTAACTGGATCAGTAA		
TTCTGGCTCTTATTGGGCCAAAGGAACCTTTCTGGCCATC	TCGTACTCGGCCACGACTGGTATTAACTGGATCAGTAA		
TTCTGGCTCTTATTGGGCCAAAGGAACCTTTCTGGCCATC:	TCGTACTCGGCCACGACTGGTATTAACTGGATCAGTAA		

(B) Sequences amplified from the junction of the tGFP cassette from pDAS000343 with Fad3C at the DSBs recognized by ZFNs 28051-2A-28052 and 28053-2A-28054. ":" indicates deletions at cut-site

Fad3	ZFN recognition site 28052	Inserted Bases	ZFN recognition site 28052	AuB110p
<u>TAGTTTATTGCCCAAGGAGAAACCTTATTGAACTTCAAC</u>			<u>TCCGACTCGCCACGACTGGTAATTAAATGGATCCACTAGTAA</u>	
TAGTTTATTGCCCAAGGAGAAACCTTATTGAACTTCAACT			:::.....:::82 bases deleted:::.....:::	
TAGTTTATTGCCCAAGGAGAAACCTTATTGAACTTCA:::			:::.....:::68 bases deleted:::.....:::	
TAGTTTATTGCCCAAGGAGAAACCTTATTGAACTTCAACT	CG		TCTGACTCGCCACGACTGGTAATTAAATGGATCCACTAGTAA	
TAGTTTATTGCCCAAGGAGAAACCTTATTGAACTTCAACT			:::TGACTCGCCACGACTGGTAATTAAATGGATCCACTAGTAA	
TAGTTTATTGCCCAAGGAGAAACCTTATTGAACTTCA:::			:::TACTCGCCACGACTGGTAATTAAATGGATCCACTAGTAA	
TAGTTTATTGCCCAAGGAGAAACCTTATTGAACTTCA:::	AT		:::.....:::AGGTAATTAAATGGATCCACTAGTAA	
.....:::121 bases deleted:::.....:::.....:::		:::.....:::	

AltOrf23t	ZFN recognition site 28053	Inserted Basses	ZFN recognition site 20054	FaddC
<u>TCCAACTGGCCAGGAAAGCTTATGCCAACTTCAC</u>			<u>TACITGGTCACTGGTGGCCACTCTGTATCATCA</u>	
TCCAACTGGTGGCCAGGAAAGCTTATGCCAACTTCAC			:ATITGGTGGTCACTGGTGGCCACTCTGTATCATCA	
TCCAACTGGTGGCCAGGAAAGCTTATGCCAACTTCAC			TACITGGTGGTCACTGGTGGCCACTCTGTATCATCA	
TCCAACTGGTGGCCAGGAAAGCTTATGCCAACTTCAC			TACITGGTGGTCACTGGTGGCCACTCTGTATCATCA	
		CTTC	GATAAAAGTGGCTGCCTGTTGGGTGTTGATGTT	
				TACAC

(A)

Sample	Fac3	ZFN recognition site 28051	# of ExtraBases Inserted	ZFN recognition site 28052	# of ExtraBases Inserted	CaMV 9sp
1:1	#5	TTCTGGCCTCTTTATTGGCGCCAAAGGAACCCCTTTCTGGCCATCT				TCGTAATCGGCCACCGACTGGTAATTAAATGGATCCAAACGGACAACCACT
1:1	#39	TTCTGGCCTCTTTATTGGCGCCAAAGGAACCCCTTTCTGG:::.....	3			TCGTAATCGGCCACCGACTGGTAATTAAATGGATCCAAACGGACAACCACT
1:1	#46	TTCTGGCCTCTTTATTGGCGCCAAAGGAACCCCTTTCTGG:::.....	153		:::13bp missing:::.....:::.....
5:1	#63	TTCTGGCCTCTTTATTGGCGCCAAAGGAACCCCTTTCTGG:::.....		370	:::CGGCACAGACTGGTAATTAAATGGATCCAAACGGACAACCACT
10:1	#16	TTCTGGCCTCTTTATTGGCGCCAAAGGAACCCCTTTCTGGCCATCT		36		TCGTAATCGGCCACCGACTGGTAATTAAATGGATCCAAACGGACAACCACT
10:1	#64	TTCTGGCCTCTTTATTGGCGCCAAAGGAACCCCTTTCTGGCCATCT		78	:::52bp missing:::.....:::.....
10:1	#66:::51bp missing:::.....:::.....			:::254bp missing:::.....:::.....

(B)

Sample	AluORF1-t-	ZFN recognition site 28051	# of ExtraBases Inserted	ZFN recognition site 28052	# of ExtraBases Inserted	Fad3C
1:1	#5	GTAATACATAGGGCGGCCAAAGGAACCCCTTTCTGGCCATCT		TCGTAATCGGCCACCGACTGGTAATTAAATGGATCCAACTTCAACTCTTA		
1:1	#39	GTAATACATAGGGCGGCCAAAGGAACCCCTTTCTGGCCATCT	4:::GGCACCGACTGGTAATTAAATGGATCCAACTTCAACTCTTA		
1:1	#46	GTAATACATAGGGCGGCCAAAGGAACCCCTTTCTGGCCATCT	112:::GACTGGTAATTAAATGGATCCAACTTCAACTCTTA		
5:1	#63	GTAATACATAGGGCGGCCAAAGGAACCCCTTTCTGGCCATCT	112:::GACTGGTAATTAAATGGATCCAACTTCAACTCTTA		
10:1	#16:::1655bp missing (possible double insertion):::.....	234:::GTAATCGGCCACCGACTGGTAATTAAATGGATCCAACTTCAACTCTTA		
10:1	#64	GTAATACATAGGGCGGCCAAAGGAACCCCTTTACTGGCCA:::.....	9:::GTAATCGGCCACCGACTGGTAATTAAATGGATCCAACTTCAACTCTTA		
10:1	#66	TGTAATACATAGGGCGGCCAAAGGAACCCCTTTACTGGCCA:::.....	7:::GTAATACATAGGGCGGCCAAAGGAACCCCTTTACTGGCCA:::.....		

FIG. 30

49/61

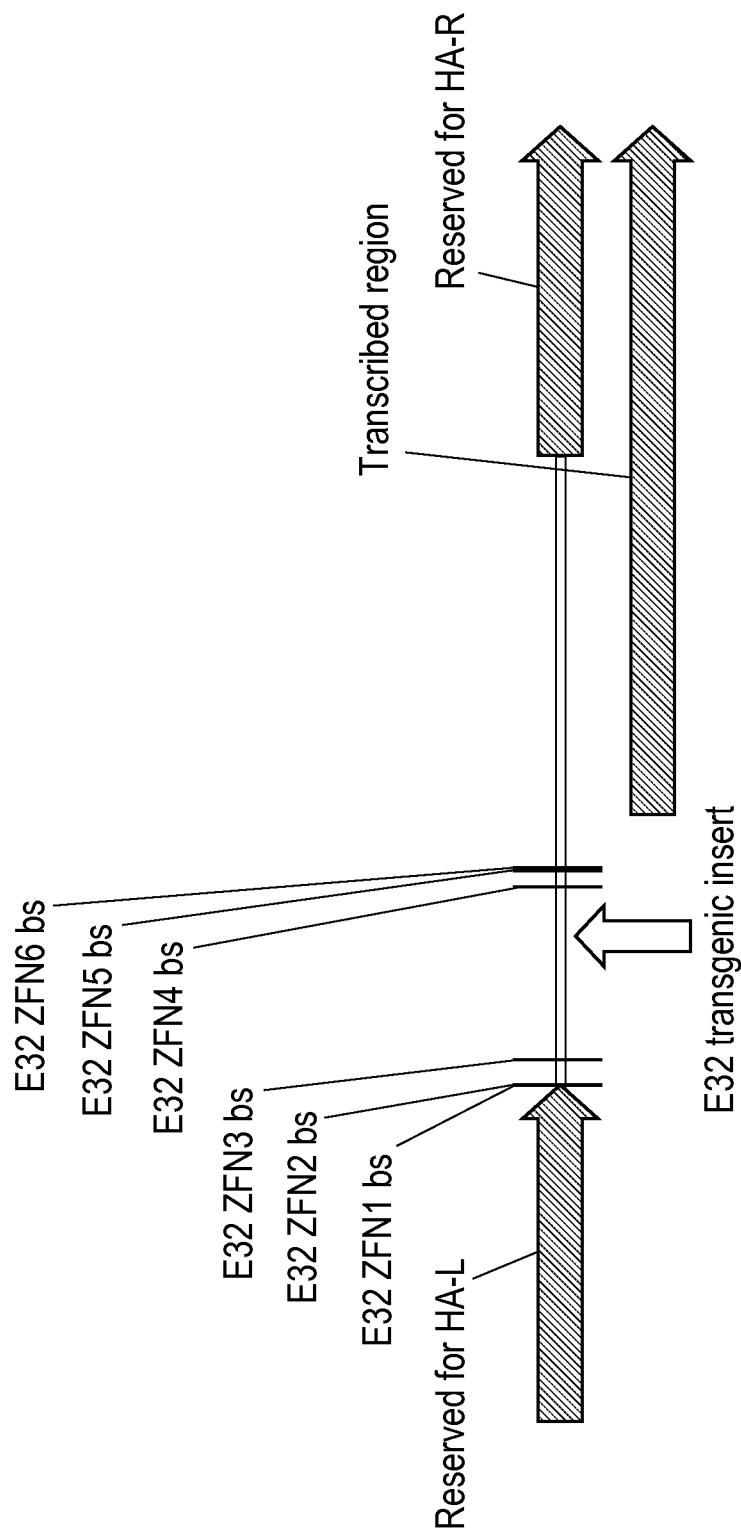
(A)

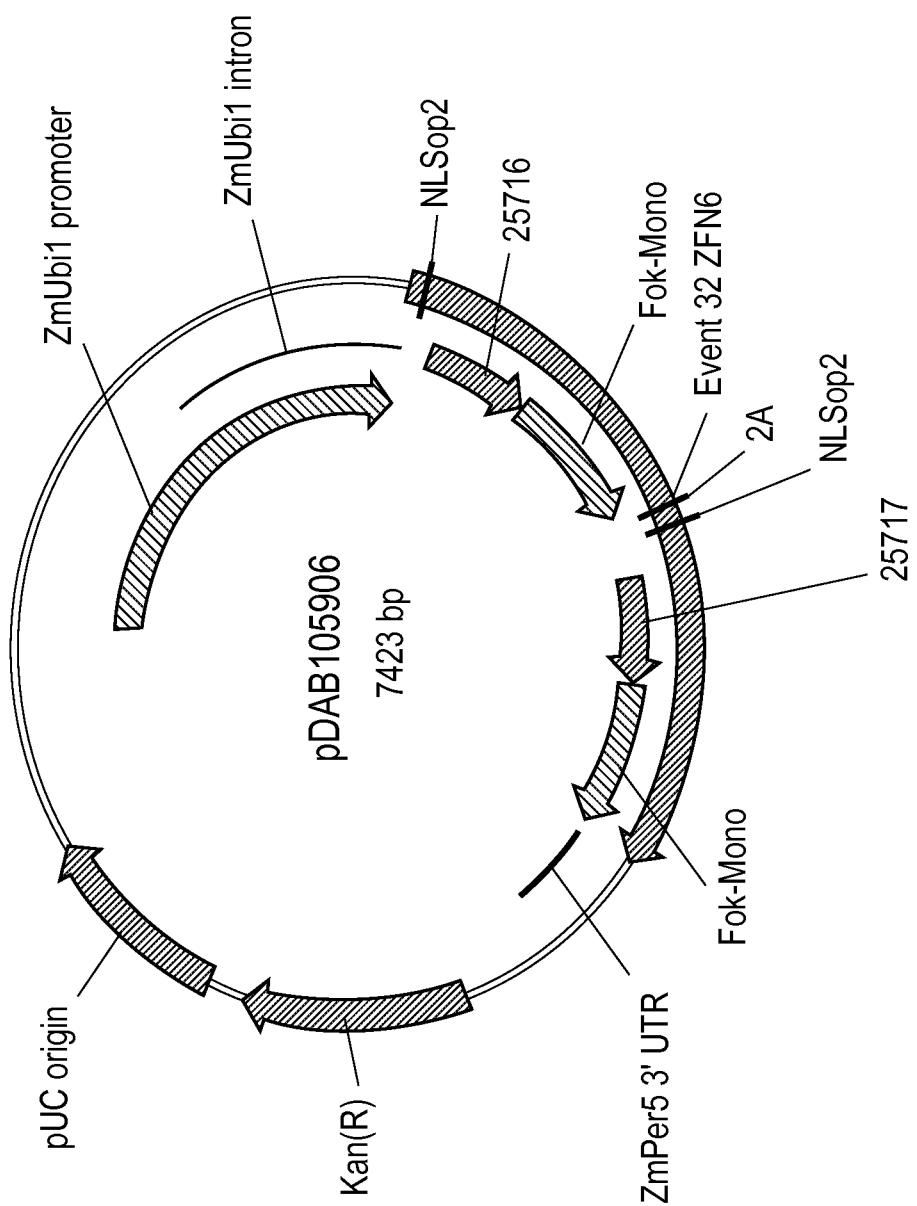
Sample	Fad3	ZFN recognition site 28051	# of ExtraBases Inserted	ZFN recognition site 28052	# of ExtraBases Inserted	ZFN recognition site 28054	# of ExtraBases Inserted	CamV19sp
		TTCTGGCCTCTTTATTGGGCCCAAGGAACCCTTTTCTGGCCATCT				TCGTACTGGCCACGACTGGTAATTAAATGGATCCAAACGGACAACCACTT		
10:1:1 #8		TTCTGGCCTCTTTATTGGGCCCAAGGAACCCTTTTCTGGCCATCT	206			TCGTACTGGCCACGACTGGTAATTAAATGGATCCAAACGGACAACCACTT		
10:1:1 #9		TTCTGGCCTCTTTATTGGGCCCAAGGAACCCTTTTCTGGCCATCT	273		CGGGCACGACTGGTAATTAAATGGATCCAAACGGACAACCACTT		
10:1:1 #21		TTCTGGCCTCTTTATTGGGCCCAAGGAACCCTTTTCTGGCCATCT	5		TCGGCACGACTGGTAATTAAATGGATCCAAACGGACAACCACTT		
10:1:1 #37		TTCTGGCCTCTTTATTGGGCCCAAGGAACCCTTTTCTGGCCATCT						

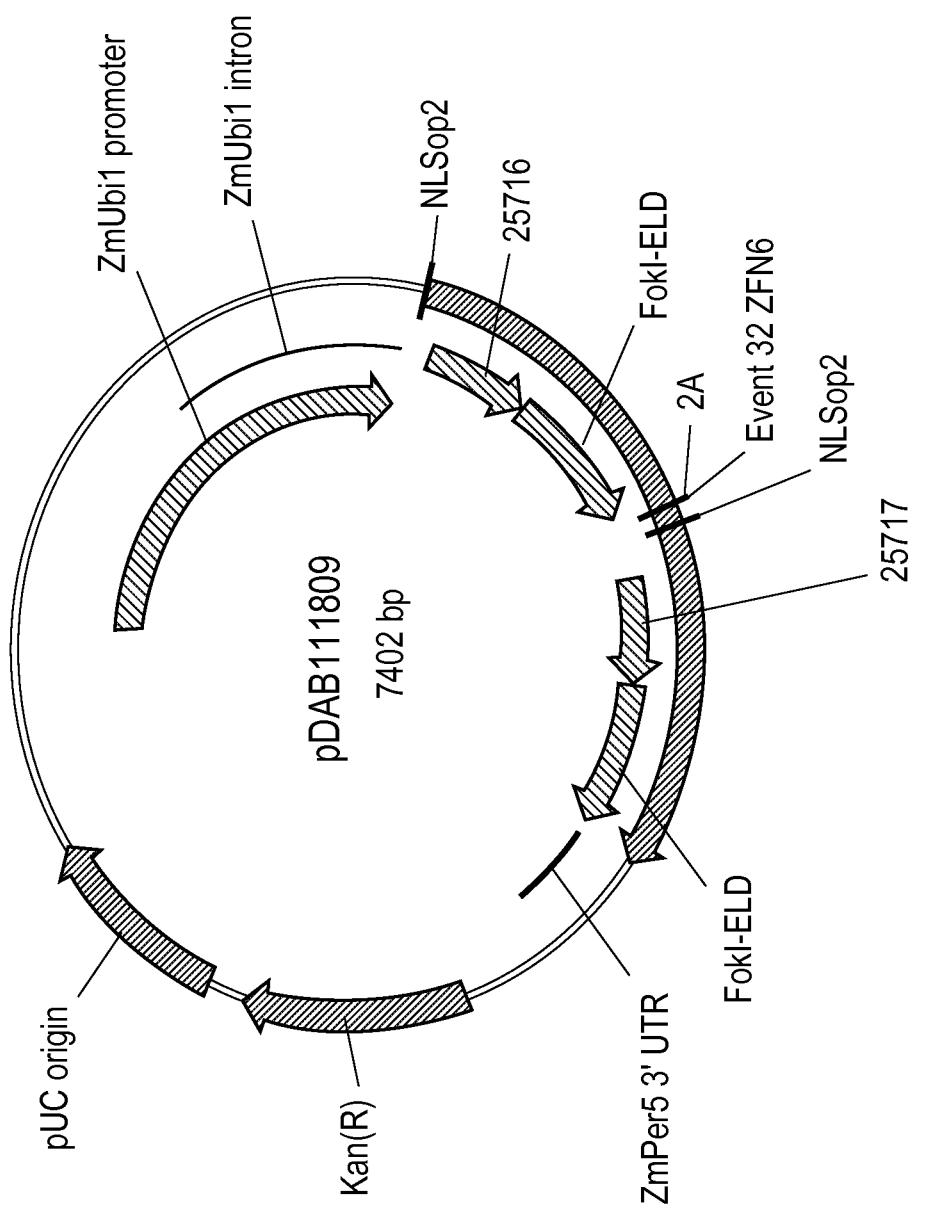
(B)

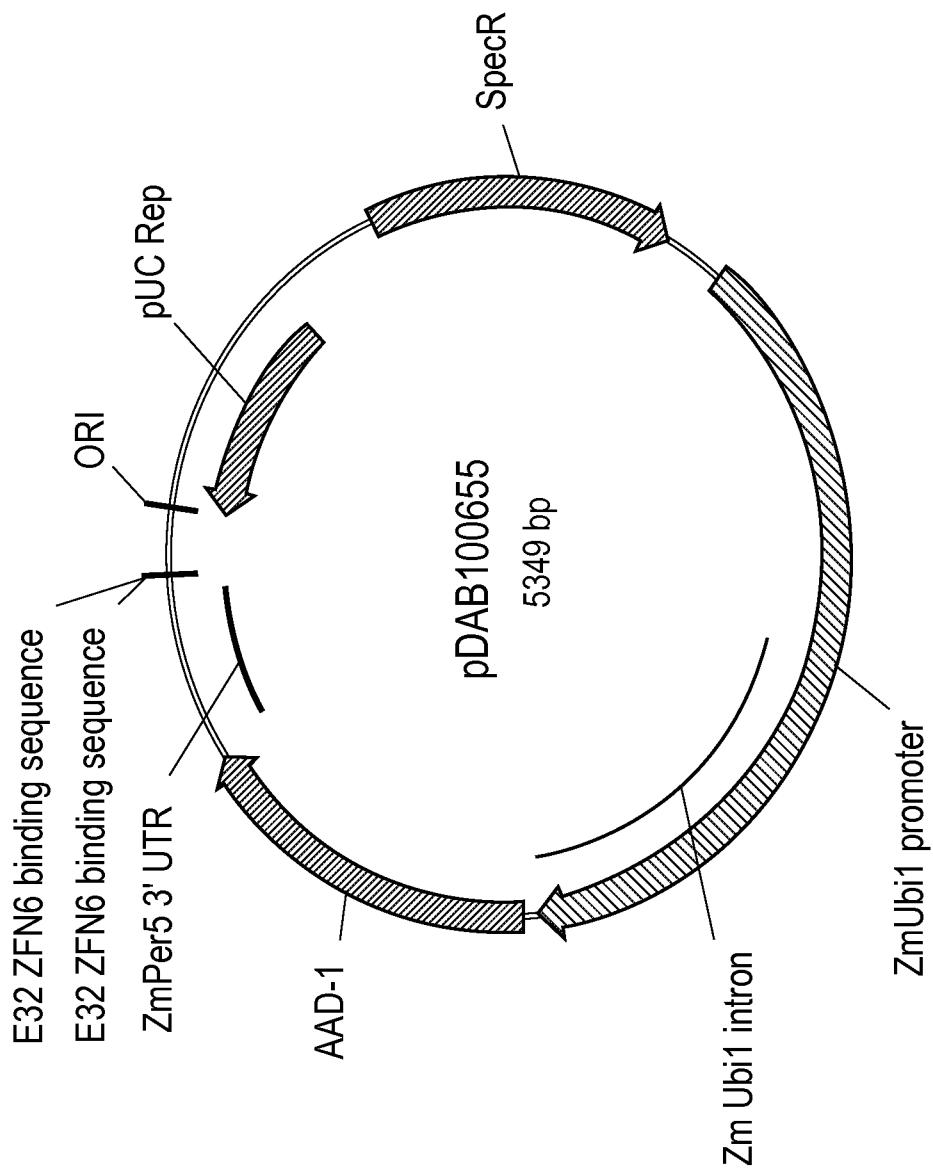
Sample	AwuORF1-	ZFN recognition site 28053	# of ExtraBases Inserted	ZFN recognition site 28054	# of ExtraBases Inserted	ZFN recognition site 28054	# of ExtraBases Inserted	Fad3C
		GTAATACATAGGGCGGAGGGAGAAAGCTTATGCAACTAAC				TACTGGCTGGCTGGATCGTGTGGCACTCTGGTTATCTATCATTCTGGTC		
5:1:1	#8	GTAATACATAGGGCGGAGGGAGAAAGCTTATGCAACTAAC	229		CTTGGCTGGCTGGATCGTGTGGCACTCTGGTTATCTATCATTCTGGTC		
10:1:1 #8		GTAATACATAGGGCGGAGGGAGAAAGCTTATGCAACTAAC	26			TACTGGCTGGCTGGATCGTGTGGCACTCTGGTTATCTATCATTCTGGTC		
10:1:1 #9		GTAATACATAGGGCGGAGGGAGAAAGCTTATGCAACTAAC	33		ACTGGCTGGCTGGATCGTGTGGCACTCTGGTTATCTATCATTCTGGTC		
10:1:1 #21		GTAATACATAGGGCGGAGGGAGAAAGCTTATGCAACTAAC	17		CTTGGCTGGCTGGATCGTGTGGCACTCTGGTTATCTATCATTCTGGTC		
10:1:1 #37		GTAATACATAGGGCGGAGGGAGAAAGCTTATGCAACTAAC						

FIG. 31

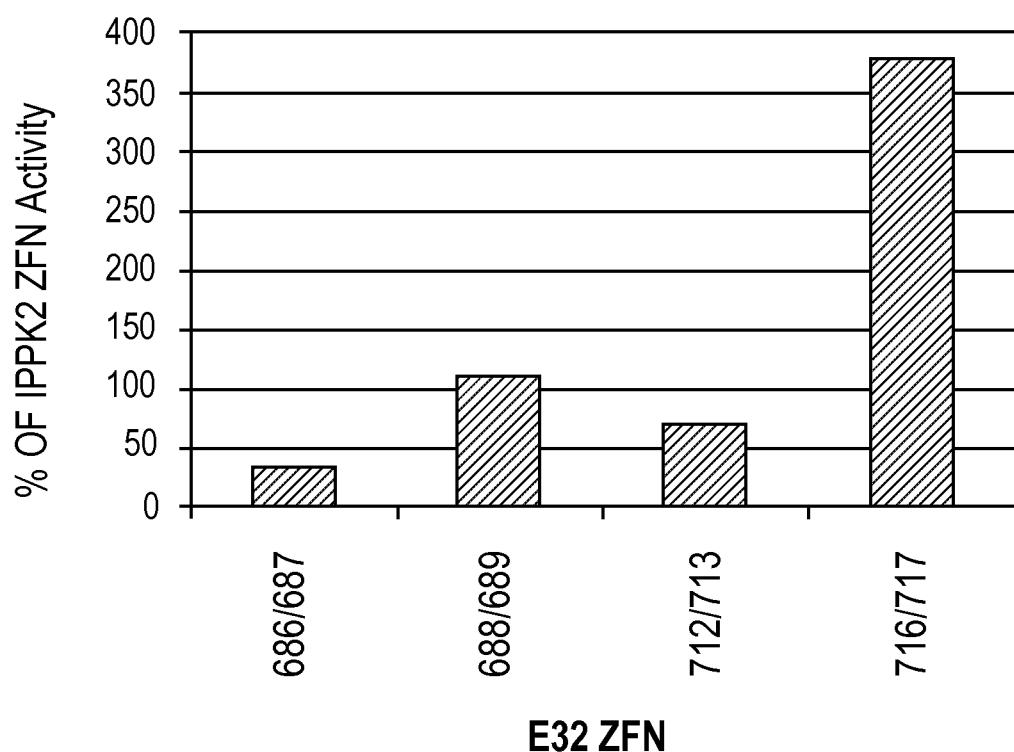
**FIG. 32**

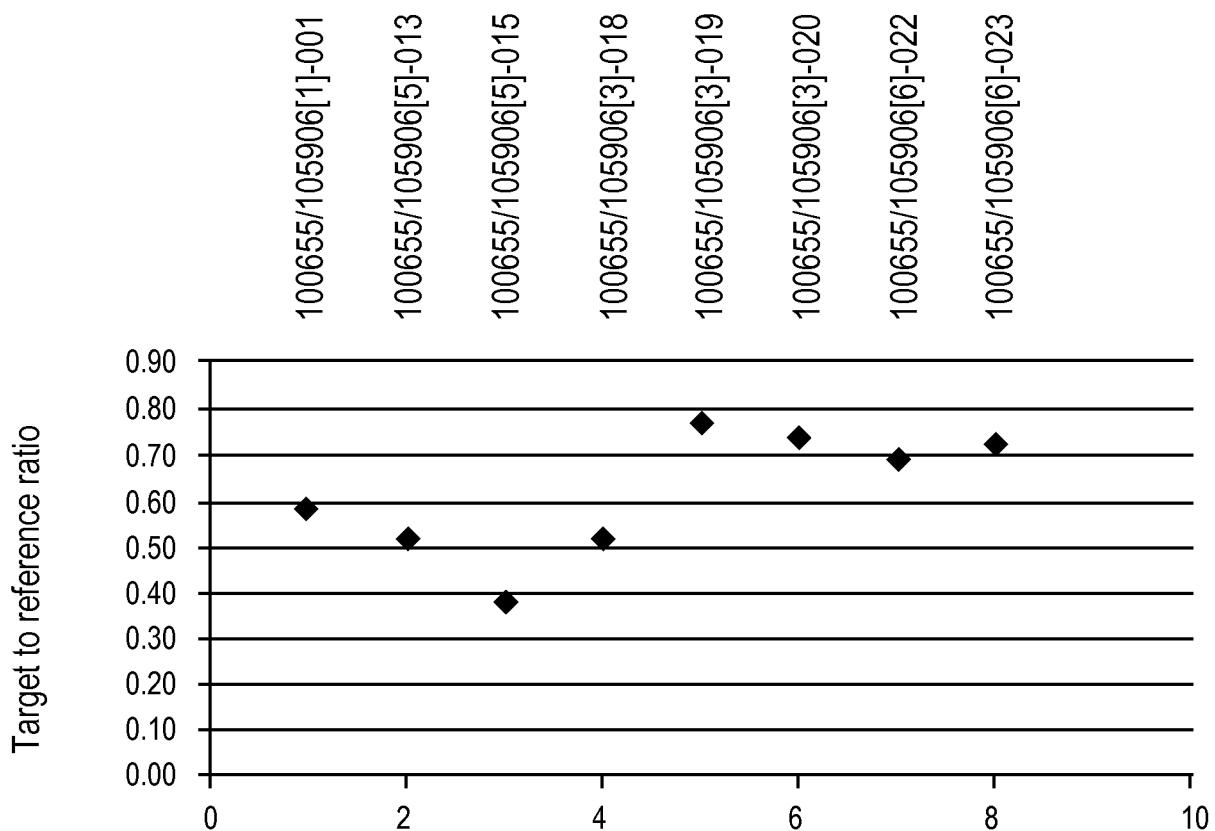
**FIG. 33**

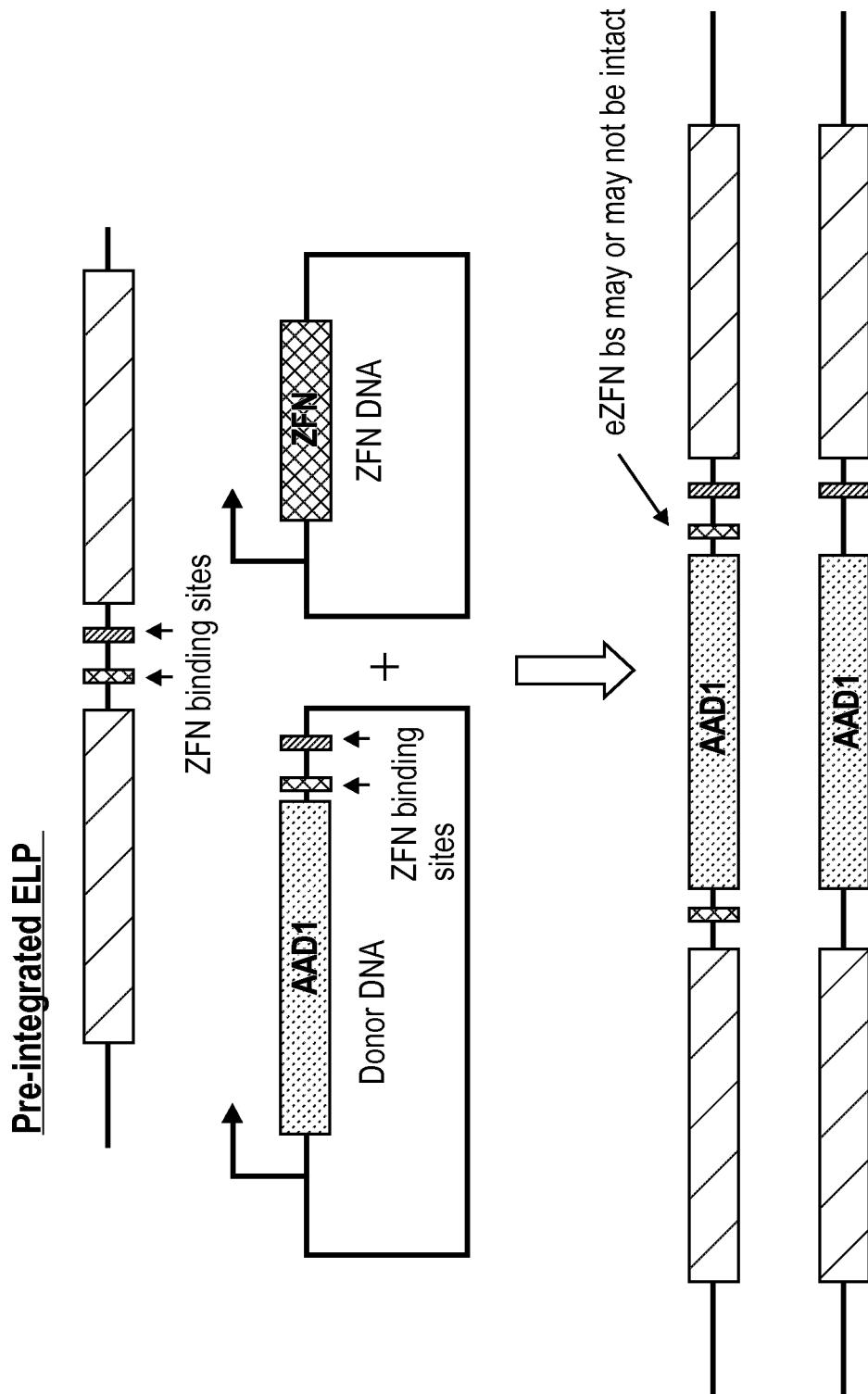
**FIG. 34**

**FIG. 35**

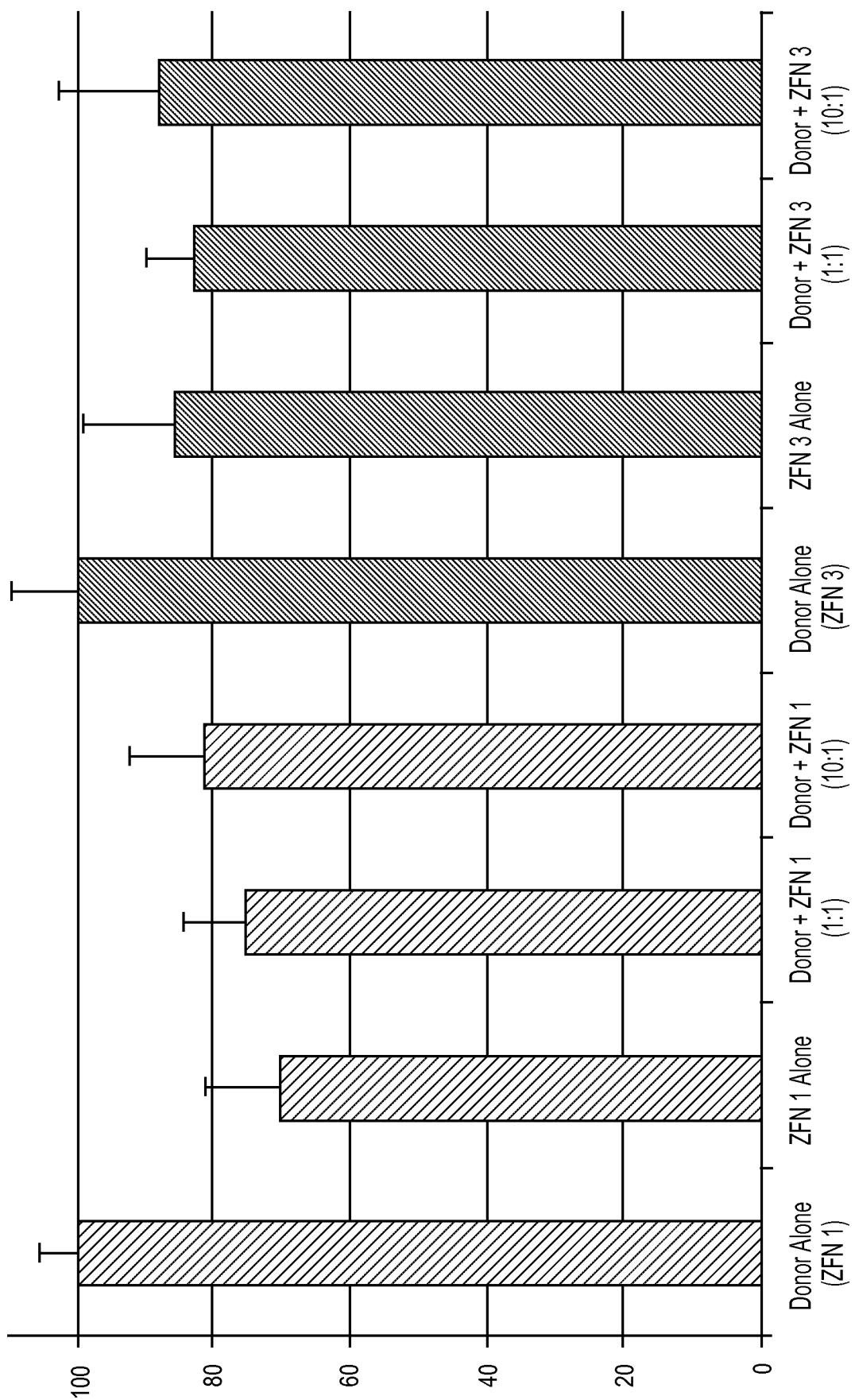
54/61

**FIG. 36**

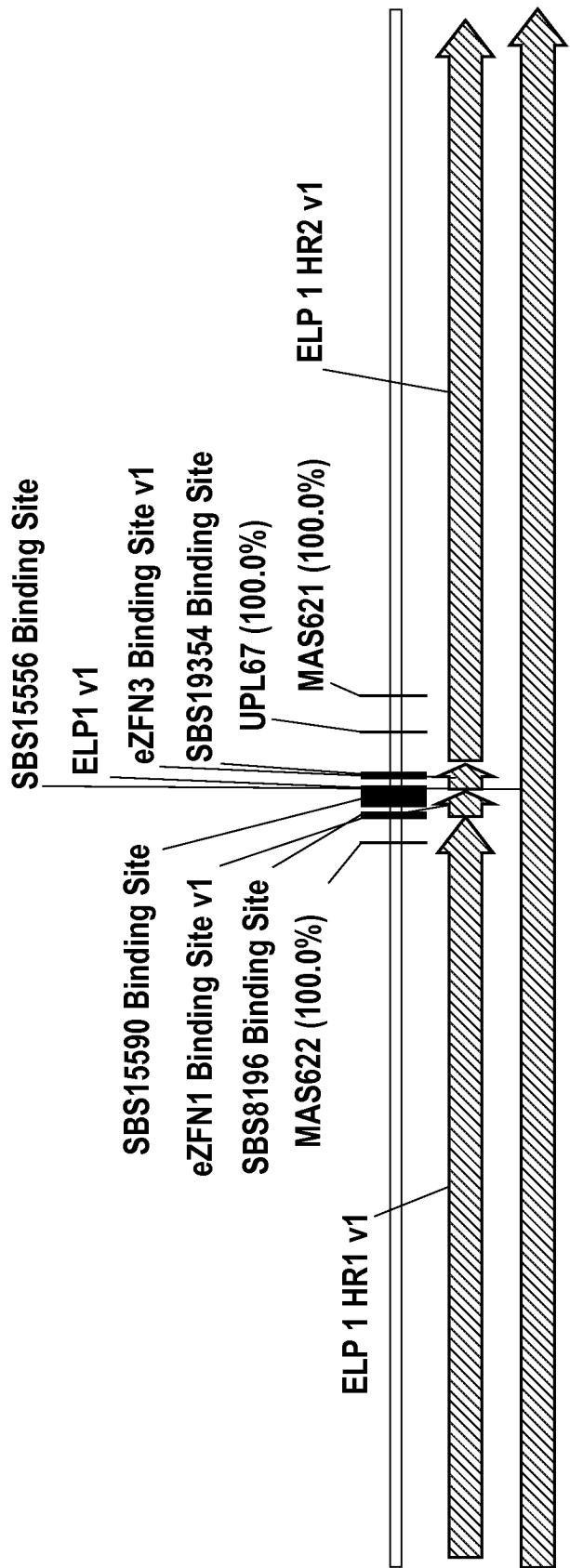
**FIG. 37**

**FIG. 38**

57/61

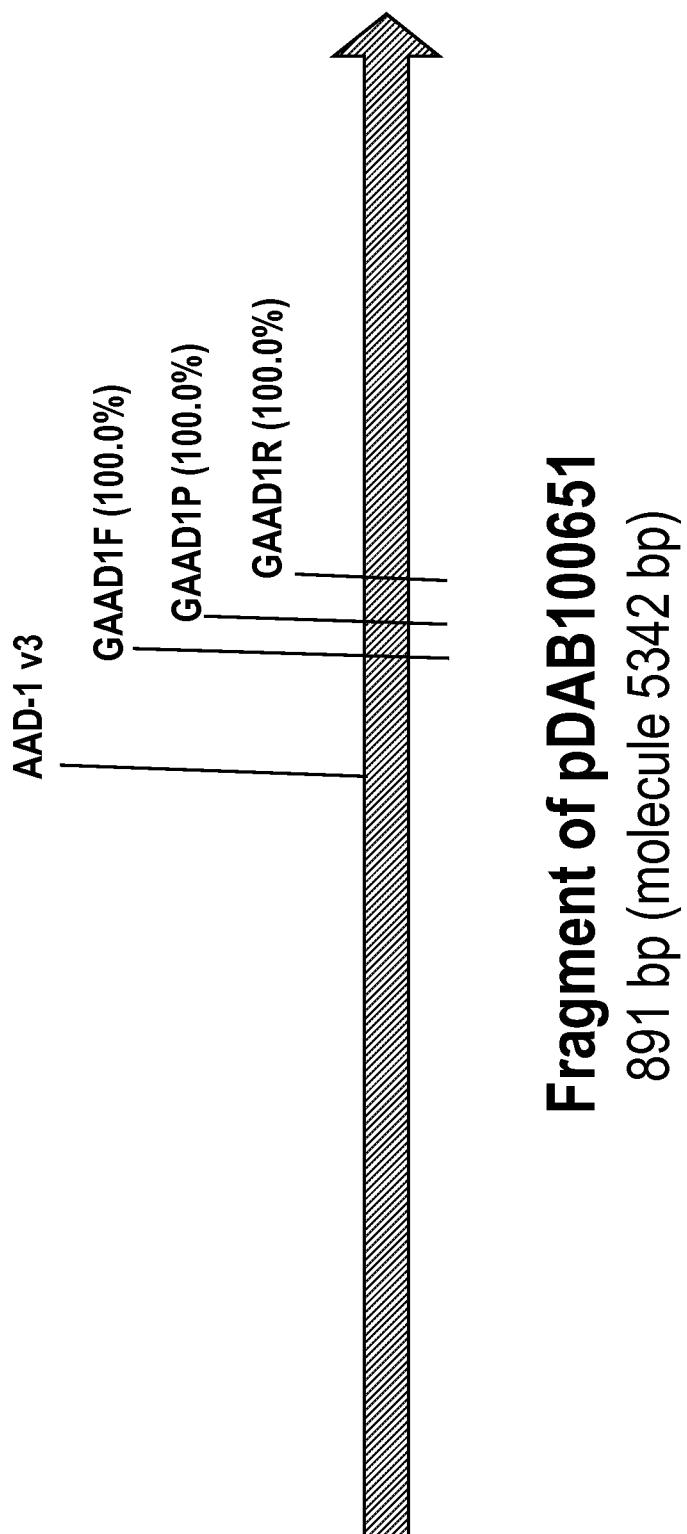
**FIG. 39**

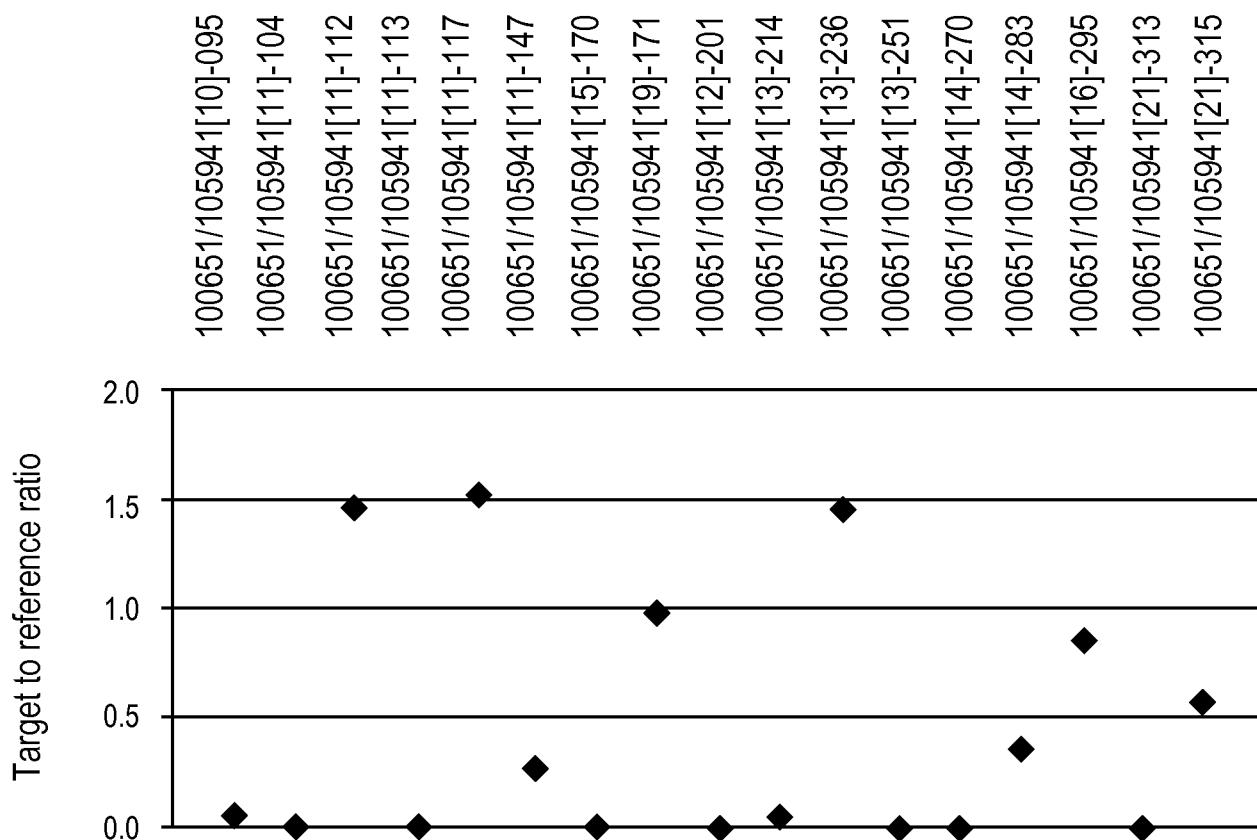
58/61



Fragment of pDAB100640
2099 bp (molecule 10694 bp)

FIG. 40



**FIG. 42**

61/61

A		SEQ ID NO: 480	C G C G C C G C T T T G C A G T T	A T C C N C T A G G G A C A G G A T T G C C A C C
A9-1	SEQ ID NO: 481	C G C G C C G C T T T G C A G T T	A T C C A C T A G G G A C A G G A T T G C C A C C	
A9-2	SEQ ID NO: 482	C G C G C C G C T T T G C A G T T	A G G G A C A G G A T T G C C A C C	
A9-5	SEQ ID NO: 483	C G C G C C G C T T T G C A G T T	A G G G A C A G G A T T G C C A C C	
A9-6	SEQ ID NO: 484	C G C G C C G C C	A C T A G G G A C A G G A T T G C C A C C	
G8-1	SEQ ID NO: 485	C G C G C C G C T T T G C A G T T	A T C C A C T A G G G A C A G G A T T G C C A C C	
G8-2	SEQ ID NO: 486	C G C G C C G C T T T G C A G T T	A T C C A C T A G G G A C A G G A T T G C C A C C	
G8-5	SEQ ID NO: 487	C G C G C C G C T T T G C A G T T	A T C C A C T A G G G A C A G G A T T G C C A C C	
G8-6	SEQ ID NO: 488	C G C G C C G C T T T G C A G T T	A T C C A C T A G G G A C A G G A T T G C C A C C	
G9-1	SEQ ID NO: 489	C G C G C C G C T T T G C A	C T A G G G A C A G G A T T G C C A C C	
G9-2	SEQ ID NO: 490	C G C G C C G C T T T G C A G T T	A T C C A C T A G G G A C A G G A T T G C C A C C	
G9-5	SEQ ID NO: 491	C G C G C C G C T T T G C A G T T	A T C C A C T A G G G A C A G G A T T G C C A C C	
G9-6	SEQ ID NO: 492	C G C G C C G C T T T G C A G T T	A T C C A C T A G G G A C A G G A T T G C C A C C	
H9-1	SEQ ID NO: 493	C G C G C C G C T T T G C A G T T	A T C C A C T A G G G A C A G G A T T G C C A C C	
H9-2	SEQ ID NO: 494	C G C G C C G C T T T G C A G T T	A G G G A C A G G A T T G C C A C C	
H9-5	SEQ ID NO: 495	C G C G C C G C T T T G C A G T T	A T C C A C T A G G G A C A G G A T T G C C A C C	
H9-6	SEQ ID NO: 496	C G C G C C G C T T T G C A G T T	A T C C A C T A G G G A C A G G A T T G C C A C C	

B		SEQ ID NO: 497	C G C G C C G C C T T G C A G T T	A A A C T G C A A A A G G C G G G A T C
A3	SEQ ID NO: 498	C G C G C C G C C T T G C A G T T	A A A C T G C A A A A G G C G G G A T C	
F1	SEQ ID NO: 499	C G C G C C G C C T T G C A G T T	A A A A C T G C A A A A G G C G G G A T C	
F2	SEQ ID NO: 500	C G C G C C G C C T T G C A G T T	A A A A C T G C A A A A G G C G G G A T C	
G1	SEQ ID NO: 501	C G C G C C G C C T T G C A G T T	A A A A C T G C A A A A G G C G G G A T C	
G2	SEQ ID NO: 502	C G C G C C G C C T T G C A G T T	A A A A C T G C A A A A G G C G G G A T C	
G4	SEQ ID NO: 503	C G C G C C G C C T T G C A G T T	A A A A C T G C A A A A G G C G G G A T C	
H1	SEQ ID NO: 504	C G C G C C G C C T T G C A G T T	A A A A C T G C A A A A G G C G G G A T C	
H4	SEQ ID NO: 505	C G C G C C G C C T T G C A G T T	A A A A G G C G G G A T C	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US13/39979

A. CLASSIFICATION OF SUBJECT MATTER IPC(8) - C12N 15/87, 15/00, 5/10 (2013.01) USPC - 435/462, 455; 536/23.1 According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC(8): C12N 15/87, 15/00 (2013.01) USPC: 435/462, 455, 320.1, 440; 536/23.1, 22.1, 18.7, 1.11		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) MicroPatent (US-G, US-A, EP-A, EP-B, WO, JP-bib, DE-C,B, DE-A, DE-T, DE-U, GB-A, FR-A); Google; Google Scholar; DialogPRO; IEEE; exogenous, targeted, integration, 'homology-independent,' 'homology arm,' length, nucleic, sequence, polynucleotide, 'DNA,' 'RNA,' nuclease		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	POVIRK, L. Role Of Braca1 In Nonhomologous DNA End Joining. U.S. Army Medical Research and Material Command. Award Number: DAMD17-03-1-0620. September 2004. pp 1-11; entire document.	1, 2, 3/1, 3/2
Y	US 2011/0207221 A1 (COST, GJ et al.) August 25, 2011; paragraph [0041]	2, 3/2
A	MALANOWSKA, K et al. CTnDOT Integrase Performs Ordered Homology-dependent And Homology-independent Strand Exchanges. Nucleic Acids Research. 24 August 2007. Vol. 35, No. 17; pp 5861-5873; entire document. DOI: 10.1093/nar/gkm637.	1-3
A	ORLANDO, SJ et al. Zinc-finger Nuclease-driven Targeted Integration Into Mammalian Genomes Using Donors With Limited Chromosomal Homology. Nucleic Acids Research. 08 June 2010. Vol. 38, No. 15; pp e152; entire document. DOI: 10.1093/nar/gkq512.	1-3
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/>		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 20 September 2013 (20.09.2013)	Date of mailing of the international search report 27 SEP 2013	
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US13/39979

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.: 4-18 because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

摘要

本文公开的是用于供体分子对细胞基因组的不依赖于同源性的靶向插入的方法和组合物。

